INTERNATIONAL TELECOMMUNICATION UNION

TELECOMMUNICATION STANDARDIZATION SECTOR

STUDY PERIOD 2022-2024

FGAI4H-P-007-A01

ITU-T Focus Group on AI for Health

Original: English

WG(s): Plen Helsinki, 20-22 September 2022

DOCUMENT

Source: TG-Derma Topic Driver

Title: Att.1 - TDD update (TG-Derma) [same as Meeting E]

Purpose: Discussion

Contact: Harsha Jayakody E-mail: harsha@flash.health
Flash Health
Sri Lanka

Abstract: This project aims to improve the existing Teledermatology screening processes through a mobile based solution and the usage of Artificial Intelligence. This document provides details on the training data and test data, as well as its format and annotations.

This version of the TDD is the same as seen in Meeting E (FGAI4H-E-011), reproduced for easier reference as a Meeting N document. The Topic Driver Maria Vasconcelos (Fraunhofer Portugal) replaced by Weihong Huang (Xiangya Hospital Central South University, China) after Meeting J, then replaced by Mr Sharad Kumar (Nurithm Labs, India) at Meeting M and he was replaced by Harsha Jayakody (Flash Health, Sri Lanka) at meeting P.

Contents`

1 Introduction

Former documents: **B-025: Proposal: Teledermatological Screening Solution via Mobile Devices (Fraunhofer Portugal)**

This submission was provided in response to the ITU-T FG-AI4H's call for proposals on use cases and data A-102. The document was presented remotely. The project aims to improve the existing teledermatology screening processes through a mobile based solution and the usage of Artificial Intelligence.

Input: image on a phone; output: diagnostics. Public and private datasets.

After presentation, the authors provided the answers to the questionnaire in B-006, as found in B-025-R1.

1.1 Topic Description

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].

1.2 Ethical Considerations

TBC

1.3 Existing AI Solutions

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.

TBC

1.4 Existing work on benchmarking

TBC

2 AI4H Topic group

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.)

3 Method

3.1 AI Inputs Data Structure

- Image File Format: jpg format;
- Image File Names: anonymized;
- Macroscopic (without magnification);
- Image size: TBD (less then 1MB).

3.1.1 Public datasets

(1) The Interactive Atlas of Dermoscopy (EDRA)[1] - A multimedia guide (Booklet + CD-ROM) intended for training medical personnel to diagnose skin lesions. It has 1000+ clinical cases, each with at least two images of the lesion: close-up clinical image and dermoscopic image. Most images are 768 pixels wide x 512 high. Each case has clinical data, histopathological results, diagnosis, and level of difficulty.

(2) ISIC Archive[2] - The International Skin Imaging Collaboration (ISIC) is an international effort to improve melanoma diagnosis, sponsored by the International Society for Digital Imaging of the Skin (ISDIS). The ISIC Archive contains the largest publicly available collection of quality controlled dermoscopic images of skin lesions.

(3) Dermofit[3] - The Dermofit Image Library is a collection of 1,300 skin lesion images and their segmentation masks divided among 10 classes. The diagnoses were provided by expert dermatologists and dermatopathologists, generating a gold standard groundtruth. Although this dataset is not publicly available, it can be purchased.

It should be noted that the previously referred datasets have more dermoscopic images than clinical and macroscopic images.

[1] http://derm.cs.sfu.ca/Welcome.html
[2] https://www.isic-archive.com/
[3] https://licensing.edinburgh-innovations.ed.ac.uk/i/software/dermofit-image-library.html

3.1.2 Dataset public (training dataset) from AICOS

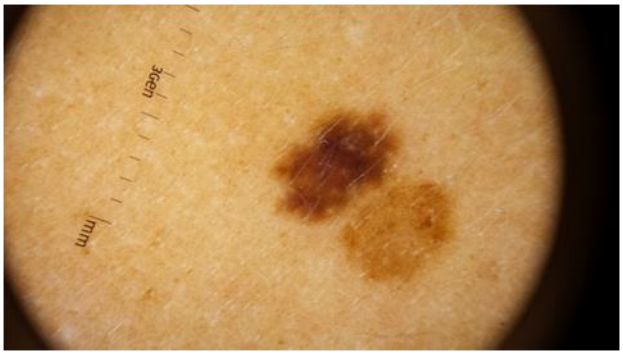
Fraunhofer AICOS has a public dataset (SMARTSKINS dataset[4]) that contains images of 106 melanocytic lesions acquired with a smartphone with and without an adaptable dermoscope.

The dataset consists of 179 macroscopic images of 106 melanocytic lesions acquired with a smartphone and its corresponding 106 dermoscopic images. Two smartphones were used to acquire the images: one HTC One and Samsung S4, and DermLite DL3 dermoscope was used coupled in the Samsung S4.

The dataset includes the annotation of one dermatology specialist of all melanocytic lesions acquired with a smartphone 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 (1- benign – low risk, 2- moderate atypia- medium risk or 3- high demarcated atypia – high risk), there is no histology result associated with the medical annotation.

The SMARTSKINS dataset has been developed for research and benchmarking purposes, in order to facilitate comparative studies of dermatological images. The dataset was acquired at the Skin Clinic of Portuguese Institute of Oncology of Porto (IPO Porto) after the approval of the IPO Ethics Committee. Informed consent was obtained from all the participants.

Example of the images, annotations and distribution of the public dataset:



[4] http://smartskins.projects.fraunhofer.pt/dataset.html

Image	HTC	S4	S4	Asymmetry		Border	Color						Overall
ID	Mobile	Mobile	Dermoscopic	Major	Minor		White	Red	Light Brown	Dark Brown	Black	Blue Grey	Risk
17	1		1	1	1	0	0	0	1	1	1	0	3
103	1	1	1	0	1	0	0	0	1	1	0	1	3

Summary	
Number of Moles	106
HTC images	103
S4 images	81
Dermoscopic S4	106
From all three devices	78

Scoring description	
Asymmetry	
0	symmetric axis
1	asymmetric axis
Border	
0 – 8	number of abrupt borders
Color	
0	color absent
1	color present
Overall risk	
1	low risk
2	medium risk
3	high risk

Doctor									
classification	0	1	2	3	4	5	6	7	8
A score	7	44	55						
B score	90	13	2	1	0	0	0	0	0
C score		59	41	6	0	0			
Risk 1-3		47	43	16					
Risk 0-1	47	59							

3.2 AI Output Data Structure

TBC

The output of the algorithms should be a CSV file in text format with the following columns:

- Name of the file processed (XXX.jpg);
- Possible classifications:

- 3 levels:
  - 1- benign – low risk;
  - 2 - moderate atypia- medium risk;
  - 3- high demarcated atypia – high risk.
- Binary classification
  - 0- benign – low risk;
  - 1- moderate or high demarcated atypia – medium or high risk.

3.3 Test Data Labels

3.3.1 Test data (private)

A set of up to 80 macroscopic images of melanocytic lesions (different from the previous dataset) obtained with smartphone can be used to classify the images in terms of level of risk (benign lesion vs moderate/high atypia). This dataset was annotated in the same conditions of the SMARTSKINS dataset).

Note: Within the next year Fraunhofer AICOS expects to extend the dataset with more 1500 macroscopic images of skin lesions and have the annotation of 3/5 dermatology specialists in terms of Overall Risk of the skin lesion. The dataset will also include dermoscopic images and some anonymized clinical information. The inclusion of this new dataset and its conditions can only be inferred in the future.

3.4 Score and Metrics

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:

- Sensitivity;
- Specificity;
- F1-score

applied to a binary classification problem (benign lesion vs moderate/high atypia).

3.5 Undisclosed Test Data Set Collection

TBC

3.6 Benchmarking Methodology and Architecture

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).

TBC

As next steps:

- disseminate the group
- schedule meeting with Weihong Huang from Xiangya Hosp. Central S. University.

4 Reporting Methodology

– Report publication in papers or as part of ITU documents

– Online reporting

– Report structure including an example

5 Results

TBC

6 Discussion

TBC

7 Declaration of Conflict of Interest

In accordance with the ITU rules in this section working on this document should define his conflicts of interest that could potentially bias his point of view and the work on this document.

TBC

References

[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.