Structural-Functional Transition in Glaucoma Assessment Edition2: Structured description of the challenge design

CHALLENGE ORGANIZATION

Title

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

Structural-Functional Transition in Glaucoma Assessment Edition2

Challenge acronym

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

STAGE2

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.

This challenge is to predict the result of the visual field (VF) test using multi-modal ophthalmic images, including volume optical coherence tomography (OCT) scans centered at the fovea and color fundus photography (CFP). Unlike STAGE challenge in last year, in 2024, we want players to use multi-modal data to make predictions of the VF. OCT is now the most widely used imaging modality used in ophthalmology tests, which provides objective cross-sectional information of the structures in the fundus, facilitating the physician's observation of structural thickness changes, and is an important basis in the diagnosis of glaucoma [1]. There is now much evidence to support the role of OCT imaging in the detection of glaucoma [2-5]. CFP is the most common surface imaging mode of fundus structure, which is convenient and non-invasive. A VF test is a reference standard examination to assess visual function. It is a subjective examination that requires the subject to remain calm and focused and to cooperate with the physician. The monocular visual field examination takes approximately 15 minutes. It is the clinical standard to decide whether there is glaucomatous optic nerve damage [6]. In contrast, a monocular volume OCT scan or CFP takes only about 3 seconds. Furthermore, there is a moderate to good correlation between retinal layer thickness calculated from OCT scans, or the cup-to-disc ratio, disc rim and retinal nerve fiber defect morphology observed in the CFP and central VF sensitivities or other markers of optic nerve function [7]. Therefore, this challenge focuses on how to predict functional VF information using objective and easy-to-acquire structural OCT images and CFPs. Three tasks are proposed for this challenge: 1) mean deviation (MD) value prediction using CFP; 2) sensitivity map prediction using multi-modal data; 3) pattern deviation probability map prediction using multi-modal data. Our challenge will provide 400 volume OCT and CFP paired data and corresponding MD value, sensitivity map, and pattern deviation probability map labels of the VF test report. Of these, 200 multi-modal data and corresponding labels will be released to the teams for model training in the preliminary round. 100 multi-modal data will also be released in the preliminary round, and the evaluation platform will be opened for the teams to validate and tune their models based on the preliminary leaderboard.

The remaining 100 multi-modal data will be released in the final round for the evaluation of the model. From the technical point of view, this challenge is concerned with computer vision studies, among which, tasks 1 and 2 involve metric regression problems, and task 3 involves a classification problem. These studies are essential in computer-aided clinical diagnosis. From a biomedical perspective, this challenge is to seek the mapping relationship between the fundus structure and visual function, which is important for understanding the underlying causes of visual defects.

Challenge keywords

List the primary keywords that characterize the challenge.challenge_

Glaucoma assessment, multi-modal data, macular OCT, color fundus photography, visual field test

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.

11th MICCAI Workshop on Ophthalmic Medical Image Analysis (OMIA XI)

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.

We expect 200 teams to register for the challenge and more than 1000 person-time to submit results. In the end, we will select 10-15 teams (approximately 30 participants) to advance to the finals and participate in the OMIA X I workshop in person.
In MICCAI 2018, we organized the 1st REFUGE challenge, where 334 participants registered and 239 submitted results and participated in, which 12 teams approximately 45 participants participated in the final at OMIA5.
In MICCAI 2019, we organized the AGE challenge, where 414 participants registered and 261 submitted results and participated, of which 8 teams approximately 30 participants participated in the final at OMIA6.
In MICCAI 2020, we organized the 2nd REFUGE challenge, where 1547 participants registered and 961 submitted results and participated, of which 22 teams approximately 60 participants participated in the final at OMIA7.
In MICCAI 2021, we organized the GAMMA contest, where 388 teams (almost 1500 participants) registered, 54 teams (almost 220 participants) submitted the online set results during the preliminary, and 10 teams (almost 40 participants) were selected to participate in the final at OMIA8.
In MICCAI 2022, we organized the GOALS challenge, a total of more than 500 teams registered, more than 3000 person-time submitted the results, and finally, 15 teams (almost 60 participants) emerged from the preliminaries to participate in the final at OMIA9.

In MICCAI 2023, we organized the STAGE challenge, a total of more than 150 teams registered, more than 630 person-time submitted the results, and finally, 7 teams (almost 20 participants) emerged from the preliminaries to participate in the final at OMIA-X.

Publication and future plans

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

We will prepare a challenge review paper for IEEE transactions on medical imaging (TMI) or Medical Image Analysis (MIA). The review papers of REFUGE1 (2018), AGE (2019), ADAM (2020), GAMMA (2021) have been published in MIA (DOI: 10.1016/j.media.2019.101570, DOI: 10.1016/j.media.2020.101798, and DOI: 10.1016/j.media.2023.102938) and TMI (DOI: 10.1109/TMI.2022.3172773). The review papers of REFUGE2(2020), and GOALS (2022) have been or will be submitted to TMI or MIA.

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.

One projector and three microphones are needed.
A computing platform that can make the evaluation based on the results submitted by each team is needed. The specific ranking evaluation algorithm which needs to consider the results of both online and onsite sets will be provided by the organizer, and only a computer is required on site.

TASK 1: Mean deviation (MD) value prediction using CFP

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.

This task is to predict the mean deviation (MD) value of the visual field test in 24-2 mode based on color fundus photographs. MD value [8] reflects the decrease in mean visual sensitivity caused by the death of ganglion cells in glaucoma, and it is around 0 dB in normal individuals. In clinical diagnosis, the MD value is a key parameter in both the H-P-A and Mills visual grading methods [9,10]. Therefore, in this challenge, we treat the prediction of this key value as the first task to predict functional information based on the structural images. From a technical point of view, this task is to regress a value based on a 2D image. In computer vision, this regression task is common research. However, the task is still challenging because the mapping relationship between fundus structure and visual function is not yet clear. From a biomedical point of view, a visual field test is the most fundamental tool for the diagnosis and management of glaucoma and can play a very critical role in the detection of early signs of glaucoma, as well as its MD value is a necessary factor to consider. Thus, the implementation of this task can simplify the time-consuming and subjective examination in clinical glaucoma diagnostic.

Keywords

List the primary keywords that characterize the task.

color fundus photograph, visual field test, mean deviation prediction, regression

ORGANIZATION

Organizers

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

Huihui Fang, Pazhou Lab.,China.
Yanwu Xu, South China University of Technology, China.
Fei Li, Zhongshan Ophthalmic Center, Sun Yat-sen University, China.
Xiulan Zhang, Zhongshan Ophthalmic Center, Sun Yat-sen University, China
Huazhu Fu, Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), Singapore.
José Ignacio Orlando, CONICET/PLADEMA-UNICEN, Argentina. Hrvoje Bogunovi, Medical University of Vienna, Austria.

b) Provide information on the primary contact person.

Huihui Fang (fanghuihuibit@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.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

Open call challenge

Challenge venue and platform

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

11th MICCAI Workshop on Ophthalmic Medical Image Analysis (OMIA XI)

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

aistudio.baidu.com

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

To Be Determined.

Participation policies

a) Define the allowed user interaction of the algorithms assessed (e.g. only (semi-) automatic methods allowed).

Only Fully automatic methods allowed.

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.

The top three teams in the whole challenge will receive the prize as in previous years. The bonus pool is not less than 4000 USD.

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.

Submissions will be scored and ranked based on evaluation metrics. All scores and rankings will be displayed on a publicly available leaderboard on the challenge website. Teams that perform well in the preliminaries will be invited to the finals and the OMIA workshop. All finalist teams are required to submit technical reports and codes,

which are authorized for publication. It is worth noting that the teams from enterprises can only publish technical reports if they cannot disclose the code due to confidentiality factors.

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 top teams in the final will be invited, with a maximum of 2 authors per team, to contribute to a joint journal paper(s) describing and summarizing the methods used and results found in this challenge. The paper will be submitted to a high-impact journal in the field. The participating teams could publish their own methods and results separately.

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 algorithm output will be requested to be sent to the AI Studio platform and then be automatically evaluated. And, after the preliminary stage, the organizer will collect the codes and results of the top 20-30 teams for verification, and the top 10 teams who are willing to submit and have valid codes will enter the final. (The number of teams collecting materials and the number of teams in the final will be adjusted according to the actual number of participants). Submission instructions will be available on the AI Studio platform's challenge page.

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.

During the preliminary, each team can submit at most 5 times per day. Once the quota of submission is exhausted, it will not be able to submit within the same day; the score for each submission will be displayed on the team's submission page, but only the best historical result will be displayed on the leaderboard page. (if the new submission results are better than the previous submission results, the performance of the participant in the leaderboard will be automatically updated and covered. On the other hand, the rankings remain the same.) During the finals, each team has only one chance to submit its 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

2024/06/01 Start the challenge registration
2024/07/01 Release training dataset for download
2024/07/08 Release preliminary dataset for download and open the entry of preliminary leaderboard evaluation
2024/08/18 Close the entrance of the preliminary leaderboard, and the top 30 teams submit materials for review
2024/09/01 Deadline for submission of review materials
2024/09/08 Determine 15 teams for the finals
2024/09/18 Release the encrypted final dataset to the final teams
2024/09/20 Release the password of the final dataset and submit the final result within a limited time
2024/10/06 Announce the ranking and award of the final in MICCAI 2024 OMIA Workshop

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).

All the data have been collected after obtaining IRB approval.

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

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.

We will provide evaluation code on Github.

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

After the challenge, we will release the finalist teams' technical reports and codes on the challenge website. It is worth noting that enterprise teams that cannot disclose code due to confidentiality factors will only release technical reports.

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 prize will be sponsored by the organizers' institute, the members of the institute are not allowed to list in the leaderboard in this challenge. The challenge dataset (original images and labels) will be provided by Prof. Zhang Xiulan's team at Zhongshan Ophthalmic Center, Sun Yat-sen University, whose members are not allowed to list in the leaderboard in this challenge. No one else will have access to the test case labels except the challenge organizers.

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

Regression

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.

People who visit the eye hospital or ophthalmology department in the general hospital.

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

People who visit the eye hospital or ophthalmology department in the general hospital.

Imaging modality(ies)

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

Color fundus photography(CFP), CFP is currently the most economical, non-invasive imaging modality for inspecting the retina.

Context information

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

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

Color fundus photograph, and left or right eye information.

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

Sex, and age

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 are color fundus photographs collected from the subjects' eyes.
The research content is the fundus structures in the color fundus photographs.

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 fundus structures, such as the optic disc, optic cup, disc rim, and retinal nerve fiber are the algorithm target, i.e., the participating algorithms will be designed to focus on them for predicting the mean visual field defect.

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.

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).

Zeiss Visucam 500 for color fundus photograph data collection, and Humphrey Field Analyzers (Carl Zeiss Meditec, Dublin, CA) for VF test report collection.

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

CFP acquisition process:
(1) Instruct the patient to sit in front of the device relaxed, place the jaw on the jaw bracket, and press the forehead forward close to the device. Adjust the height of the jaw bracket so that the height of the patient's outer canthal is at the level of the eye position marker, and instruct the patient to gaze at the fixed cursor inside the lens of the device.
(2) Aim the scanning head at the center of the patient's pupil, and then gradually advance until a clear image appears on the display screen, and adjust the image until it is clear (including the edge).
(3) Automatic focusing or manual focusing can be selected to adjust the target diopter, iris focusing, pupil position and color fundus image centering.
(4) Adjust the scope of the fundus covered by the scanning frame to be consistent with the site required for clinical examination. The center of the field of view of the photographs were either placed in the optic disc, the macula, or the midpoint of the optic disc and macula.
(5) Ensure that the scanned retina light band is clear and placed in the middle of the display screen observation window, with high signal intensity and uniform brightness.

VF test process:
(1) Instruct the patient to sit in front of the device relaxed, and the eye not being tested will be covered with a patch.
(2) Guiding the patient to look into the center of a bowl-shaped instrument called a perimeter, and the testing eye will have the lens prescription placed in front of it to make sure the patient is seeing as well as possible.
(3) The patient should keep looking at a center target throughout the test. Small, dim lights will begin to appear in different places throughout the bowl, and the patient will press a button whenever the patient sees a light. The machine tracks which lights the patient did not see.
(4) The patient may blink normally during the test and pause the test if the patient feels the need to take a moment.
(5) Reliable VF tests must have fixation loss of less than 20%, a false positive rate lower than 15%, and a false negative rate lower than 33%. If the VF test is unreliable, the patient should rest for a period and then retest.

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.

Zhongshan Ophthalmic Center, Sun Yat-sen University, 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).

Color fundus photographs collection, and VF test were operated by ophthalmologists and technicians with at least 5 years of collecting experience.

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 color fundus image of a human eye and the subject's age and sex information. All these cases have a label of the mean deviation value, which is read from the visual field test report. The visual field test was performed at the same time as the color fundus image collection by the same ophthalmologist.

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

A total of 400 color fundus images are available. The dataset is split into 3 subsets for training (200), preliminary validation (100), and final test (100).

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

Data for this challenge will be randomly selected from the glaucoma study cohort (200) and the high myopia study cohort (200) at the Zhongshan Ophthalmic Center, Sun Yat-sen University. Referring to the previous STAGE Challenge, it is still planned to provide 200 samples in the training step, and 100 samples each in the preliminary and final rounds to verify the model performance.

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.

To have a balanced distribution of samples across the data subsets, the 400 samples will be assigned to ensure a consistent distribution of age groups, gender, and disease conditions. Among them, the distribution of non-glaucoma samples, early glaucoma samples, intermediate glaucoma samples, and advanced glaucoma samples will be at 50%, 25%, 15%, and 10% in each subset. This allocation takes into account both the real-world distribution of glaucomatous samples and the equal class distribution during model training and testing. That is, we will maintain trends in the distribution of samples with varying degrees of glaucoma and also keep that the difference in their numbers is not very large.

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 labels for this MD value prediction task are derived from the report of the visual field (VF) test in clinical practice. All the VFs are automated white-on-white perimetry SITA 24-2 standard/fast VFs acquired by Humphrey Field Analyzers (Carl Zeiss Meditec, Dublin, CA). VF test results are considered reliable if the VF with fixation losses of less than 2/13, and false-positive rate less than 15%, and false-negative rate less than 25% [11]. The mean deviation (MD) value in the VF report will be the ground truth of this task.

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.

None.

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.

None.

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.

None.

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.

Personal information except age and sex will be removed.

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.

None.

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

None.

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)

Symmetric Mean Absolute Percentage Error (SMAPE) will be used to calculate the difference between the predicted value and the target MD value. If both the target and predicted values are 0, then SMAPE for this case will be recorded as 0.

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

For this regression task, the SMAPE measure will be used. Compared to the MAE commonly used for regression

tasks, SMAPE are bound, with the lower bound of 0% implying a perfect fit, and the upper bound of 200% [12,13]. It is friendlier for the score calculation in the challenge and is more reflective of the degree of error. SMAPE is progressively gaining momentum in biomedical applications due to its interesting properties [14,15].

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.

Since SMAPE takes values in [0, 2] and the smaller the SMAPE value the better the prediction, we take the inverse of SMAPE and add an extra 0.1 to the denominator to prevent the denominator from being 0. Therefore, the score of task 1 will be ranged in [0.476, 10].

In the final ranking, we will consider both the preliminary and final performances. Because the teams can adjust their models after seeing the performance on the preliminary dataset through the leaderboard, we set the preliminary weight at 0.3. Compared with the predicted results in the preliminary round, those in the final round can reflect the real generalizability performance of the models due that the final dataset is unknown to the teams. Hence, we set the final weight at 0.7.

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

the missing results will report error information.

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

MD_score = 1/(SMAPE + 0.1). Each team is ranked according to their MD scores on the test dataset. The higher the score the higher the ranking.
In the challenge, the final score of this task will be considered in a combination of the preliminary round and the final round. The final score is MD_score(total) = 0.3*MD_score(preliminary round) + 0.7*MD_score(final round).

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.

Missing data not allowed (incomplete submissions not evaluated).
The statistical significance of the differences in performance of the top-ranked teams was assessed by means of Wilcoxon signed-rank tests (= 0.05).

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

The Wilcoxon signed-rank test [16] is a non-parametric statistical hypothesis test used either to test the location of a set of samples or to compare the locations of two populations using a set of matched samples.

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.

In the challenge review paper, we will have further analysis of the above four points, including combining algorithms via ensembling, inter-algorithm variability, common problems/biases of the submitted methods, and ranking variability.

TASK 2: Sensitivity map prediction using multi-modal data

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 purpose of this task is to use multi-modal data, including macular OCT volumes and color fundus photographs to predict sensitivity maps in visual field tests. The sensitivity map [17], also known as the threshold map, is the raw data obtained during visual field tests, i.e., the photosensitivity observable at each test point. The automatic visual field analyzer uses a threshold cursor intensity to express photosensitivity in dB, with 0 dB being the strongest photosensitivity and 100 dB being the weakest photosensitivity. There are 52 test points in the 24-2 visual field test, and each point in the sensitivity map has a corresponding photosensitivity value, which represents the weakest photosensitivity to which this test point can respond. In glaucoma grading methods, such as H-P-A method and Millfs method [9,10], the sensitivity map of the visual field test is an important grading basis. Therefore, this task requires teams to directly use objective and rapidly available color fundus images and OCT images to predict sensitivity maps of the visual field test. From a technical point of view, this task is a regression task to predict 52 values in each sample. From a biomedical perspective, this task is intended to address structural image to function prediction, which is a topic of great concern in computer-aided disease diagnosis and treatment. If the results are promising for clinical practice, then in the future, instead of subjective, time-consuming, and less reproducible functional tests, physicians can determine a patient's functional loss using functional results predicted based on objective structural images.

Keywords

List the primary keywords that characterize the task.

color fundus photograph, macular OCT, visual field test, sensitivity map prediction

ORGANIZATION

Organizers

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

Huihui Fang, Pazhou Lab.,China.
Yanwu Xu, South China University of Technology, China.
Fei Li, Zhongshan Ophthalmic Center, Sun Yat-sen University, China.
Xiulan Zhang, Zhongshan Ophthalmic Center, Sun Yat-sen University, China
Huazhu Fu, Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), Singapore.
José Ignacio Orlando, CONICET/PLADEMA-UNICEN, Argentina. Hrvoje Bogunovi, Medical University of Vienna, Austria.

b) Provide information on the primary contact person.

Huihui Fang (fanghuihuibit@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.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

Open call challenge

Challenge venue and platform

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

11th MICCAI Workshop on Ophthalmic Medical Image Analysis (OMIA XI)

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

aistudio.baidu.com

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

To Be Determined.

Participation policies

a) Define the allowed user interaction of the algorithms assessed (e.g. only (semi-) automatic methods allowed).

Only Fully automatic methods allowed.

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.

The top three teams in the whole challenge will receive the prize as in previous years. The bonus pool is not less than 4000 USD.

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.

Submissions will be scored and ranked based on evaluation metrics. All scores and rankings will be displayed on a publicly available leaderboard on the challenge website. Teams that perform well in the preliminaries will be invited to the finals and the OMIA workshop. All finalist teams are required to submit technical reports and codes, which are authorized for publication. It is worth noting that the teams from enterprises can only publish technical reports if they cannot disclose the code due to confidentiality factors.

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 top teams in the final will be invited, with a maximum of 2 authors per team, to contribute to a joint journal paper(s) describing and summarizing the methods used and results found in this challenge. The paper will be submitted to a high-impact journal in the field. The participating teams could publish their own methods and results separately.

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 algorithm output will be requested to be sent to the AI Studio platform and then be automatically evaluated. And, after the preliminary stage, the organizer will collect the codes and results of the top 20-30 teams for verification, and the top 10 teams who are willing to submit and have valid codes will enter the final. (The number of teams collecting materials and the number of teams in the final will be adjusted according to the actual number of participants). Submission instructions will be available on the AI Studio platform's challenge page.

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.

During the preliminary, each team can submit at most 5 times per day. Once the quota of submission is exhausted, it will not be able to submit within the same day; the score for each submission will be displayed on the team's submission page, but only the best historical result will be displayed on the leaderboard page. (if the new submission results are better than the previous submission results, the performance of the participant in the leaderboard will be automatically updated and covered. On the other hand, the rankings remain the same.) During the finals, each team has only one chance to submit its 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

2024/06/01 Start the challenge registration
2024/07/01 Release training dataset for download
2024/07/08 Release preliminary dataset for download and open the entry of preliminary leaderboard evaluation
2024/08/18 Close the entrance of the preliminary leaderboard, and the top 30 teams submit materials for review
2024/09/01 Deadline for submission of review materials
2024/09/08 Determine 15 teams for the finals
2024/09/18 Release the encrypted final dataset to the final teams
2024/09/20 Release the password of the final dataset and submit the final result within a limited time
2024/10/06 Announce the ranking and award of the final in MICCAI 2024 OMIA Workshop

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).

All the data have been collected after obtaining IRB approval.

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

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.

We will provide evaluation code on Github.

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

After the challenge, we will release the finalist teams' technical reports and codes on the challenge website. It is worth noting that enterprise teams that cannot disclose code due to confidentiality factors will only release technical reports.

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 prize will be sponsored by the organizers' institute, the members of the institute are not allowed to list in the leaderboard in this challenge. The challenge dataset (original images and labels) will be provided by Prof. Zhang Xiulan's team at Zhongshan Ophthalmic Center, Sun Yat-sen University, whose members are not allowed to list in the leaderboard in this challenge. No one else will have access to the test case labels except the challenge organizers.

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

Regression

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.

People who visit the eye hospital or ophthalmology department in the general hospital.

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

People who visit the eye hospital or ophthalmology department in the general hospital.

Imaging modality(ies)

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

(1) Optical Coherence Tomography (OCT), OCT uses the basic principles of the weakly coherent optical interferometer to detect the back reflection or several scattering signals of the incident weakly coherent light at different depths of biological tissue. OCT is a noninvasive and non-contact imaging modality for morphological analysis and diagnosis of retinal abnormality.
(2) Color fundus photography(CFP), CFP is currently the most economical, non-invasive imaging modality for inspecting the retina.

Context information

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

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

Macular OCT Volume, color fundus photograph, and left or right eye information.

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

Sex, and age

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 are macular OCT volumes and color fundus photographs collected from the subjects' eyes.
The research content is the fundus structures in multi-modal data.

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 fundus structures, such as the optic nerve fiber layer and ganglion complex cell layer in OCT scans and the optic disc, optic cup, and disc rim in color fundus photographs, are the algorithm target, i.e., the participating algorithms will be designed to focus on them for predicting the sensitivity map.

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.

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).

Zeiss Visucam 500 for color fundus photograph data collection, Topcon DRI OCT Triton for OCT data collection, and Humphrey Field Analyzers (Carl Zeiss Meditec, Dublin, CA) for VF test report collection.

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

CFP acquisition process:
(1) Instruct the patient to sit in front of the device relaxed, place the jaw on the jaw bracket, and press the forehead forward close to the device. Adjust the height of the jaw bracket so that the height of the patient's outer canthal is at the level of the eye position marker, and instruct the patient to gaze at the fixed cursor inside the lens of the device.
(2) Aim the scanning head at the center of the patient's pupil, and then gradually advance until a clear image appears on the display screen, and adjust the image until it is clear (including the edge).

(3) Automatic focusing or manual focusing can be selected to adjust the target diopter, iris focusing, pupil position and color fundus image centering.
(4) Adjust the scope of the fundus covered by the scanning frame to be consistent with the site required for clinical examination. The center of the field of view of the photographs were either placed in the optic disc, the macula, or the midpoint of the optic disc and macula.
(5) Ensure that the scanned retina light band is clear and placed in the middle of the display screen observation window, with high signal intensity and uniform brightness.

OCT acquisition process:
(1) Instruct the patient to sit in front of the device relaxed, place the jaw on the jaw bracket, and press the forehead forward close to the device. Adjust the height of the jaw bracket so that the height of the patient's outer canthal is at the level of the eye position marker, and instruct the patient to gaze at the fixed cursor inside the lens of the device.
(2) Aim the scanning head at the center of the patient's pupil, and then gradually advance until a clear image appears on the display screen, and adjust the image until it is clear (including the edge).
(3) Automatic focusing or manual focusing can be selected to adjust the target diopter, iris focusing, pupil position and OCT image centering.
(4) Adjust the scope of the fundus covered by the scanning frame to be consistent with the site required for clinical examination.
(5) Ensure that the scanned retina light band is clear and placed in the middle of the display screen observation window, with high signal intensity and uniform brightness.

VF test process:
(1) Instruct the patient to sit in front of the device relaxed, and the eye not being tested will be covered with a patch.
(2) Guiding the patient to look into the center of a bowl-shaped instrument called a perimeter, and the testing eye will have the lens prescription placed in front of it to make sure the patient is seeing as well as possible.
(3) The patient should keep looking at a center target throughout the test. Small, dim lights will begin to appear in different places throughout the bowl, and the patient will press a button whenever the patient sees a light. The machine tracks which lights the patient did not see.
(4) The patient may blink normally during the test and pause the test if the patient feels the need to take a moment.
(5) Reliable VF tests must have fixation loss of less than 20%, a false positive rate lower than 15%, and a false negative rate lower than 33%. If the VF test is unreliable, the patient should rest for a period and then retest.

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.

Zhongshan Ophthalmic Center, Sun Yat-sen University, 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).

Original macular OCT images, color fundus photographs collection, and VF test were operated by ophthalmologists and technicians with at least 5 years of collecting experience.

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 OCT volume and a color fundus image of a human eye, and the subject's age and sex information. All these cases have a label of the sensitivity map, which is read from the visual field test report. The visual field test was performed at the same time as the color fundus image and OCT collection by the same ophthalmologist.

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

A total of 400 macular OCT volumes and color fundus images paired data are available. The dataset is split into 3 subsets for training (200), preliminary validation (100), and final test (100).

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

Data for this challenge will be randomly selected from the glaucoma study cohort (200) and the high myopia study cohort (200) at the Zhongshan Ophthalmic Center, Sun Yat-sen University. Referring to the previous STAGE Challenge, it is still planned to provide 200 samples in the training step, and 100 samples each in the preliminary and final rounds to verify the model performance.

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.

To have a balanced distribution of samples across the data subsets, the 400 samples will be assigned to ensure a consistent distribution of age groups, gender, and disease conditions. Among them, the distribution of non-glaucoma samples, early glaucoma samples, intermediate glaucoma samples, and advanced glaucoma samples will be at 50%, 25%, 15%, and 10% in each subset. This allocation takes into account both the real-world distribution of glaucomatous samples and the equal class distribution during model training and testing. That is, we will maintain trends in the distribution of samples with varying degrees of glaucoma and also keep that the difference in their numbers is not very large.

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 labels for this sensitivity map prediction task are derived from the report of the visual field (VF) test in clinical practice. All the VFs are automated white-on-white perimetry SITA 24-2 standard/fast VFs acquired by Humphrey Field Analyzers (Carl Zeiss Meditec, Dublin, CA). VF test results are considered reliable if the VF with fixation losses of less than 2/13, and false-positive rate less than 15%, and false-negative rate less than 25% [11]. The sensitivity map in the VF report will be the ground truth of this task. The label contains 52 values (value of 0-100) corresponding to 52 positions in the sensitivity map.

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.

None.

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.

None.

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.

None.

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.

Each raw 3D OCT volume will be stored in 256 images and released to the participating teams. And personal information except age and sex will be removed.

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.

None.

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

None.

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)

Symmetric Mean Absolute Percentage Error (SMAPE) will be used to calculate the difference between each predicted value and the target value in the sensitivity map. The average of these SMAPEs (ASMAPE) will be the final evaluation metric for every case in this task.
If both the target and predicted values are 0, then the SMAPE for the point will be recorded as 0.

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

For this regression task, the SMAPE measure will be used. Compared to the MAE commonly used for regression

tasks, SMAPE are bound, with the lower bound of 0% implying a perfect fit, and the upper bound of 200% [12,13]. It is friendlier for the score calculation in the challenge and is more reflective of the degree of error. SMAPE is progressively gaining momentum in biomedical applications due to its interesting properties [14,15].

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.

Since ASMAPE takes values in [0, 2] and the smaller the ASMAPE value the better the prediction, we take the inverse of ASMAPE and add an extra 0.1 to the denominator to prevent the denominator from being 0. Therefore, the score of task 2 will be ranged in [0.476, 10].

In the final ranking, we will consider both the preliminary and final performances. Because the teams can adjust their models after seeing the performance on the preliminary dataset through the leaderboard, we set the preliminary weight at 0.3. Compared with the predicted results in the preliminary round, those in the final round can reflect the real generalizability performance of the models due that the final dataset is unknown to the teams. Hence, we set the final weight at 0.7.

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

the missing results will report error information.

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

Sensitivity-map_pred_score = 1/(ASMAPE + 0.1). Each team is ranked according to their sensitivity map prediction scores on the test dataset. The higher the score the higher the ranking.
In the challenge, the final score of this task will be considered in a combination of the preliminary round and the final round. The final score is Sensitivity-map_pred_score(total) = 0.3*Sensitivity-map_pred_score(preliminary round) + 0.7*Sensitivity-map_pred_score(final round).

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.

Missing data not allowed (incomplete submissions not evaluated).
The statistical significance of the differences in performance of the top-ranked teams was assessed by means of Wilcoxon signed-rank tests (= 0.05).

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

The Wilcoxon signed-rank test [16] is a non-parametric statistical hypothesis test used either to test the location of a set of samples or to compare the locations of two populations using a set of matched samples.

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.

In the challenge review paper, we will have further analysis of the above four points, including combining algorithms via ensembling, inter-algorithm variability, common problems/biases of the submitted methods, and ranking variability.

TASK 3: Pattern deviation probability map prediction using multi-modal data

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 purpose of this task is to predict the pattern deviation probability map in the visual field test report using multi-modal data, including macular OCT images and color fundus images. In the visual field report, the pattern deviation probability map [17] reflects the statistical significance of the remaining local sensitivity reduction after excluding the generalized sensitivity reduction from the overall sensitivity reduction. That is, the probability that the patient has a visual field defect in a certain area after removing interference factors such as anterior segments (e.g., cataracts, small pupils), which can highlight a meaningful local visual deficit. There are 52 test points in the 24-2 visual field test, and each point belongs to one of the five categories of visual loss in the pattern deviation probability map. These 5 categories of cases are normal visual field, probability of normal visual field less than 5%, less than 2%, less than 1%, and less than 0.5%. Among them, the probability of a normal visual field is less than 0.5% is the most serious case of visual field loss. In H-P-A and Mills glaucoma grading methods [9,10], the results of pattern deviation probability map on visual field test are a critical grading basis. Therefore, this task requires teams to directly use objective and quickly available OCT images and color fundus images to predict the pattern deviation probability map without the subjective and time-consuming visual field test. From a technical point of view, this task is a multi-classification task and is a well-researched topic in computer vision.

Keywords

List the primary keywords that characterize the task.

color fundus photograph, macular OCT, visual field test, pattern deviation probability map prediction

ORGANIZATION

Organizers

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

Huihui Fang, Pazhou Lab.,China.
Yanwu Xu, South China University of Technology, China.
Fei Li, Zhongshan Ophthalmic Center, Sun Yat-sen University, China.
Xiulan Zhang, Zhongshan Ophthalmic Center, Sun Yat-sen University, China
Huazhu Fu, Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), Singapore.
José Ignacio Orlando, CONICET/PLADEMA-UNICEN, Argentina. Hrvoje Bogunovi, Medical University of Vienna, Austria.

b) Provide information on the primary contact person.

Huihui Fang (fanghuihuibit@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.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

Open call challenge

Challenge venue and platform

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

11th MICCAI Workshop on Ophthalmic Medical Image Analysis (OMIA XI)

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

aistudio.baidu.com

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

To Be Determined.

Participation policies

a) Define the allowed user interaction of the algorithms assessed (e.g. only (semi-) automatic methods allowed).

Only Fully automatic methods allowed.

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.

The top three teams in the whole challenge will receive the prize as in previous years. The bonus pool is not less than 4000 USD.

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.

Submissions will be scored and ranked based on evaluation metrics. All scores and rankings will be displayed on a publicly available leaderboard on the challenge website. Teams that perform well in the preliminaries will be invited to the finals and the OMIA workshop. All finalist teams are required to submit technical reports and codes, which are authorized for publication. It is worth noting that the teams from enterprises can only publish technical reports if they cannot disclose the code due to confidentiality factors.

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 top teams in the final will be invited, with a maximum of 2 authors per team, to contribute to a joint journal paper(s) describing and summarizing the methods used and results found in this challenge. The paper will be submitted to a high-impact journal in the field. The participating teams could publish their own methods and results separately.

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 algorithm output will be requested to be sent to the AI Studio platform and then be automatically evaluated. Submission instructions will be available on the AI Studio platform's challenge page.

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.

During the preliminary, each team can submit at most 5 times per day. Once the quota of submission is exhausted, it will not be able to submit within the same day; the score for each submission will be displayed on the team's submission page, but only the best historical result will be displayed on the leaderboard page. (if the new submission results are better than the previous submission results, the performance of the participant in the leaderboard will be automatically updated and covered. On the other hand, the rankings remain the same.) During the finals, each team has only one chance to submit its 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

2024/06/01 Start the challenge registration
2024/07/01 Release training dataset for download
2024/07/08 Release preliminary dataset for download and open the entry of preliminary leaderboard evaluation
2024/08/18 Close the entrance of the preliminary leaderboard, and the top 30 teams submit materials for review
2024/09/01 Deadline for submission of review materials
2024/09/08 Determine 15 teams for the finals
2024/09/18 Release the encrypted final dataset to the final teams
2024/09/20 Release the password of the final dataset and submit the final result within a limited time
2024/10/06 Announce the ranking and award of the final in MICCAI 2024 OMIA Workshop

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).

All the data have been collected after obtaining IRB approval.

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

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.

We will provide evaluation code on Github.

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

After the challenge, we will release the finalist teams' technical reports and codes on the challenge website. It is worth noting that enterprise teams that cannot disclose code due to confidentiality factors will only release technical reports.

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 prize will be sponsored by the organizers' institute, the members of the institute are not allowed to list in the leaderboard in this challenge. The challenge dataset (original images and labels) will be provided by Prof. Zhang Xiulan's team at Zhongshan Ophthalmic Center, Sun Yat-sen University, whose members are not allowed to list in the leaderboard in this challenge. No one else will have access to the test case labels except the challenge organizers.

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

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.

People who visit the eye hospital or ophthalmology department in the general hospital.

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

People who visit the eye hospital or ophthalmology department in the general hospital.

Imaging modality(ies)

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

(1) Optical Coherence Tomography (OCT), OCT uses the basic principles of the weakly coherent optical interferometer to detect the back reflection or several scattering signals of the incident weakly coherent light at different depths of biological tissue. OCT is a noninvasive and non-contact imaging modality for morphological analysis and diagnosis of retinal abnormality.
(2) Color fundus photography(CFP), CFP is currently the most economical, non-invasive imaging modality for inspecting the retina.

Context information

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

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

Macular OCT Volume, color fundus photograph, and left or right eye information.

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

Sex, and age

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 are macular OCT volumes and color fundus photographs collected from the subjects' eyes.
The research content is the fundus structures in multi-modal data.

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 fundus structures, such as the optic nerve fiber layer and ganglion complex cell layer in OCT scans and the optic disc, optic cup, and disc rim in color fundus photographs, are the algorithm target, i.e., the participating algorithms will be designed to focus on them for predicting the pattern deviation probability map.

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)).

Precision, 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).

Zeiss Visucam 500 for color fundus photograph data collection, Topcon DRI OCT Triton for OCT data collection, and Humphrey Field Analyzers (Carl Zeiss Meditec, Dublin, CA) for VF test report collection.

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

CFP acquisition process:
(1) Instruct the patient to sit in front of the device relaxed, place the jaw on the jaw bracket, and press the forehead forward close to the device. Adjust the height of the jaw bracket so that the height of the patient's outer canthal is at the level of the eye position marker, and instruct the patient to gaze at the fixed cursor inside the lens of the device.
(2) Aim the scanning head at the center of the patient's pupil, and then gradually advance until a clear image appears on the display screen, and adjust the image until it is clear (including the edge).

(3) Automatic focusing or manual focusing can be selected to adjust the target diopter, iris focusing, pupil position and color fundus image centering.
(4) Adjust the scope of the fundus covered by the scanning frame to be consistent with the site required for clinical examination. The center of the field of view of the photographs were either placed in the optic disc, the macula, or the midpoint of the optic disc and macula.
(5) Ensure that the scanned retina light band is clear and placed in the middle of the display screen observation window, with high signal intensity and uniform brightness.

OCT acquisition process:
(1) Instruct the patient to sit in front of the device relaxed, place the jaw on the jaw bracket, and press the forehead forward close to the device. Adjust the height of the jaw bracket so that the height of the patient's outer canthal is at the level of the eye position marker, and instruct the patient to gaze at the fixed cursor inside the lens of the device.
(2) Aim the scanning head at the center of the patient's pupil, and then gradually advance until a clear image appears on the display screen, and adjust the image until it is clear (including the edge).
(3) Automatic focusing or manual focusing can be selected to adjust the target diopter, iris focusing, pupil position and OCT image centering.
(4) Adjust the scope of the fundus covered by the scanning frame to be consistent with the site required for clinical examination.
(5) Ensure that the scanned retina light band is clear and placed in the middle of the display screen observation window, with high signal intensity and uniform brightness.

VF test process:
(1) Instruct the patient to sit in front of the device relaxed, and the eye not being tested will be covered with a patch.
(2) Guiding the patient to look into the center of a bowl-shaped instrument called a perimeter, and the testing eye will have the lens prescription placed in front of it to make sure the patient is seeing as well as possible.
(3) The patient should keep looking at a center target throughout the test. Small, dim lights will begin to appear in different places throughout the bowl, and the patient will press a button whenever the patient sees a light. The machine tracks which lights the patient did not see.
(4) The patient may blink normally during the test and pause the test if the patient feels the need to take a moment.
(5) Reliable VF tests must have fixation loss of less than 20%, a false positive rate lower than 15%, and a false negative rate lower than 33%. If the VF test is unreliable, the patient should rest for a period and then retest.

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.

Zhongshan Ophthalmic Center, Sun Yat-sen University, 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).

Original macular OCT images, color fundus photographs collection, and VF test were operated by ophthalmologists and technicians with at least 5 years of collecting experience.

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 OCT volume and a color fundus image of a human eye, and the subject's age and sex information. All these cases have a label of the pattern deviation probability map, which is read from the visual field test report. The visual field test was performed at the same time as the color fundus image and OCT collection by the same ophthalmologist.

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

A total of 400 macular OCT volumes and color fundus images paired data are available. The dataset is split into 3 subsets for training (200), preliminary validation (100), and final test (100).

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

Data for this challenge will be randomly selected from the glaucoma study cohort (200) and the high myopia study cohort (200) at the Zhongshan Ophthalmic Center, Sun Yat-sen University. Referring to the previous STAGE Challenge, it is still planned to provide 200 samples in the training step, and 100 samples each in the preliminary and final rounds to verify the model performance.

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.

To have a balanced distribution of samples across the data subsets, the 400 samples will be assigned to ensure a consistent distribution of age groups, gender, and disease conditions. Among them, the distribution of non-glaucoma samples, early glaucoma samples, intermediate glaucoma samples, and advanced glaucoma samples will be at 50%, 25%, 15%, and 10% in each subset. This allocation takes into account both the real-world distribution of glaucomatous samples and the equal class distribution during model training and testing. That is, we will maintain trends in the distribution of samples with varying degrees of glaucoma and also keep that the difference in their numbers is not very large.

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 labels for this pattern deviation probability map prediction task are derived from the report of the visual field (VF) test in clinical practice. All the VFs are automated white-on-white perimetry SITA 24-2 standard/fast VFs acquired by Humphrey Field Analyzers (Carl Zeiss Meditec, Dublin, CA). VF test results are considered reliable if the VF with a fixation loss of less than 2/13, a false-positive rate of less than 15%, and a false-negative rate of less than 25% [11]. The pattern deviation probability map in the VF report will be the ground truth of this task. The label also contains 52 values corresponding to 52 positions in the pattern deviation probability map. Unlike task 2, these 52 values are taken to be 0-4, representing the degree of visual field deficit at that location falling into one of the five categories.

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.

None.

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.

None.

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.

None.

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.

Each raw 3D OCT volume will be stored in 256 images and released to the participating teams. And personal information except age and sex will be removed.

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.

None.

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

None.

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 micro-F1 score will be used to evaluate the multi-classification results (points belong to the normal category, the normal likelihood less than 5% category, the normal likelihood less than 2% category, the normal likelihood less than 1% category, and normal likelihood less than 0.5% category) of the 52 points in the pattern deviation probability map. Sensitivity and specificity will be used to evaluate the binary-classification accuracy of the pattern deviation probability map with 52 points, including normal points and the points with visual loss.

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

Task 3 is a multi-classification task, and here, we first chose to evaluate the submission results using micro-F1 scores. For the evaluation of the multi-classification task, the micro-average is more concerned with the imbalance of the category distribution than the macro-average [18]. While in our pattern deviation probability map prediction task, points with a severe degree of visual field loss are less likely to occur compared to points with slight loss, i.e., the categories are unbalanced. Therefore, we use the micro-average to calculate the F1 score. In addition, to determine whether the model is accurate in discriminating the presence of visual loss, we will consider category 0 and other categories 1-4 of the classification results submitted by the participating teams as two categories and evaluate the classification results of these two categories with sensitivity and specificity.

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.

Task 3 is a multi-classification task, so the score calculation formula assigns a weight of 0.5 to the Micro F1 metric. In addition, due to the difficulty of the multi-classification task, we also designed two metrics of sensitivity and specificity to test the model's binary prediction results on whether the test point has visual loss. The weights of these two metrics in the score calculation formula are 0.25. Since the Micro-F1 score, sensitivity, and specificity take values in the range [0,1], the task 3 calculation formula is multiplied by 10 for all metrics to ensure that the score takes value in the interval [0, 10].

In the final ranking, we will consider both the preliminary and final performances. Because the teams can adjust their models after seeing the performance on the preliminary dataset through the leaderboard, we set the preliminary weight at 0.3. Compared with the predicted results in the preliminary round, those in the final round can reflect the real generalizability performance of the models due that the final dataset is unknown to the teams. Hence, we set the final weight at 0.7.

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

the missing results will report error information.

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

The score calculation for the pattern deviation probability map (PDPM) is PDPM_score = 5*Micro_F1+2.5*Sensitivity + 2.5*Specificity.
In the challenge, the final score of this task will be considered in a combination of the preliminary round and the final round. The final score is PDPM_score(total) = 0.3*PDPM_score(preliminary round) + 0.7*PDPM_score(final round).

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.

Missing data not allowed (incomplete submissions not evaluated).
The statistical significance of the differences in performance of the top-ranked teams was assessed by means of Wilcoxon signed-rank tests (= 0.05).

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

The Wilcoxon signed-rank test [16] is a non-parametric statistical hypothesis test used either to test the location of a set of samples or to compare the locations of two populations using a set of matched samples.

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.

In the challenge review paper, we will have further analysis of the above four points, including combining algorithms via ensembling, inter-algorithm variability, common problems/biases of the submitted methods, and 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] Schuman J S, Hee M R, Arya A V, et al. Optical coherence tomography: a new tool for glaucoma diagnosis[J]. Current opinion in ophthalmology, 1995, 6(2): 89-95.
[2] Chen M J, Chang Y F, Kuo Y S, et al. Macular ganglion cell-inner plexiform vs retinal nerve fiber layer measurement to detect early glaucoma with superior or inferior hemifield defects[J]. Journal of the Chinese Medical Association, 2019, 82(4): 335-339.
[3] Kim K E, Park K H. Macular imaging by optical coherence tomography in the diagnosis and management of glaucoma[J]. British Journal of Ophthalmology, 2017.
[4] Wu Z, Weng D S D, Thenappan A, et al. Evaluation of a region-of-interest approach for detecting progressive glaucomatous macular damage on optical coherence tomography[J]. Translational Vision Science & Technology, 2018, 7(2): 14-14.

[5] Asrani S, Rosdahl J A, Allingham R R. Novel software strategy for glaucoma diagnosis: asymmetry analysis of retinal thickness[J]. Archives of ophthalmology, 2011, 129(9): 1205-1211.
[6] Garway-Heath D F, Poinoosawmy D, Fitzke F W, et al. Mapping the visual field to the optic disc in normal tension glaucoma eyes[J]. Ophthalmology, 2000, 107(10): 1809-1815.
[7] Mohammadzadeh V, Fatehi N, Yarmohammadi A, et al. Macular imaging with optical coherence tomography in glaucoma[J]. Survey of ophthalmology, 2020, 65(6): 597-638.
[8] Banegas S A, Antón A, Morilla A, et al. Evaluation of the retinal nerve fiber layer thickness, the mean deviation, and the visual field index in progressive glaucoma[J]. Journal of glaucoma, 2016, 25(3): e229-e235.
[9] Hodapp, E., Parrish, R.K. and Anderson, D.R. Clinical Decisions in Glaucoma. Mosby, Maryland Heights, 52-61, 1993.
[10] Mills R P, Budenz D L, Lee P P, et al. Categorizing the stage of glaucoma from pre-diagnosis to end-stage disease[J]. American journal of ophthalmology, 2006, 141(1): 24-30.
[11] Li F, Song D, Chen H, et al. Development and clinical deployment of a smartphone-based visual field deep learning system for glaucoma detection[J]. NPJ digital medicine, 2020, 3(1): 123.
[12] Makridakis S. Accuracy measures: theoretical and practical concerns[J]. International journal of forecasting, 1993, 9(4): 527-529.
[13] Chicco D, Warrens M J, Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation[J]. PeerJ Computer Science, 2021, 7: e623.
[14] Purkayastha S, Bhattacharyya R, Bhaduri R, et al. A comparison of five epidemiological models for transmission of SARS-CoV-2 in India[J]. BMC infectious diseases, 2021, 21(1): 1-23.
[15] Lucano E, Liberti M, Mendoza G G, et al. Assessing the electromagnetic fields generated by a radiofrequency MRI body coil at 64 MHz: Defeaturing versus accuracy[J]. IEEE Transactions on Biomedical Engineering, 2015, 63(8): 1591-1601.
[16] Conover W J. Practical nonparametric statistics[M]. john wiley & sons, 1999.
[17] Bengtsson B, Heijl A. A visual field index for calculation of glaucoma rate of progression[J]. American journal of ophthalmology, 2008, 145(2): 343-353.
[18] Takahashi K, Yamamoto K, Kuchiba A, et al. Confidence interval for micro-averaged F 1 and macro-averaged F 1 scores[J]. Applied Intelligence, 2022, 52(5): 4961-4972.

Further comments

Further comments from the organizers.

None.