|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ITU Logo | INTERNATIONAL TELECOMMUNICATION UNION  **TELECOMMUNICATION STANDARDIZATION SECTOR**  STUDY PERIOD 2017-2020 | | FG-AI4H-H-007-A02 | |
| **ITU-T Focus Group on AI for Health** | |
| **Original: English** | |
| **WG(s):** | | Plenary | Brasilia, 22-24 January 2020 | |
| **DOCUMENT** | | | | |
| **Source:** | | TG-Derma Topic Driver | | |
| **Title:** | | Att.2 – CfTGP (TG-Derma) [same as Meeting E] | | |
| **Purpose:** | | Engagement | | |
| **Contact:** | | Topic driver: Maria Vasconcelos | | Email: [maria.vasconcelos@fraunhofer.pt](mailto:maria.vasconcelos@fraunhofer.pt) |

|  |  |
| --- | --- |
| **Abstract:** | Calling on members of the medical and artificial intelligence communities with a vested interest in AI for dermatology! Become engaged in the group dedicated to establishing a standardized benchmarking platform for AI for dermatology within the International Telecommunication Union (ITU)/World Health Organization (WHO) Focus Group on “Artificial Intelligence for Health” (FG-AI4H). This document is the same as seen in meeting E, reproduced for meeting H for easier reference. |

**Call for Topic Group Participation: AI for Dermatology**

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 dermatology.

# 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 diseases, 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 dermatology

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 labelled 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 dedicated to AI for dermatology. One in every three cancers diagnosed is a skin cancer, and every year approximately 3 million new cases of skin cancer is detected worldwide, more than breast cancer, prostate cancer, lung cancer and colon cancer combined [1]. According to Skin Cancer Foundation Statistics, one in every five Americans will develop skin cancer in their lifetime. Although Malignant Melanoma (MM) accounts only for a small percentage of this type of cancer, it is responsible for the most skin cancer related deaths. Early diagnosis of MM is, therefore, extremely important considering the high success rates of recovery if the malignancy is detected during the early stages of its development. Therefore, awareness activities and screening procedures are of high importance.

Several recent studies have been focused in evaluating the usage of mobile teledermatology for skin cancer screening. High management concordance of 81% between in-person and teledermatology evaluations have been previously reported [2], with similar results being obtained by a recent study in a low prevalence population (78% concordance) [3]. This last study also explored the impact of adding a dermoscopic image to conventional images, being reported only a slight impact in terms of concordance (85% vs 78%), but still important when the goal is to maximize the number of correctly ruled-out cases for further follow-up (82% vs 74%). Still in terms of concordance, these studies are in line with previous research demonstrating high cancer detection rates of mobile teledermatology in high-prevalence settings [4, 5].

Significant advances in the automatic risk assessment of skin lesions through computer vision have been recently reported. However, most of these works were made on an academic level and mainly focused on specific parts of the problem. There is, therefore, a shortage of systems that combines this knowledge, in order to create an integrated tool with effective practical utility.

Fraunhofer AICOS has been performing research in the field of Mobile Dermatology since 2011, more details about our research can be obtained in the website. So far, our work has been focused in building strong scientific and technical competences in the area of computer vision and machine learning that can be directly applied to dermatology, specifically in the area of automated risk assessment of melanoma. Fraunhofer AICOS has a public dataset ([SMARTSKINS](http://smartskins.projects.fraunhofer.pt/dataset.html) dataset) that contains images of 106 melanocytic lesions acquired with a smartphone with and without an adaptable dermoscope. The dataset includes medical annotation of all smartphone acquired images without using a dermoscope, namely the assessment of Asymmetry, Border and Color criteria score according to the ABCD rule, as well as the Overall Risk of the skin lesion (benign, moderate atypia or high demarcated atypia), with no histology result associated.

Moreover, we are currently developing an AI module for automatic risk categorization of dermatological referral requests to be later tested in the Portuguese National Health Service, through “derm.AI - Usage of Artificial Intelligence to power Teledermatological Screening”, a joint project with SPMS (Shared Services for Ministry of Health).

For the benchmarking task, the participants should be able to submit an AI model capable of classifying macroscopic images of skin lesions obtained via smartphones in terms of level of risk (benign lesion vs moderate/high atypia). As possible metrics we are currently considering the Sensitivity, Specificity and F1-score applied to a binary classification problem (benign lesion vs moderate/high atypia).

Regarding data availability, there are 106 melanocytic lesions annotated. There are other open datasets available such as The Interactive Atlas of Dermoscopy ([EDRA](http://derm.cs.sfu.ca/Welcome.html)), [ISIC Archive](https://www.isic-archive.com/) and [Dermofit](https://licensing.edinburgh-innovations.ed.ac.uk/i/software/dermofit-image-library.html).

[1] “World health organization.” URL: http://www.who.int/, Apr. 2018

[2] Lamel, S. A., Haldeman, K. M., Ely, H., Kovarik, C. L., Pak, H., & Armstrong, A. W. (2012). Application of mobile teledermatology for skin cancer screening. Journal of the American Academy of Dermatology, 67(4), 576-581.

[3] Markun, S., Scherz, N., Rosemann, T., Tandjung, R., & Braun, R. P. (2017). Mobile teledermatology for skin cancer screening: A diagnostic accuracy study. Medicine, 96(10).

[4] Kroemer, S., Frühauf, J., Campbell, T. M., Massone, C., Schwantzer, G., Soyer, H. P., & Hofmann‐Wellenhof, R. (2011). Mobile teledermatology for skin tumour screening: diagnostic accuracy of clinical and dermoscopic image tele‐evaluation using cellular phones. British Journal of Dermatology, 164(5), 973-979.

[5] Börve, A., Terstappen, K., Sandberg, C., & Paoli, J. (2013). Mobile teledermoscopy—there’s an app for that! Dermatology practical & conceptual, 3(2), 41.

More details about the activities of the topic group can be found in the document [FG-AI4H-C-015](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7b3D1B3D44-B3E7-4CCB-B104-29E2C087003C%7d&file=FGAI4H-C-015.docx&action=default), which can be accessed with a free ITU account (cf. “Get involved”).

Current members of the topic group on AI for dermatology include Maria Vasconcelos, PhD in Informatics Engineer, Senior Researcher at Fraunhofer AICOS. The topic group would benefit from further expertise of the medical and AI communities and from additional data.

# 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 ([maria.vasconcelos@fraunhofer.pt](mailto:maria.vasconcelos@fraunhofer.pt)). Please use a descriptive e-mail subject (e.g. "Participation topic group AI for dermatology"), 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.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_