|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ITU Logo | INTERNATIONAL TELECOMMUNICATION UNION  **TELECOMMUNICATION STANDARDIZATION SECTOR**  STUDY PERIOD 2017-2020 | | FG-AI4H-M-051 | |
| **ITU-T Focus Group on AI for Health** | |
| **Original: English** | |
| **WG(s):** | | Plenary | E-meeting, 28-30 September 2021 | |
| **DOCUMENT** | | | | |
| **Source:** | | Nurithm Labs Private Limited (India) | | |
| **Title:** | | TG-Derma: Clinical validation of a machine learning driven mobile phone application for diagnosis of forty common dermatological conditions | | |
| **Purpose:** | | Discussion | | |
| **Contact:** | | Sharad Kumar Nurithm Labs Private Limited India | | Tel: +91.9871876653 E-mail: sharad.kumar@nurithmlabs.tech |

|  |  |
| --- | --- |
| **Abstract:** | The author will present results from a clinical study conducted for a machine learning driven mobile phone app, Derma AidTM, for diagnosis of 40 skin diseases in India. A total of 5014 patients were evaluated by the m-health app and the results compared against dermatologist diagnosis. The app demonstrated an overall top-1 accuracy of 76.93 +/- 0.88% and a mean area-under-curve of 0.95 +/- 0.02. The study underscores the utility of the AI-driven smartphone applications as a point-of-care, clinical decision support tool for dermatological diagnosis of skin diseases in patients of colour. The results presented here are from the full paper with DOI: 10.1111/jdv.16867. Please refer to this for the full details. The next steps is where the author has started to identify follow-up items that can be part of the Derma TG. |

Introduction & Background

The study aimed to develop a convolutional neural network (CNN)-based algorithm trained with clinical images of 40 different skin diseases. A user-friendly, smartphone app was also generated, and a clinical validation study on 5014 patients was done by physicians in urban, rural primary care and tertiary care set- tings. The app’s analysis of a single image of the lesion was compared to the consensus diagnosis made by two board-certified dermatologists. Only patients, for whom the treating board-cer tified dermatologist was confident of the diagnosis and the diagnosis was cross-verified by another board-certified dermatologist, were included. If there was no consensus then the patient was excluded from the study.

Methodology & Results

Model performance was measured in terms of per disease sensitivity, specificity, positive predictive values (PPV) and negative predictive values (NPV), area-under-the-curve (AUC) and mean AUC. Mean values with standard deviation were shown for disease-specific sensitivities, specificities, PPV, NPV and overall accuracy of model and app with 95% confidence intervals (CI) were calculated. Receiver operating characteristics (ROC) curves were plotted using probability scores for each of the disease classes by varying the cutoff threshold. The threshold varied from 0 to 1 in equally spaced intervals and the sensitivity and specificity for each disease were measured as they changed. We used this data to calculate the true positive rate (same as sensitivity) and the false-positive rate (1 – specificity).

Diagram

Description automatically generated

Figure 1: This ﬂowchart depicts the different steps in data collection, algorithm generation, algorithm testing, app generation and app validation in clinical studies.

Graphical user interface, application, table, Excel

Description automatically generated

Figure 2: This shows the number of patients, AUC, Top‐1/Top‐3 sensitivity, speciﬁcity, positive predictive and negative predictive values for combined clinical data from three different clinical settings

Next Steps

The above study followed a typical flow for clinical validation of an AI-driven decision support application. However, there are several things that need more discussion, robustness for adoption in real clinical settings and provide positive patient outcomes. Some of the things we identified:

1. Images have a significant impact on the quality of results. The image zoom level, resolution, lighting, distractions such as clothing all impact the quality of results.
2. Training of Physicians also has a significant impact. Just being able to capture the images is not sufficient. A physician or a patient (in a telemedicine setting) has to understand or be guided about how to capture the disease lesion information.
3. There needs to be mechanism to capture patient symptomatic, history and other aspects that are typically captured when a patient sees a dermatologist in person.

Here are some of the Needs we identified. There are probably others that this WG should brainstorm on.

1. Support API and algorithms for Image quality checks for appropriateness for diagnosis
2. Support API and algorithms for additional patient clinical symptomatic information and patient history to supplement diagnosis. This should include meta-data for patient history and optionally support voice data.
3. Support algorithms for disease progression analysis
4. Support for prescription, treatment, history.

Several of these aspects can also be generalised to how a one can standardise and drive interoperability for images and patient history captured on mobile device.

**References**

1. A machine learning-based, decision support, mobile phone application for diagnosis of common dermatological diseases, DOI 10.1111/jdv.16867

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