2023 IEEE SPS Video and Image Processing (VIP) Cup

Ophthalmic Biomarker Detection

Overview

Ophthalmic clinical trials that study treatment efficacy of eye diseases are performed with a specific purpose and a set of procedures that are predetermined before trial initiation. Hence, they result in a controlled data collection process with gradual changes in the state of a diseased eye. In general, these data include 1D clinical measurements and 3D optical coherence tomography (OCT) imagery. Physicians interpret structural biomarkers for every patient using the 3D OCT images and clinical measurements to make personalized decisions for every patient.
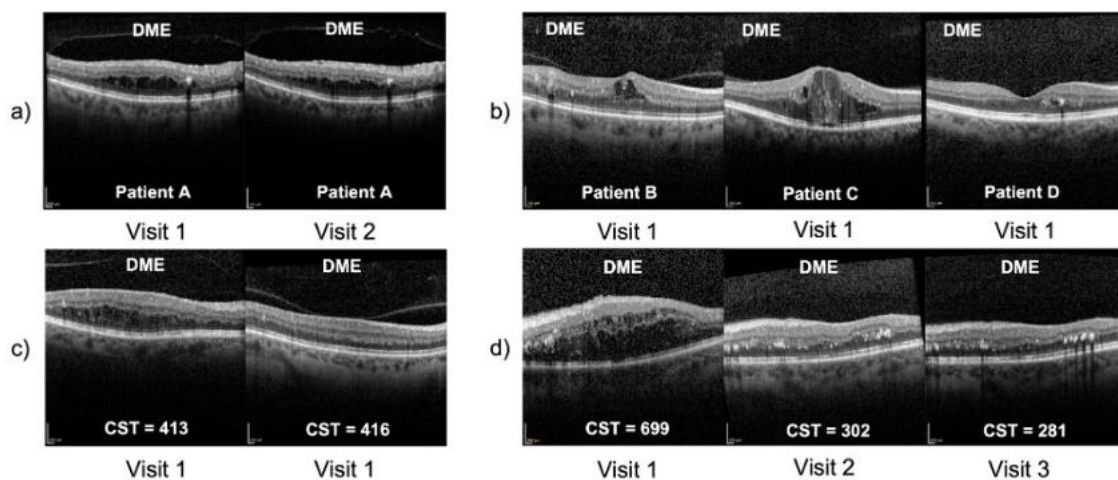

Fig 1: An illustration of personalization challenges within the dataset.

Two main challenges in medical image processing has been _generalization_ and _personalization_.

Generalization aims to develop algorithms that work well across diverse patients and scenarios, providing standardized and widely applicable solutions. Personalization, in

contrast, tailors algorithms to individual patients based on their unique characteristics, optimizing diagnosis and treatment planning. Generalization offers broad applicability but may overlook individual variations. Personalization provides tailored solutions but requires patient-specific data. While deep learning has shown an affinity towards generalization, it is lacking in personalization.

The presence and absence of biomarkers is a personalization challenge rather than a generalization challenge. This is shown in Fig.1. The variation within OCT scans of patients between visits can be minimal while the difference in manifestation of the same disease across patients may be substantial. The domain difference between OCT scans can arise due to pathology manifestation across patients (Fig.1a and Fig.1b), clinical labels (Fig.1c), and the visit along the treatment process when the scan is taken (Fig.1d). Morphological, texture, statistical and fuzzy image processing techniques through adaptive thresholds and preprocessing may prove substantial to overcome these fine-grained challenges. This challenge provides the data and application to address personalization.

Task

PHASE 1:

The overall task is to predict the presence or absence of six different biomarkers simultaneously on every OCT scan in the held-out test set. The training set will consist of the labeled OCT scans for biomarkers and associated clinical information, along with the biomarker labels as ground-truth. Participants are encouraged to make use of all available modalities. Additionally, we will provide a testing dataset that is derived from a recent clinical study in collaboration with RCT. Teams will be provided with the OCT scans and their clinical information **only** for the test set. Participants will use the available data to predict the biomarkers associated with each OCT scan in this test set.

Train and Test Summary

Modality	Per Visit	Per Eye	Train Total	Test Total
OCT	49	Np*49	78189	3871
Clinical	4	Np*4	5082	320
Biomarker	6	Np*49*6	469134	23226

Table 1: Statistics of the available data modalities in the train and test sets, where Np is the total number of visits.

PHASE 2:

In PHASE 1 of the competition, each slice in the test set was treated as its own independent entity. However, every set of 49 slices within the test set was associated with a specific patient's eye. In practice, practitioners may be interested in performance with respect to the patient's eye, rather than performance with respect to isolated slices of the retina. Thus, in PHASE 2 of the competition we want to assess how well the model can *personalize*. To perform the personalization aspect of the competition, the invited **TOP 10 TEAMS** from PHASE 1 will have the opportunity to re-train their models and submit the same biomarker prediction csv files for each image in the test set.

Suggested methods for improving detection:

Potential ideas for improvement include:

1. Using multi-modal integration of OCT scans and clinical labels
   - Clinical labels, such as BCVA and CST, correspond to the health of a patient's eye, and these clinical values can indicate structural changes within the eye. Therefore, a multi-modal approach to integrate clinical labels via a joint loss may improve accuracy. See reference [1]
   - Example Code for [1]
2. Using self-supervised learning with clinical labels
   - Exploring self-supervised techniques may be an effective way to leverage clinical information for biomarker classification. Since clinical labels can exhibit relationships with biomarkers, [2] presents a supervised contrastive loss using clinical labels (BCVA, CST, or a combination of both) for biomarker identification.

- [Example Code](#) for [3]
3. Exploiting information about disease severity label.
   - Exploiting information about OCT disease severity may help with biomarker identification, as patients exhibiting similar disease severity levels are more likely to have similar structural characteristics. [4] quantifies how anomalous OCT images are relative to a healthy distribution, using this approach to choose pairs for a contrastive learning approach for biomarker classification. Other information that can potentially be leveraged is DRSS, which indicates the severity of the disease. Binning images into similar DRSS scores may also be an approach to understand how anomalous OCT images are.

Data

Head over to zenodo to download the dataset: https://zenodo.org/record/8040573

Participation

PHASE 1:

PLEASE REGISTER FOR THE CHALLENGE THROUGH THE [CMS WEBSITE](#) AND BY FILLING UP THIS [FORM](#)

The competition will be hosted on Codalab: **[HERE](#)**

Instructions:

- All teams must register under the same name in the above [form](#) as well as in the participating Codalab challenge.

- Each team needs to have an advising faculty member assigned.

- Once, you have filled out the above form, go to the Codalab competition below and register to participant. The organizers will verify your registration and grant you access to the challenge to start your submission.

PHASE 2:

The **TOP 10 TEAMS** from the PHASE 1 leaderboard will be invited to re-train their models and submit the same biomarker prediction csv files for each image in the test set to assess how well the model is able to *personalize*.

Submission

PHASE 1:

Once participants have trained a model with the provided training data, they are expected to complete the following steps:

- Perform inference for the presence or absence of each biomarker on each individual OCT scan in the test set.
- To submit please fill out the provided template [also available in the Participate tab of Codalab] using the model output for each image in the test set. There should be the file path followed by a one or zero for the presence or absence of each of 6 biomarkers for the associated image.
- **Please zip ONLY your csv file and submit the zip file on codalab for evaluation.**
- SUBMIT THE ZIP FILE **HERE**

Starter Code Usage:

The link to the starter kit is: [HERE]

```
python train.py –batch_size 128 –model 'resnet18' –dataset 'OLIVES' –epochs 1 –device 'cuda:0' –train_image_path " –test_image_path " –
test_csv_path './csv_dir/test_set_submission_template.csv' –train_csv_path './csv_dir/Training_Biomarker_Data.csv'
```

Fill this out with the appropriate file path fields for the training and test data to train a model and produce a numpy that can act as a valid submission once the file paths are appended and saved as a csv.

PHASE 2:

At the end of PHASE 1, the **TOP 10 Teams** who will qualify for PHASE 2 will be required to submit a minimum 1-page description of the team composition (organization, member names), the general approach used in the work, system specifications of the devices used to train and test

the model, and any other detail that relates to the implementation and the deployed algorithms, **along with their CSV submission for PHASE 2**

The submission form for PHASE 2 will be available soon.

Evaluation

To measure the performance of the biomarker detection task, we will make use of the macro averaged F1-score.

Calculation:

$$F1 \ Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

The equation for F1 score

If we express it in terms of True Positive (TP), False Positive (FP), and False Negative (FN), we get this equation:

$$F1 \ Score = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

The alternative equation for F1 score

The macro-averaged F1 score (or macro F1 score) is computed using the arithmetic mean (aka **unweighted** mean) of all the per-class F1 scores.

This method treats all classes equally regardless of their **support** values.

PHASE 1:

There are several characteristics of the dataset that makes this metric desirable. The first is that the test set is imbalanced as it is impossible to guarantee an equal distribution of biomarkers within each OCT scan. Therefore, we need F1-score as it is the harmonic mean of precision and

recall which are both sensitive to dataset imbalances. Furthermore, we want our metric to treat each class of biomarker as equally important, rather than reflecting a bias towards classes with more instances present. In PHASE 1 of the competition, each slice in the test set will be treated as its own independent entity. **Therefore, we will select the TOP 10 TEAMS for PHASE 1 with the highest macro-averaged F1-score across all 6 biomarkers.**

PHASE 2:

We will compute the macro f1-score for each slice in the same way as PHASE 1. However, we will now average the f1-scores with respect to each set of slices associated with an individual patient, rather than averaging across the test set as a whole. This will result in an associated macro f1-score with respect to each individual patient. Teams will be ranked with respect to performance on each of the 40 patients in the test set. Afterwards, the ranking of each team will be averaged across all patients. **The team with the highest average ranking will win the personalization task of the competition and VIP CUP 2023 as a whole.**

Leaderboard

PHASE 1:

The leaderboard for PHASE 1 will be hosted on Codalab HERE

Due to the allowance of multiple submissions for each team, there may be chances of overfitting of the model on the test data. For this reason, the public leaderboard, during the running phase of the competition [July 1st to Aug. 20th, 2023], will display the macro F1 scores for validation on 70% of the test set, and the final decision for the **TOP 10** for PHASE 1 will be made based on the private leaderboard [which will be made public at the end of the challenge] which will rank the teams based on their performance on the other 30% of the set.

PHASE 2:

Evaluation will be conducted on the submissions made by the TOP 10 teams and their reports will be reviewed by September 3, 2023. **The team with the highest average ranking in this phase will win the personalization task of the competition and VIP CUP 2023 as a whole.**

Results for PHASE 2 will be published on our website.

Note: In the event of a tie, the performance in phase 1 will determine the winner.

Important Dates

Phase 1	Date
Registration opens *(open till last day of submission)*	July 1, 2023
Training, Validation & Testing (hidden ground truth) data + Starter code availability	July 1, 2023
Submission & Public Leaderboard opens	July 1, 2023
Phase 1 + Submission ends (Top 10 TEAM announced)	Aug 20, 2023

Phase 2	Date
Phase 2 starts *(Submission form available)*	Aug 21, 2023
Phase 2 submission ends *(both csv and report)*	Sep 1, 2023
Top 3 TEAMS announced	Sep 3, 2023

Prizes

Please check the IEEE VIP Cup Website for prizes: https://signalprocessingsociety.org/community-involvement/video-image-processing-cup.

The **THREE TEAMS** with highest performance in the FINAL competition (PHASE 2) will be invited to present their work in the final phase at ICIP 2023. The champion team will receive a grand prize of $5,000. The first and the second runner-up will receive a prize of $2,500 and $1,500, respectively, in addition to travel grants and complimentary conference registrations.

- Up to three student members from each finalist team will be provided travel support to attend the conference in-person. In-person attendance of the physical conference is required for reimbursement.
- Complimentary conference registration for the three finalist team members from each team who present at ICIP. These complimentary conference registrations cannot be used to cover any papers accepted by the conference. If you are one of the three finalist team

members from each team and wish to receive complimentary registration and/or conference banquet access, you must email Jaqueline Rash, Jaqueline.rash@ieee.org, with this information once your team has been selected as a finalist.

- The three finalist team members from each team will also be invited to join the Conference Banquet and the SPS Student Job Fair, so that they can meet and talk to SPS leaders and global experts. Please note registration to the Conference Banquet and Student Job Fair is limited and based on availability.

Cite

1. M. Prabhushankar, K. Kokilepersaud*, Y. Logan*, S. Trejo Corona*, G. AlRegib, C. Wykoff, "OLIVES Dataset: Ophthalmic Labels for Investigating Visual Eye Semantics," in *Advances in Neural Information Processing Systems* (NeurIPS 2022) Track on Datasets and Benchmarks, New Orleans, LA,, Nov. 29 – Dec. 1 2022 [PDF][Code]
2. K. Kokilepersaud, M. Prabhushankar, and G. AlRegib, "Clinical Contrastive Learning for Biomarker Detection," in *NeurIPS 2022 Workshop: Self-Supervised Learning – Theory and Practice*, Oct. 16 2022, [PDF].
3. K. Kokilepersaud, S. Trejo Corona, M. Prabhushankar, G. AlRegib, C. Wykoff, "Clinically Labeled Contrastive Learning for OCT Biomarker Classification," in *IEEE Journal of Biomedical and Health Informatics, 2023*, May. 15 2023. , [PDF], [Code].
4. K. Kokilepersaud, M. Prabhushankar, G. AlRegib, S. Trejo Corona, C. Wykoff, "Gradient Based Labeling for Biomarker Classification in OCT," in *IEEE International Conference on Image Processing (ICIP)*, Bordeaux, France, Oct. 16-19 2022, [PDF].

Organizers

Ghassan AlRegib (alregib@gatech.edu)is currently the John and Marilu McCarty Chair Professor in the School of Electrical and Computer Engineering at the Georgia Institute of Technology. He was a recipient of the ECE Outstanding Junior Faculty Member Award, in 2008, and the 2017 Denning Faculty Award for Global Engagement. His research group, the Omni Lab for Intelligent Visual Engineering and Science (OLIVES) works on research projects related to machine learning, image and video processing, image and video understanding, seismic interpretation, machine learning for ophthalmology, and video analytics. He has participated in several service activities within the IEEE. He served as the TP co-Chair for ICIP 2020 and GlobalSIP 2014. He is an IEEE Fellow.

Mohit Prabhushankar (mohit.p@gatech.edu) received his Ph.D. degree in electrical engineering from the Georgia Institute of Technology (Georgia Tech), Atlanta, Georgia, 30332, USA, in 2021. He is currently a Postdoctoral Research Fellow in the School of Electrical and Computer Engineering at the Georgia Institute of Technology in the Omni Lab for Intelligent Visual Engineering and Science (OLIVES). He is working in the fields of image processing, machine learning, active learning, healthcare, and robust and explainable AI. He is the recipient of the Best Paper award at ICIP 2019 and Top Viewed Special Session Paper Award at ICIP 2020. He is the recipient of the ECE Outstanding Graduate Teaching Award, the CSIP Research award, and of the Roger P Webb ECE Graduate Research Excellence award, all in 2022.

Kiran Kokilepersaud (kpk6@gatech.edu) is a Ph.D. student in electrical and computer engineering at the Georgia Institute of Technology (Georgia Tech), Atlanta, Georgia, 30332, USA. He is currently a Graduate Research Assistant in the School of Electrical and Computer Engineering at the Georgia Institute of Technology in the Omni Lab for Intelligent Visual Engineering and Science (OLIVES) lab. He is a recipient of the Georgia Tech President's Fellowship for excellence amongst incoming Ph.D. students. His research interests include digital signal and image processing, machine learning, and its associated applications within the medical field.

Prithwijit Chowdhury (pchowdhury6@gatech.edu) received his B.Tech. degree from KIIT University, India in 2020. He joined Georgia Institute of Technology as an MS student in the department of Electrical and Computer Engineering in 2021. He is currently pursuing his Ph.D. degree as a researcher in The Center for Energy and Geo Processing (CeGP) as a member of the Omni Lab for Intelligent Visual Engineering and Science (OLIVES). His interests lie in the areas of digital signal and image processing and machine learning with applications to geophysics.

Zoe Fowler (zfowler3@gatech.edu) is a PhD student in electrical and computer engineering at Georgia Institute of Technology (Georgia Tech), Atlanta, Georgia, 30332, USA. She is currently a researcher in the Omni Lab for Intelligent Visual Engineering and Sciences (OLIVES). She is a recipient of the National Science Foundation Graduate Research Fellowship Program (NSF GRFP) that recognizes graduate students based on their research and academic achievements, as well as the Georgia Tech President's Fellowship. Her interests lie in the areas of digital signal and image processing and machine learning with applications to healthcare.

Extra Reading: Clinical Labels and their Generation

This information is for the medically curious minded and that this is not relevant for performing the task itself.

Full clinical Labels: The clinical labels obtained from the PRIME trials include BCVA, DRSS, CST, eye ID, patient ID, diabetes type, BMI, age, race, gender, HbA1c, leakage index, years with diabetes, and injection arm. The clinical labels from the TREX-DME trials include BCVA, Snellen score, CST, Eye ID, and Patient ID. Since OLIVES is a combination of the two, we use only the common labels from both trials as our clinical labels in our experiments. These common labels include BCVA, CST, Patient ID and Eye ID

The Early Treatment Diabetic Retinopathy Study (ETDRS) diabetic retinopathy severity scale (DRSS) has 13 levels describing DR severity and change over time based on color fundus photograph grading. The scale starts at level 10 and ends at level 90 with irregular scale numbering. Nonproliferative diabetic retinopathy (NPDR) DRSS levels on the scale are below 61 and proliferative diabetic retinopathy (PDR) levels are 61 and above. Diabetes type refers to the patient's diagnosis of either type one or type two diabetes mellitus. HbA1c is the measurement of glycated hemoglobin, commonly referred to as blood sugar, which serves as an indicator for diabetes diagnosis or diabetic control. Leakage index refers to the panretinal leakage index used in the PRIME trial in which areas of leakage, regions of hyperfluorescence in fluorescein angiography images, were divided by areas of

interest, region of total analyzable retinal area, and converted to a percentage. Injection arm refers to either the DRSS-guided cohort or the PLI-guided cohort in the PRIME trial. Snellen score is the visual acuity testing procedure commonly used in ophthalmic clinical settings. The first number indicates the distance in feet that the letter chart was read, in U.S., this number is commonly 20, followed by a number indicating the distance a person with "normal" vision (20/20) would have to be to read something the person tested could read at 20 feet. Thus, a larger denominator would indicate poorer vision.

Other self-explanatory demographic information including body mass index (BMI), age, race, and gender are provided. We caution the users regarding the societal impact of using these labels since the underlying PRIME trial did not study the causality of these labels.

ML Centric Clinical Labels: We describe BCVA and CST in this section. ETDRS best-corrected visual acuity (BCVA) is a visual function assessment performed by certified examiners where a standard vision chart is placed 4-meters away from the patient. The patient is instructed to read the chart from left to right from top to bottom until the subject completes 6 rows of letters or the subject is unable to read any more letters. The examiner marks how many letters were correctly identified by the patient. Central subfield thickness (CST) is the average macular thickness in the central 1-mm radius of the ETDRS grid. CST was obtained from the automated macular topographic information in the Heidelberg Eye Explorer OCT software.

The remaining clinical labels of Patient ID and Eye ID are self-explanatory and collected on clinical visits.


Extra Reading: Structural Biomarkers

This information is for the medically curious minded and that this is not relevant for performing the task itself.

The term "biomarker", a portmanteau of "biological marker", refers to a broad subcategory of medical signs – that is, objective indications of medical state observed from outside the patient – which can be measured accurately and reproducibly. Medical signs stand in contrast to medical symptoms, which are limited to those indications of health or illness perceived by patients themselves.

Biomarkers are measurable indicators of a patient's medical state, helping detect diseases. While they can be surrogate endpoints in clinical trials, caution is advised unless trials are specifically

designed for that purpose. Biomarkers indicate disease presence but are not causal. Medical causality can be singular, involving linked events, or general, examining event relationships.

The six different biomarkers studied in this experiment are Intraretinal Hyperreflective Foci (IRHRF), Partially Attached Vitreous Face (PAVF), Fully Attached Vitreous Face (FAVF), Intraretinal Fluid (IRF), and Diffuse Retinal Thickening or Diabetic Macular Edema (DRT/ME) and Vitreous debris (VD).
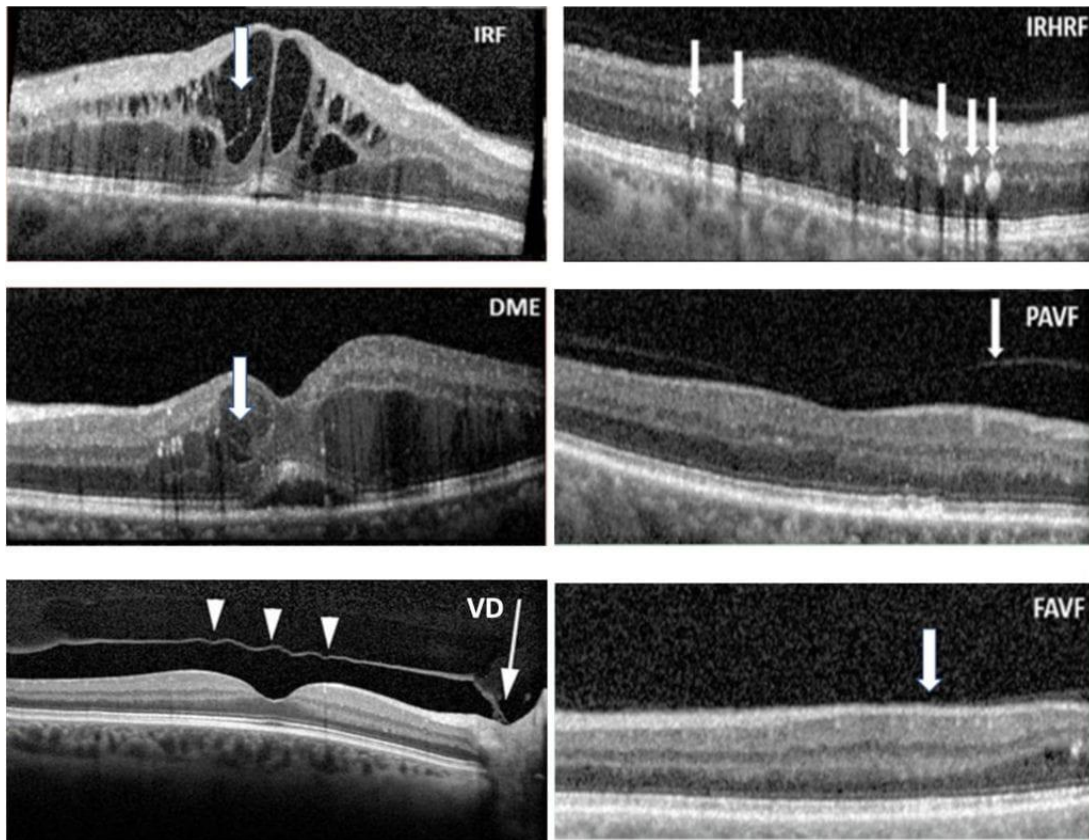
Fig 2: The six different biomarkers studied in this experiment are Intraretinal Hyperreflective Foci (IRHRF), Partially Attached Vitreous Face (PAVF), Fully Attached Vitreous Face (FAVF), Intraretinal Fluid (IRF), and Diffuse Retinal Thickening or Diabetic Macular Edema (DRT/ME) and Vitreous debris (VD).

Biomarker Labelling

Intraretinal Hyperreflective Foci (IRHRF) were indicated as present with the appearance of intraretinal, highly reflective spots, which correspond pathologically to microaneurysms or hard exudates, with or without shadowing of the more posterior retinal layers.

A **Partially Attached Vitreous Face (PAVF)** was indicated as present with evidence of perifoveal detachment of the vitreous from the internal limiting membrane (ILM) with a macular attachment point within a 3-mm radius of the fovea.

A **Fully Attached Vitreous Face (FAVF)** was indicated as present with no evidence of perifoveal or macular detachment from the ILM.

Intraretinal Fluid (IRF) was indicated as present when intraretinal hyporeflective areas or cysts had a minimum fluid height of 20 µm

Diffuse Retinal Thickening or Diabetic Macular Edema (DRT/ME) was indicated as present when there was increased retinal thickness of 50 µm above the otherwise flat retina surface with associated reduced reflectivity in the intraretinal tissues

Vitreous Debris (VD) was indicated as present with evidence of hyperreflective foci in the vitreous or shadowing of the retinal layers in the absence of an intraretinal hemorrhage.

Terms & Conditions

- Please check the team formation and eligibility rules at: https://signalprocessingsociety.org/community-involvement/video-image-processing-cup
- Specifically "Each team participating should be composed of one faculty member or someone with a PhD degree employed by the university (the Supervisor), at most one graduate student (the Tutor), and at least three, but no more than ten undergraduate students. At least three of the undergraduate team members must hold either regular or student memberships of the IEEE Signal Processing Society. Undergraduate students who are in the first two years of their college studies, as well as high school students who are capable to contribute are welcome to participate in a team. A participant cannot be on more than one team."
- Participants must register with a contact email, team name, and e-signing a document that acknowledges an honest result reporting policy to receive the held out clinical trial test set.
- Previously published methods that are adapted to this challenge must be properly cited by the participants. Contributions of the participants on top of these existing methods must be properly documented and highlighted.