|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ITU Logo | INTERNATIONAL TELECOMMUNICATION UNION  **TELECOMMUNICATION STANDARDIZATION SECTOR**  STUDY PERIOD 2017-2020 | | FGAI4H-E-005-A07 | |
| **ITU-T Focus Group on AI for Health** | |
| **Original: English** | |
| **WG(s):** | | N/A | Geneva, 30 May – 1 June 2019 | |
| **DOCUMENT** | | | | |
| **Source:** | | TG-Ophtalmo Driver | | |
| **Title:** | | Call for Topic Group Participation: Standardized benchmarking of AI for Ophthalmology (Retinal Imaging Diagnostics) | | |
| **Purpose:** | | Engagement | | |
| **Contact:** | | Topic driver: Arun Shroff | | Email: [arunshroff@gmail.com](mailto:arunshroff@gmail.com) |

|  |  |
| --- | --- |
| **Abstract:** | Calling on members of the medical and artificial intelligence communities with a vested interest in AI for Ophthalmology (retinal imaging diagnostics)! Become engaged in the group dedicated to establishing a standardized benchmarking platform for AI for Ophthalmology (retinal imaging diagnostics) within the International Telecommunication Union (ITU)/World Health Organization (WHO) Focus Group on “Artificial Intelligence for Health” (FG-AI4H). |

**Call for Topic Group Participation: AI for** **Ophthalmology (retinal imaging diagnostics)**

The International Telecommunication Union (ITU)/World Health Organization (WHO) Focus Group on “Artificial Intelligence for Health” (FG-AI4H; <https://www.itu.int/go/fgai4h>) seeks engagement from members of the medical and artificial intelligence (AI) communities (including clinicians, technologists, entrepreneurs, potential benchmarking data providers, machine learning experts, software developers, researchers, regulators, policy-makers, companies/institutions, and field experts) with a vested interest in shaping the benchmarking process of AI for Ophthalmology (retinal imaging diagnostics)

# About FG-AI4H

Over the past decade, considerable resources have been allocated to exploring the use of AI for health, which has revealed an immense potential. Yet, due to the complexity of AI models, it is difficult to understand their strengths, weaknesses, and limitations. If the technology is poorly designed or the underlying training data are biased or incomplete, errors or problematic results can occur. AI technology can only be used with complete confidence if it has been quality controlled through a rigorous evaluation in a standardized way. Towards developing this standard assessment framework of AI for health, the ITU has established FG-AI4H in partnership with the WHO.

Thus far, FG-AI4H has established thirteen topic groups. These are concerned with: AI and cardiovascular disease risk prediction, child growth monitoring, dermatology, falls among the elderly, histopathology, neuro-cognitive disorders, ophthalmology (retinal imaging diagnostics), psychiatry, radiotherapy, snakebite and snake identification, symptom checkers, tuberculosis, and volumetric chest computed tomography.

Each topic group agrees upon representative benchmarking tasks in a pragmatic, best-practice approach, which can later be scaled and expanded to similar tasks. Every benchmarking task should address a health problem of relevance (e.g. impacting a large and diverse part of the global population or challenging to treat) and for which AI technology would provide a tangible improvement relative to the current practice (e.g. better care, results, and/or cost/time effectiveness).

For a rigorous and sound evaluation, undisclosed test data sets must be available (or have to be collected) for each task. All data must be of high quality and compliant with ethical and legal standards. In addition, the data must originate from a variety of sources so that it can be determined whether an AI algorithm can generalize across different conditions, locations, or settings (e.g. across different people, hospitals, and/or measurement devices). The format/properties of the data serving as input to the AI and of the output expected from the AI, as well as the benchmarking metrics are agreed upon and specified by the topic group.

Finally, the AI-to-be-evaluated will be benchmarked with the undisclosed test data on FG-AI4H computing infrastructure. Here, the AI will process single samples of the undisclosed test data set and predict output variables, which will be compared with the "ground truth." The results of the benchmarking will be provided to the AI developers and will appear on a (potentially anonymized) leaderboard.

# Topic group: AI for Ophthalmology (retinal imaging diagnostics)

A topic group is a community of stakeholders from the medical and AI communities with a shared interest in a topic. The objectives of the topic groups are manifold:

1. to provide a forum for open communication among various stakeholders,
2. to agree upon the benchmarking tasks of this topic and scoring metrics,
3. to facilitate the collection of high quality labeled test data from different sources,
4. to clarify the input and output format of the test data,
5. to define and set-up the technical benchmarking infrastructure, and
6. to coordinate the benchmarking process in collaboration with the Focus Group management and working groups.

The primary output of a topic group is one document that describes all aspects of how to perform the benchmarking for this topic. (The document will be developed in a cooperative way by suggesting changes as input documents for the next FG-AI4H meeting that will then be discussed and integrated into an official output document of this meeting. The process will continue over several meetings until the topic description document is ready for performing the first benchmarking.)

This topic group is devoted to using artificial intelligence for the detection and diagnostics of ophthalmological diseases and conditions, in particular Diabetic Retinopathy (DR), from retinal images. Other retinal diagnostics for Glaucoma and Age-related Macular Degeneration (AMD) are also within the scope of the AI for Ophthalmology topic group.

DR is a serious eye-disease caused by diabetes that can lead to vision impairment and blindness if it is not detected and treated early enough. The WHO estimates that there are over 422 million people with diabetes worldwide. Of these 35% or over 148 million are estimated to have DR with potential for vision impairment and 11% or 48 million are estimated to have Vision Threating DR that can lead to blindness. Prevention of vision loss requires early detection of DR via regular eye exams and screening by a trained ophthalmologist or eye care professional. However, given the large numbers of people affected worldwide, there are not sufficient specialists globally to screen everyone at risk. The shortfall is particularly acute in developing countries, including India, and many countries in Asia and Africa. In addition, many affected people live in remote areas with little or no access to clinics and screening facilities. This makes Diabetic Retinopathy a global healthcare challenge that needs urgent resolution.

Recent advances in Artificial Intelligence algorithms for image recognition and medical diagnostics have been shown to be effective in detecting Diabetic Retinopathy at an accuracy comparable to an ophthalmologist. A standard way of benchmarking the performance of various AI applications to detect and diagnose Diabetic Retinopathy is therefore important. It will help in the evaluation and selection of appropriate solutions to address this global healthcare challenge.

**Benchmarking:**

The benchmarking of the algorithms for detecting DR would be done on a sufficiently large undisclosed test data set - representative of number of categories that the algorithm seeks to classify.

DR images are usually classified into five DR stages: Normal (no presence of DR), Mild Non-proliferative DR (NPDR), Moderate NPDR, Severe NPDR, Proliferative DR (PDR). In the binary DR classification, moderate, severe, and proliferative DR are considered as referable DR, while none and mild are deemed non-referable. If the image is not of sufficient quality to be graded for severity it is classified as Non-gradable – a non-diagnostic category.

The following benchmarks are proposed for algorithms for detecting DR:

1. An algorithm may classify an image as: Non-gradable, Non-referable DR, Referable DR
2. An algorithm may classify an image as: Non-gradable, Normal, Mild NPDR, Moderate NPDR, Severe NPDR, PDR

In all cases, the following key metrics will be used for benchmarking:

* Accuracy
* Area under the Receiver Operating Characteristics (ROC) curve or AUC,
* Sensitivity and Specificity
* For the multi-class cases we will also use a Confusion Matrix for comparing performance.

**Available Datasets:**

Publicly available datasets include the EyePACS dataset (around 90,000 fundus images, 5 levels of severity), MESSIDOR dataset (1,200 images, 4 levels of severity), the DIARETDB dataset (around 200 images marked with lesions). Currently, there are no private datasets available for testing.

**Topic Group Activity:**

More details about the activities of the topic group can be found in the documents:

* [FGAI4H-C-026-R1.docx](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/docs/FGAI4H-C-026-R1.docx) (Medindia.net)
* [FGAI4H-C-022.docx](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/docs/FGAI4H-C-022.docx) (Baidu)

These can be accessed with a free ITU account (cf. “Get involved”).

**Topic Group Members:**

Current members of the topic group on Ophthalmology (retinal imaging diagnostics) include

* Arun Shroff, Xtend.AI and Medindia.net (USA, India),

Founder of Xtend.ai and Topic Lead for Ophthalmology (retinal imaging diagnostics)

* Yanwu XU, Artificial Intelligence Innovation Business, Chief Scientist, Baidu, China
* Jingyu WANG, Artificial Intelligence Group, Baidu, China
* Shan Xu, CAICT, China

The topic group would benefit from further expertise of the medical and AI communities and from additional data.

The requirement for this topic group is – to be(come) an active member of the FGAI4H group and have a background, interest or expertise in this topic - Ophthalmology (retinal image diagnostics) either as a healthcare professional or an AI practitioner with a model or algorithm for DR or in some other capacity.

# Get involved

To join this topic group, please send an e-mail to the focus group secretariat ([tsbfgai4h@itu.int](mailto:tsbfgai4h@itu.int)) and the topic driver ([arunshroff@gmail.com](mailto:arunshroff@gmail.com)). Please use a descriptive e-mail subject (e.g. "Participation topic group AI for Ophthalmology"), briefly introduce yourself and your organization, concisely describe your relevant experience and expertise, and explain your interest in the topic group.

Participation in FG-AI4H is free of charge and open to all. To attend the workshops and meetings, please visit the Focus Group website (<https://www.itu.int/go/fgai4h>), where you can also find the whitepaper, get access to the documentation, and sign up to the mailing list.