|  |  |  |
| --- | --- | --- |
| ITU Logo | INTERNATIONAL TELECOMMUNICATION UNION**TELECOMMUNICATIONSTANDARDIZATION SECTOR**STUDY PERIOD 2017-2020 | FG-AI4H-K-033 |
| **ITU-T Focus Group on AI for Health** |
| **Original: English** |
| **WG(s):** | TG-Derma | E-meeting, 27-29 January 2021 |
| **DOCUMENT** |
| **Source:** | Dermatology AI for Global Health, USA |
| **Title:** | TG-Derma: Dermatology AI for global health (DAIGH) proposal for image recognition challenge for skin diseases  |
| **Purpose:** | Discussion |
| **Contact:** | Mr. Lincoln ManziDermatology AI for Global Health, USA | Tel: +1.202.617.0686Email: dermatology.global.health@gmail.com  |

|  |  |
| --- | --- |
| **Abstract:** | Skin disease is known to have a significant impact on quality of life, productivity and mental health, especially in the developing world. Despite this, skin diseases in developing countries often remain undiagnosed. The application of convolutional neural networks (CNNs) to the diagnosis of skin disease is an active field of research but has been predominantly focused on developed-world diseases such as melanoma. We assembled a database of more than twenty thousand open-source photos of human skin affected with a wide variety of conditions and have built a CNN-based model for image-based diagnoses for skin diseases prevalent in the developing world. We propose to execute an image recognition challenge for skin diseases in which we would make the dataset and baseline model available to multiple participants, allow each participant to build alternative CNN architectures, evaluate the performance of each model, and publish the highest-performing model on an open-source basis. |

# Overview

Skin disease is known to have a significant impact on quality of life, productivity and mental health, especially in the developing world. Despite this, skin diseases in developing countries often remain undiagnosed. The application of convolutional neural networks (CNNs) to the diagnosis of skin disease is an active field of research but has been predominantly focused on developed-world diseases such as melanoma. We assembled a database of more than twenty thousand open-source photos of human skin affected with a wide variety of conditions and have built a CNN-based model for image-based diagnoses for skin diseases prevalent in the developing world. We propose to execute an image recognition challenge for skin diseases in which we would make the dataset and baseline model available to multiple participants, allow each participant to build alternative CNN architectures, evaluate the performance of each model, and publish the highest-performing model on an open-source basis.

# Relevance

Skin disease is known to have a significant impact on quality of life, productivity and mental health. In fact, the Global Burden of Disease (GBD) Study 2013, found skin disease to be the fourth (non-fatal) cause of disability worldwide. The impact of any skin disease is typically larger in lower socioeconomic populations with limited access to health care and an agricultural economy. The World Health Organization's 2001 report on the global burden of disease indicated that skin diseases were associated with mortality rates of 20,000 in Sub-Saharan Africa in 2001. This burden was comparable to mortality rates attributed to meningitis, hepatitis B, obstructed labor, and rheumatic heart disease in the same region. Using a comparative assessment of disability-adjusted life years (DALYs) from the same report, the World Health Organization estimated a total of 896,000 DALYs for the region in the same year, similar to that attributed to gout, endocrine disease, panic disorders, and war-related injuries.

# Impact

Early detection of many skin diseases is essential to achieving the best possible clinical outcomes. Despite this, skin diseases in developing countries often remain undiagnosed because limited healthcare resources are devoted to problems that are more life-threatening. There is increasing attention being devoted to smartphone applications that can allow self-diagnosis of skin conditions, but none are yet available that can reliably diagnose a wide range of skin diseases and these applications are also expensive by developing world standards. Development and deployment of smartphone-based general-purpose skin disease diagnostic tools focused on skin conditions prevalent in the developing world can lead to earlier diagnosis, and therefore more effective treatment, for skin diseases.

# Existing Work

The application of convolutional neural networks to medical diagnosis, including diagnosis of skin disease, is an active field of research. Recent academic publications that describe approaches to applying neural networks to skin disease identification using digital images include “A Web-Based Skin Disease Diagnosis Using Convolutional Neural Networks” (I.J. Information Technology and Computer Science, 2019, 11, 54-60) and “A Method Of Skin Disease Detection Using Image Processing And Machine Learning” (Procedia Computer Science. 163. 85-92). In addition, there are several commercially available smartphone

applications focused on identification of melanoma using images of skin obtained via embedded smartphone cameras, including UMSkinCheck, MoleMapper, Miiskin, MoleScope and SkinVision.

# Feasibility

Estimation of probability of melanoma using smartphone photos has been demonstrated repeatedly in the field. The work executed to date by Dermatology AI for Global Health (DAIGH) applies these concepts to skin diseases prevalent in the developing world. DAIGH has assembled a database of more than twenty thousand open-source photos of human skin affected with a wide variety of conditions. We have further built and implemented a CNN that can detect the presence of any of four conditions (scabies, ringworm, eczema or acne) with approximately 80 percent accuracy using only a single photo of affected area of skin. We hold a pending U.S. patent for this system (U.S. Patent 63/034564: "Image-Based Skin Disease Identification") that we will make available without cost to the project.

The project would extend this work in two dimensions. First, it would test a variety of alternative CNN architectures to improve accuracy on these disease classes. Second, it would extend the diagnostic models to a wider variety of prevalent skin conditions. Prior work in a variety of areas has demonstrated that testing a wide variety of alternative network architectures by multiple independent teams can drive large gains in accuracy and generalizability.

# Data Availability

The dataset is composed of approximately 20,000 photos of areas of skin with ground-truth disease state labels for 23 skin conditions. All images have been harvested from open sources. The images have been randomly divided into a train set (90% of images) and test hold-out (10% of images) stratified by disease state. Test data can be withheld from analytical groups and used as a pure hold-out test group.

An example subset of the train data can be made available to developers. An open data set can be made available for training purposes.

The images have all been annotated by licensed dermatologists.

# Data Quality

The images are by design “in the wild” photos, as these are the best proxies for photos taken by non-professionals under non-clinical conditions using smartphone cameras.

# Annotation/Label Quality

The images have all been annotated by licensed dermatologists.

# Data Provenance

The images have been collected from open sources. All are available for non-commercial use without restriction other than attribution.

# Benchmarking

AI algorithms would be alternative implemented CNN model architectures, much like the well-known ImageNet Large Scale Visual Recognition Challenge. The challenge would be to most accurately assign images to each of the 23 skin condition categories. Participants would be provided a training set of approximately 18,000 images. Each training set image would be annotated with one ground-truth label of one of the 23 targeted skin conditions. Each participant would construct a model to generate a probability of membership in each of the 23 classes for an image.

The performance of the models constructed by each participant would be evaluated on the unseen test set of approximately 2,000 images. Each model will be applied to the test set images and generate 23 values representing the probability of the image representing each targeted condition. Evaluation metrics would be: (1) Top 1 Accuracy (% of images for which the skin condition with the highest predicted probability is correct); and (2) Top 5 Accuracy (% of images for which any one of the five skin conditions with the highest predicted probabilities is correct).

# Organizer

DAIGH has been created to encourage the exploitation and extension of the initial technology that we have constructed specifically to provide zero-cost software to improve skin disease diagnosis using smartphones. We have previously assembled datasets, constructed multiple CNN diagnostic models and benchmarked their performance. We have not previously run a challenge process with external participants.

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