|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ITU Logo | INTERNATIONAL TELECOMMUNICATION UNION  **TELECOMMUNICATION STANDARDIZATION SECTOR**  STUDY PERIOD 2017-2020 | | FG-AI4H-I-024-A01 | |
| **ITU-T Focus Group on AI for Health** | |
| **Original: English** | |
| **WG(s):** | | Plenary | E-meeting, 7-8 May 2020 | |
| **DOCUMENT** | | | | |
| **Source:** | | TG-Diabetes Driver | | |
| **Title:** | | Updated TDD (TG-Diabetes) | | |
| **Purpose:** | | Discussion | | |
| **Contact:** | | Andrés Valdivieso Ahnfelt Anastasia.ai, Chile | | Tel: +56996330961 Email: [avaldivieso@anastasia.ai](mailto:avaldivieso@anastasia.ai) |

|  |  |
| --- | --- |
| **Abstract:** | This document serves as initial Draft for the topic group TG- Primary and secondary diabetes prediction. This TDD will be created in a joint effort by the topic group and continuously improved over the upcoming meetings until it is finally approved by the focus group. TG- Primary and secondary diabetes prediction focus group which concerned the standardized benchmarking of AI applications aimed to improve a broad spectrum of pre-diabetes, diabetes care. Including, diabetes screening and detection of their complications, monitoring and treatment. Benchmarking is expected to yield more robust models and algorithms and to allow transparent comparisons of different models and algorithms from Automated Retinal Screening, Diabetic food, Predictive Population Risk Stratification, Clinical Decision Support and Patient Self-Management Tools. |

**CONTENTS**

| Page |
| --- |
| [1 Introduction 4](#_Toc39736271)  [1.1 Topic Description 4](#_Toc39736272)  [1.1.1 Topic Thematic Classification: 5](#_Toc39736273)  [1.1.2 Relevance 6](#_Toc39736274)  [1.1.3 Current approaches and gold standards for detection 6](#_Toc39736275)  [1.1.4 Impact of AI 7](#_Toc39736276)  [1.1.5 Impact of benchmarking AI Solutions 9](#_Toc39736277)  [1.2 Ethical considerations on usage of AI 9](#_Toc39736278)  [1.2.1 Ethical consideration of and benchmarking including its data acquisition 9](#_Toc39736279)  [1.3 Existing AI solutions (includes datasets, systems and benchmarks) 9](#_Toc39736280)  [1.3.1 Datasets 9](#_Toc39736281)  [1.3.2 Systems and Benchmarks 10](#_Toc39736282)  [2 AI4H Topic Group 11](#_Toc39736283)  [2.1 General mandate of the Topic Group 11](#_Toc39736284)  [2.2 Topic description document 12](#_Toc39736285)  [2.3 Subtopics 12](#_Toc39736286)  [2.4 Topic group participation 12](#_Toc39736287)  [2.5 Status of this Topic Group 12](#_Toc39736288)  [2.5.1 Status Update for Meeting H (Brasilia): 12](#_Toc39736289)  [2.5.2 Current members of the topic group: 12](#_Toc39736290)  [2.6 Next meetings 12](#_Toc39736291)  [2.6.1 Next steps to work on this document 13](#_Toc39736292)  [3 Method 13](#_Toc39736293)  [3.1 Overview of the benchmarking 13](#_Toc39736294)  [3.2 AI Input Data Structure 13](#_Toc39736295)  [3.3 AI Output Data Structure 14](#_Toc39736296)  [3.4 Test Data Labels 14](#_Toc39736297)  [3.5 Scores and Metrics: 15](#_Toc39736298)  [3.5.1 Classification Tasks: 15](#_Toc39736299)  [3.5.2 Segmentation Tasks: 16](#_Toc39736300)  [3.6 Undisclosed test data set collection (TBC) 17](#_Toc39736301)  [3.7 Benchmarking methodology and architecture (TBC) 17](#_Toc39736302)  [4 Reporting methodology (TBC) 17](#_Toc39736303)  [5 Results 17](#_Toc39736304)  [6 Discussion 18](#_Toc39736305)  [7 Declaration of conflict of interest 18](#_Toc39736306)  [References 18](#_Toc39736307) |

**List of Tables**

| Page |
| --- |
| [Table n1: AI tools for the clinical topics within TG-Diabetes 4](#_Toc39735746)  [Table 1 – FG-AI4H thematic classification scheme 5](#_Toc39735747) |

**List of Figures**

| Page |
| --- |
| **No table of figures entries found.** |

# Introduction

As part of the work of the WHO/ITU Focus Group (FG) AI for health (AI4H), this document specifies a standardized benchmarking approach for AI-based application for Diabetes and Pre-diabetes.

## Topic Description

This topic group is devoted to standardized benchmarking of artificial intelligence for Diabetes and Pre-diabetes, certain complications associated a self-assessment management tool. The specific conditions and diseases of diabetes and orediabetes and self-assessment management tools approach included in this document are categorized in the following categories (see Table 1):

1. **Predictive Population Risk Stratification and Clinical Decision Support**: Identification of diabetes sub-populations at higher risk for complications, hospitalization, readmissions, and detection and monitoring of diabetes and comorbidities.
2. **Diabetic retinopathy**: Diabetic retinopathy is a serious eye-disease caused by diabetes that affects blood vessels in the light-sensitive tissue called the retina that lines the back of the eye. It is the most common cause of vision loss among people with diabetes and the leading cause of vision impairment and blindness among working-age adults Worldwide. (Source TG-Ophthalmo Topic Diabetic retinopathy)
3. **Diabetic Foot**: Diabetic foot is a condition in which [foot ulcers](https://www.sciencedirect.com/topics/medicine-and-dentistry/foot-ulcer) are form on patients with [diabetes](https://www.sciencedirect.com/topics/medicine-and-dentistry/diabetes-mellitus). People with diabetic foot ulcers (DFUs) have a decreased quality of life and a 8% higher incidence of possible [lower extremity amputation](https://www.sciencedirect.com/topics/medicine-and-dentistry/leg-amputation) (LEA).
4. **Patient Self-Management Tools**: These are tools that provide more active self-management, but in highly defined ways. Examples include sound and text reminders from a tabletop appliance or perhaps a personal digital assistant or telephone, or devices allowing a patient to transmit data such as blood pressure readings with these people have to make choices and decisions about how to manage their life and their diabetes. Through good self-management, people with diabetes can improve their quality of life and reduce the risk of developing complications.

Table n1: AI tools for the clinical topics within TG-Diabetes



Additional complications and conditions that are relevant to this Topic Group may be added in the future.

### Topic Thematic Classification:

According to the current version of the thematic classification scheme document C-104 of the FG, the categorization of this topic “Primary and secondary diabetes prediction " is applicable as described in Table 1:

Table 1 – FG-AI4H thematic classification scheme

| **Level** | | **Thematic classification** |
| --- | --- | --- |
| Level 1 | Public Health (Level-1A) | 1.1. Health service  1.2. Health systems  1.3 Health expenditure  1.8 Cause-specific mortality and mobility  1.10. Non-communicable diseases  sub-classes applicable:  1. Epidemiology  4. Health services delivery  6. Community health  7. Behavioral health  8. Health economics  9. Informatics  10. Public health interventions  11. Public policy |
| Clinical Health  (Level-1B) | 1.1. Prevention  1.2. Diagnosis  1.3 Treatment  sub-classes applicable:  4. Cardiology  9. Endocrinology  10. Laboratory medicine  23.Ophtalmology |
| Level-2 Thematic Classification (Artificial Intelligence)  AI-benchmarking class type | | 1. Machine Learning  1.1. Classification  1.3. Clustering  1.4 Recommendation system  1.7. Anomaly detection  5. Perception  5.1. Visual recognition (photo/video)  5.2 Natural Language Processing (text/voice) |
| Level-3 Thematic Classification  (nature of data types) | | 3.1. Anonymized Electronic Health Record data  3.2. Medical Images, photographs  3.3 Non-medical data (socio economic, environmental, etc.)  3.5 Lab test result  3.6 Questionnaire responses |
| Level-4 Thematic Classification (origin of the data) | | 4.1 PACS  4.2 EHR  4.4 LIS  4.6 mHealth App  4.7 Medical Device |
| Level-5 Thematic Classification (data collectors) | | 5.1 Service provider (technologist or doctor)  5.2 Patient (or proxy person)  5.3 Machine-generated |

### Relevance

Diabetes is a global pandemic. An estimated 425 million people worldwide have diabetes, accounting for 12% of the world's health expenditures, and yet 1 in 2 person remain undiagnosed and untreated **(1)**. Type 2 diabetes is Diabetic retinopathy by the global obesity epidemic and a sedentary lifestyle that overwhelms the body's internal glucose control requiring exogenous insulin**(2)**. Millions of newborns are born to mothers with gestational diabetes. Pre-diabetic retinopathies born with type 1 diabetes mellitus, in which the body cannot produce insulin, require life-long insulin therapy. In the United States, diabetes is the leading cause of kidney failure, lower limb amputations, adult-onset blindness, and almost doubles the risk of heart attack and all-cause mortality, leading to hospitalization, long-term complications, and higher costs **(3)(4)**.

### Current approaches and gold standards for detection

1. **Predictive Population Risk Stratification and Clinical decision Support:**
2. **Diabetic retinopathy:**

Detection requires capturing a photograph of the retina using specialized equipment such as a slit-lamp and fundus camera. The image is then examined by an ophthalmologist, optometrist or a trained professional to detect abnormalities such as microaneurysms, exudates, hemorrhages, macular edema, etc. to determine if Diabetic retinopathy is present and its severity and stage of progression.

In general, Diabetic retinopathy can be classified as mild, moderate or vision-threatening, which includes severe non-proliferative Diabetic retinopathy (NPDiabetic retinopathy), proliferative Diabetic retinopathy (PDiabetic retinopathy) and diabetic macular edema (DME). Accurate diagnosis of Diabetic retinopathy from fundus camera images and grading its severity requires professional expertise and training.

The UK National Institute for Clinical Excellence (NICE) guideline states that a Diabetic retinopathy screening test should have sensitivity and specificity of at least 80% and 95% respectively, with a technical failure rate of less than 5%.[[1]](#endnote-1)

The gold standard photography method for the detection of Diabetic retinopathy is stereoscopic colour fundus photography in 7 standard fields (30°) as defined by the Early Treatment Diabetic retinopathy Study (ETDiabetic retinopathyS) group. (Source TG-Ophthalmo Topic Diabetic retinopathyr FG-AI4H-H-017-A01))

1. **Diabetic Foot TBC**
2. **Patient Self-Management Tools: TBC**

### Impact of AI

1. **Diabetes and Prediabetes**

Decades of well-designed studies have established that intensive therapy effectively delays the onset and slows the progression of diabetes-related complications, such as retinopathy, nephropathy, and neuropathy.[4](https://www.liebertpub.com/doi/full/10.1089/pop.2018.0129#B4) Yet, a recent study of 300,000 patients with type 2 diabetes who were started on medical therapy found that after 3 months, 31% of patients had discontinued their diabetes medications altogether: this increased to 44% by 6 months, and to 58% by 1 year. Only 40% eventually restarted diabetes medications.[5](https://www.liebertpub.com/doi/full/10.1089/pop.2018.0129#B5) Optimal care for persons with diabetes (PWDs) often is hampered by the absence of real-time, key health information necessary to make informed choices associated with intensive therapy and tight diabetes control. Although advances in technology offer unprecedented and inexpensive access to essential information for many individuals in many fields, its impact in the care of patients with diabetes seems rather limited. The challenges of real-time diabetes care information are compounded by the rapid expansion of medical knowledge. The index of biomedical literature contains more than 28 million articles as of June 2018 and is growing at a rate of more than 850,000 new citations each year.[6](https://www.liebertpub.com/doi/full/10.1089/pop.2018.0129#B6) Each person will generate more than 1 million gigabytes of health-related data in his or her lifetime, the equivalent of about 300 million books. An estimated 80% of health data is unstructured. This includes clinician notes, clinical trials, hospital records and discharge summaries, imaging and laboratory reports, and nonclinical data sources, including device and sensor data (often referred to as Internet of Things data), genomic data, and social determinants of health data.[7](https://www.liebertpub.com/doi/full/10.1089/pop.2018.0129#B7) Ninety percent of a person's health outcomes may be attributed to genomics and exogenous data, underscoring the importance of PWDs and their clinicians collecting and leveraging these data to make informed health choices.[8](https://www.liebertpub.com/doi/full/10.1089/pop.2018.0129#B8)

However, this large amount of data cannot be stored by the doctor's memory and will have to rely on solutions that will improve their abilities to be effective and efficient to do a better job in the future and on the other hand, seeks to change from curative medicine to a preventive and to be able to anticipate the complications of diabetes and pre-diabetes.

Rapid advances in artificial intelligence (AI) offer the promise of making both real-time structured and unstructured health data available for the care of PWDs. The Turing Archive for the History of Computing defines AI as “the science of making computers do things that require intelligence when done by humans.”[9](https://www.liebertpub.com/doi/full/10.1089/pop.2018.0129#B9) AI covers a broad range of approaches to simulating human intelligence and performing various reasoning tasks, such as visual perception, speech recognition, analytics, decision making, and translation between languages. Cognitive systems employ the spectrum of AI approaches to extend and scale human knowledge and expertise by enabling humans to leverage vast knowledge sources rapidly to solve problems.

Today, AI is harnessing massive amounts of vital information to meet consumer demand in every business, including health care. A 2017 survey found that 68% of mobile health app developers and publishers believe that diabetes continues to be the single most important health care field with the best market potential for digital health solutions within the near future, and that 61% see AI as the most disruptive technology shaping the digital health sector.[10](https://www.liebertpub.com/doi/full/10.1089/pop.2018.0129#B10) Although advances in AI for health care are being reported in the literature[11](https://www.liebertpub.com/doi/full/10.1089/pop.2018.0129#B11) and new AI-powered devices are being approved for diabetes care,[12](https://www.liebertpub.com/doi/full/10.1089/pop.2018.0129#B12) a systematic review of clinically relevant diabetes AI applications is missing.

The tables below show the possible cost impact of preventing complications due to better actions by having more segmented knowledge of patients during their total life cycle.

*Insert Economic Table of Impact*

1. **Diabetic retinopathy**

For Diabetic retinopathy diseases, vision loss and blindness can be delayed or prevented by early detection and treatment of the condition. This requires an examination and screening by a trained ophthalmologist or an eye care professional.

However, given the large numbers of people affected worldwide by these conditions, there are not sufficient specialists globally to screen everyone at risk. The shortfall is particularly acute in developing countries, including India, China and many countries in Asia and Africa. In addition to the dire shortage of trained professionals, many of the affected people live in remote areas with little or no access to an eye care clinic or a screening center.

In India, for example, there are over 72 million people with diabetes and an estimated 25 million have some stage of Diabetic retinopathy and about 7 million have VT Diabetic retinopathy. However, India only has 15,000 trained ophthalmologists, which in a nation with 1.3 billion people amounts to a mere 9 specialists per million. Kenya, with a population of 48 million has less than 100 ophthalmologists, and Angola, less than 20 for 29 million people.[[2]](#endnote-2)

*(The following is from Calligo Technologies contribution FG-AI4H-G-028 received during meeting G.)*

With the advent of Edge Computing, health care industry has transformed itself considerably, while hospitals and clinics are gearing up to take better and faster care of their patients. In fact, Edge Computing has permeated the industry in such a powerful manner that clinicians and doctors heavily rely on them to treat patients. As more and more devices get connected in the health care industry, networking among them all has really become huge because the data that keeps comes in is never going to slow down.

A frequent problem in mass eyecare checkups is that the quality of images captured might not always be usable for an ophthalmologist to grade for Diabetic retinopathy. In such situations, the patients are asked to come back and undergo the process again. Now with AI, the system checks the image as soon as it is clicked and prompts the technician to click another image in case it is not good enough. Now, even a minimally skilled technician can take usable images of the eye fundus.

Once usable images are captured, the system grades the images, again in real-time, and identifies if the images have Diabetic retinopathy. In case a patient is found to be Diabetic retinopathy positive, they are advised to consult an ophthalmologist to determine the next course of action.

Checking on patients with high risk problems and ensuring a more effective, customized treatment approach can thus be facilitated. Lack of data makes the creation of patient-centric care programs more difficult, so one can clearly understand why utilizing big data can be so highly important in the industry ***(Source TG-Ophthalmo Topic Diabetic retinopathyr FG-AI4H-H-017-A01))***

1. **Diabetic Foot**

The cost of health care for ulceration and amputation in diabetes in 2014-2015 is estimated at between £837 million and £962 million; 0.8% to 0.9% of the National Health Service (NHS) budget for England. More than 90% of expenditure was related to ulceration, and 60% was for care in community, outpatient and primary settings. For inpatients, multiple regression analysis suggested that ulceration was associated with a length of stay 8.04 days longer (95% confidence interval 7.65 to 8.42) than that for diabetes admissions without ulceration.

1. **Patient Self-Management Tools**

TBC

### Impact of benchmarking AI Solutions

1. **Predictive Population Risk Stratification and Clinical decision Support: TBC**
2. **Diabetic retinopathy:**

An accurate way of benchmarking the performance of AI solutions to detect and diagnose Diabetic retinopathy, can have a major impact on selecting and implementing the best solution to ad Diabetic retinopathies the global healthcare challenge posed by these diseases specially in the LMICs. This can in turn improve the lives of millions at risk for vison impairment and vision loss globally because they do not have access to human experts and infrastructure to get screened. This also fulfils the important objective of achieving the UN’s SDGs in health. (Source TG-Ophthalmo Topic Diabetic retinopathy FG-AI4H-H-017-A01)

1. **Diabetic Foot TBC**
2. **Patient Self-Management Tools: TBC**

## Ethical considerations on usage of AI

* Technical robustness, safety, and accuracy
* Data governance (storage, access and security) and privacy
* Bias and fairness of training datasets
* Explainable
* Accountability

### Ethical consideration of and benchmarking including its data acquisition

* Ethical acquisition of data, including necessary IRB reviews
* Privacy: No personally identifiable information
* Bias and fairness: The benchmarking dataset must capture sufficient variations and diversities (of subjects and settings) which are clearly outlined

(for more information see working groups [Ethical considerations on AI for health (WG-Ethics)](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/wg/SitePages/WG-Ethics.aspx).

## Existing AI solutions (includes datasets, systems and benchmarks)

### Datasets

#### Predictive Population Risk Stratification and Clinical decision Support Datasets: TBC

#### Diabetic retinopathy Datasets:

* Publicly available datasets include the EyePACS dataset (around 90,000 fundus images, 5 levels of severity), [[3]](#endnote-3)
* MESSIDOR dataset (1,200 images, 4 levels of severity), [[4]](#endnote-4)
* The DIARETDB dataset (around 200 images marked with lesions), etc.[[5]](#endnote-5)
* Diabetic retinopathy datasets from Kaggle:
* Kaggle Diabetic retinopathy Challenge 2015: 35,000 images of Diabetic retinopathy classified into 5 levels of severity (No Diabetic retinopathy, Mild, Moderate, Severe, Proliferative Diabetic retinopathy).
* APTOS 2019 Blindness Detection Challenge: 3664 Images classified into 5 levels of severity ((No Diabetic retinopathy, Mild, Moderate, Severe, Proliferative Diabetic retinopathy).

#### Diabetic Foot Datasets TBC

#### Patient Self-Management Tools Datasets: TBC

### Systems and Benchmarks

1. **Predictive Population Risk Stratification and Clinical decision Support: TBC**
2. **Diabetic retinopathy Systems and Benchmarks:**

A team at Google published results in 2016 of a study for detecting Diabetic retinopathy working with doctors in India and the US. The results show that their AI model’s performance for Diabetic retinopathy detection and grading its severity was on-par with that of ophthalmologists. Their model had a combined accuracy score of 0.95, which was slightly better than the median of the 8 ophthalmologists consulted (measured at 0.91). [[6]](#endnote-6)

Currently, IDx-Diabetic retinopathy is the first FDA approved device for AI Diabetic retinopathy screening. Based on a customized CNN architecture and lesion characteristics, this device can achieve a sensitivity of 96.8% and a specificity of 87%.[[7]](#endnote-7)

The best reported performance on binary classification of no Diabetic retinopathy/non-referable Diabetic retinopathy vs. referable Diabetic retinopathy is a sensitivity of 94% and specificity of 98% [[8]](#endnote-8)

This work combined features both from deep ResNet and from meta-data and classified the features with a gradient boosting decision tree.

For five level classification of no Diabetic retinopathy, mild, moderate, severe non-proliferative Diabetic retinopathy, and proliferative Diabetic retinopathy [[9]](#endnote-9) [[10]](#endnote-10) [[11]](#endnote-11), the best accuracy reported is 96% by a combination of GoogleNet and ResNet model.

In the APTOS 2019 Blindness Detection challenge organized by Kaggle in Sep 2019, 2931teams competed, and the top solution achieved a QuaDiabetic retinopathyatic Kappa weighted score of 0.9361 on an undisclosed test data set.

*(The following is from Calligo Technologies contribution FG-AI4H-G-028 received during meeting G.)*

“Calligo Health Engine” is an Edge Analytics solution which is easy to use, industry gradable, low cost & low resource and is capable of identifying Diabetic retinopathy using Artificial Intelligence with an accuracy over 96% and within Seconds. (Source TG-Ophthalmo Topic Diabetic retinopathy: FG-AI4H-H-017-A01)

1. **Diabetic Foot Systems and Benchmarks TBC**
2. **Patient Self-Management Tools Systems and Benchmarks: TBC**

# AI4H Topic Group

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 (13) topic groups. These are concerned with: AI and cardiovascular disease risk prediction, dermatology, falls among the elderly, histopathology, neuro-cognitive disorders, outbreak detection, ophthalmology, psychiatry, radiotherapy, snakebite and snake identification, symptom assessment, tuberculosis and Volumetric chest computed tomography. A current list can be found at the FGAI4H website at:

<https://www.itu.int/en/ITU-T/focusgroups/ai4h/Pages/default.aspx>

As the work by the Focus Group continues, new Topic Groups will be created. To organize the Topic Groups, for each topic the Focus Group members choose a topic. The responsibilities of the topic **Primary and secondary diabetes prediction** are still to be defined and are likely to change over time. The preliminary and yet-to-confirm list of the responsibilities includes:

* Creating the initial **Primary and secondary diabetes prediction** of the topic description document.
* Reviewing the input documents for the topic and moderating the integration in a dedicated session at each Focus Group meeting.
* Organizing regular phone calls to coordinate work on the topic description document between meetings.

## General mandate of the Topic Group

The Topic Group is a concept specific to the AI4H-FG. The preliminary responsibilities of the Topic Group members are:

1. Provide a forum for open communication among various stakeholders.
2. Agree upon the benchmarking tasks of this topic and scoring metrics.
3. Facilitate the collection of high-quality labeled test data from different sources.
4. Clarify the input and output format of the test data.
5. Define and set-up the technical benchmarking infrastructure.
6. Coordinate the benchmarking process in collaboration with the Focus Group management and working groups.

## Topic description document

The primary output of each Topic Group is the topic description document (TDD) specifying all relevant aspects of the benchmarking for the individual topics. **This document is the TDD for the Topic Group on “Primary and secondary diabetes prediction)” (TG-Diabetes and Prediabetes)** The document has been developed one FG-AI4H meetings in Brasilia. Any contributions from members have been incorporated in the relevant sections. Suggested changes to the document will be submitted as input documents for each meeting. The relevant changes will then be discussed and integrated into an official output document until the TDD ready for the first official benchmarking.

## Subtopics

Topic groups summarize similar AI benchmarking use cases to limit the number of use case specific meetings at the Focus Group meetings and to share similar parts of the benchmarking. However, in some cases, it is expected that inside a Topic Group different subtopic Groups can be established to pursue different topic-specific specializations. TG- Diabetes and Prediabetes will start with separate areas of analysis.

1. Predictive Population Risk Stratification and Clinical decision Support
2. Diabetic retinopathy: (Automated Retinal Screening). (See information and description of (Source TG-Ophthalmo Topic Diabetic retinopathy: FG-AI4H-H-017-A01)
3. Diabetic Foot
4. Patient Self-Management Tools

## Topic group participation

The participation in both the focus and Topic Group is generally open and free of charge. Anyone who is from a member country of the ITU may participate. On the 14th of March 2019 the ITU published an official “call for participation” document outlining the process for joining the Focus Group and the Topic Group. For this topic, the corresponding call can be found at <https://www.itu.int/en/ITU-T/focusgroups/ai4h/Pages/ophthalmo.aspx>.

## Status of this Topic Group

### Status Update for Meeting H (Brasilia):

The group is currently being founded.

### Current members of the topic group:

1. Andrés Valdivieso Ahnfelt Director of Innovation of Anastasia.ai
2. Marcos Lacayo CEO of Estación Vital
3. Marcelo Guerra COO Tecnigen

## Next meetings

The Focus Groups meets about every two months at changing locations. The upcoming meetings are:

* I: 7-8 May 2020 – TBC; An up to date list can be found at the official [ITU FG AI4H website](https://www.itu.int/en/ITU-T/focusgroups/ai4h/Pages/default.aspx).
* Internal Meetings of the Topic group TBD

### Next steps to work on this document

The document is first to be fully populated. Then, specific subsections are to be expanded on, depending on the specific activities and directions this group takes.

# Method

## Overview of the benchmarking

1. **Predictive Population Risk Stratification and Clinical decision Support:**

TBD

1. **Diabetic retinopathy:**

The benchmarking of the algorithms for detecting Diabetic retinopathy, will be done on a sufficiently large and previously undisclosed test data set. All data will be provided as per the data acceptance guidelines published by the focus group. All data will be labelled by licensed ophthalmologists or eye-care professionals. (Source TG-Ophthalmo Topic Diabetic retinopathy: FG-AI4H-H-017-A01)

1. **Diabetic Foot TBD**
2. **Patient Self-Management Tools**

TBD

## AI Input Data Structure

1. **Predictive Population Risk Stratification and Clinical decision Support:**

TBC

1. **Diabetic retinopathy:**

Images of each retina captured with fundus cameras should be submitted as separate files in the following format: *(Source TG-Ophthalmo Topic Diabetic retinopathy: FG-AI4H-H-017-A01)*

Image File Format: JPG, PNG or Dicom format

Image File Names: Images names will be anonymized to exclude any patient identifying information.

Image Resolution: the images will be supplied in their original resolution as captured from the fundus cameras.

1. **Diabetic Foot :**

Images of each retina captured with fundus cameras should be submitted as separate files in the following format:

* 1. Image File Format: JPG, PNG format or Dicom
  2. Image File Names: Images names will be anonymized to exclude any patient identifying information.
  3. Image Resolution: the images will be supplied in their original resolution as captured from the fundus cameras.

1. **Patient Self-Management Tools**

## AI Output Data Structure

1. **Predictive Population Risk Stratification and Clinical decision Support: TBC**
2. **Diabetic retinopathy:**

The output of the algorithm should be a CSV file in text format with the following columns: (Source TG-Ophthalmo Topic Diabetic retinopathy: FG-AI4H-H-017-A01)

Name of the image file processed (for example: I705656.JPG OR L566768.PNG)

The diagnosis of the retinal image as per the algorithm. The labels will depend upon the specific condition and the type of classification that is being benchmarked:

Classification: All Diabetic retinopathy severity levels:

0 (Non-gradable Image) (TBD: should this be one of the classifications.)

1(No Diabetic retinopathy)

2 (Mild)

3 (Moderate NPDiabetic retinopathy)

4 (Severe NPDiabetic retinopathy)

5 (PDiabetic retinopathy)

Binary Classification: Referable or Non-referable Diabetic retinopathy:

0 (Ungradable Image. TBD)

1 (Non-referable Retinopathy – No Diabetic retinopathy or Mild Diabetic retinopathy)

(Referable Retinopathy - Moderate, Severe, PDiabetic retinopathy)

1. **Diabetic Foot TBC**
2. **Patient Self-Management Tools TBC**

## Test Data Labels

1. **Predictive Population Risk Stratification and Clinical decision Support: TBC TBD**
2. **Diabetic retinopathy**

A separate CSV file in text format will be provided containing the following columns:

Name of the Image File (example: R705656.JPG OR L566768.PNG)

Label or Annotation of the Image that contains the diagnosis of the retinal image. The labels will depend upon the specific condition that is being benchmarked and also the type of classification. Currently, the following are being proposed:

1. Classification: All Diabetic retinopathy severity levels:

* 0 (Non-gradable Image) (TBD: should this be one of the classifications.)
* 1(No Diabetic retinopathy)
* 2 (Mild)
* 3 (Moderate NPDiabetic retinopathy)
* 4 (Severe NPDiabetic retinopathy)
* 5 (PDiabetic retinopathy)

1. Classification: Referable or Non-referable Diabetic retinopathy:

* 0 (Ungradable Image. TBD)
* 1 (Non-referable Retinopathy – No Diabetic retinopathy or Mild Diabetic retinopathy)
* 2 (Referable Retinopathy - Moderate, Severe, PDiabetic retinopathy)

1. **Diabetic Foot TBC**
2. **Patient Self-Management Tools TBC**

## Scores and Metrics:

All metrics will be computed based on the performance of the algorithm on the undisclosed test dataset. The scores and metrics used for benchmarking AI will depend upon the type of task performed by the AI, which for the conditions in this topic group would generally be either classification or segmentation.

### Classification Tasks:

1. **Predictive Population Risk Stratification and Clinical decision Support: TBD**
2. **Diabetic retinopathy**

Classification of the conditions being considered may be either binary (2 classes) – for example Diabetic retinopathy or no Diabetic retinopathy, for example, in the case of Diabetic retinopathy an image may be classified as having No Diabetic retinopathy, mild, moderate, severe or PDiabetic retinopathy (5 classes).

We start with a few definitions:

* An instance is either a single image (for classification tasks), or a patch or a pixel of an image (for segmentation tasks).
* True Positive (TP) is the number of positive (disease) instances which are correctlyclassified.
* True Negative (TN) is the number of negative (normal) instances which are correctly classified.
* False Positive (FP) is the number of positive (disease) instances which are incorrectly classified.
* False Negative (FN) is the number of negative (normal) instances which are incorrectly classified.

Based on the above definitions, the following are the most common metrics used to evaluate performance of algorithms (Source TG-Ophthalmo Topic Diabetic retinopathy: FG-AI4H-H-017-A01)

1. **Diabetic Foot TBC**
2. **Patient Self-Management Tools TBC**

#### Binary Classification Tasks:

1. **Predictive Population Risk Stratification and Clinical decision Support: TBD**
2. ***Diabetic retinopathy***

* Sensitivity (Recall) or True Positive Rate is the proportion of correctly classified positive (disease) instances. It is calculated as TP / (TP + FN)
* Specificity or True negative rate is the proportion of correctly classified negative (normal) instances. It is calculated as TN / (TN + FP)
* Precision or Positive Predictive Value is the fraction of positive (disease) instances that are correctly classified. It is calculated as TP / (TP + FP).
* F1-Score combines Precision and Recall into a single metric. It is calculated as the harmonic mean of Precision and Recall. It is calculated as 2 x (Precision x Recall) / (Precision + Recall)
* Accuracy is the proportion of instances that are correctly classified. It is calculated as (TP + TN) / (TP + FP + TN + FN)

AUC (Area Under Receiver Operating Curve or ROC): The ROC is a plot of True Positive Rate (Sensitivity) vs. False Positive Rate (1- Specificity)) at different predictive thresholds of the classifier. The AUC has a value between 0 and 1. The closer it is to 1 the better the performance. (Source TG-Ophthalmo Topic Diabetic retinopathy: FG-AI4H-H-017-A01)

1. **Diabetic Foot TBC**
2. **Patient Self-Management Tools TBC**

#### Multi-label Classification Tasks:

1. **Predictive Population Risk Stratification and Clinical decision Support: TBD**
2. **Diabetic retinopathy**

In this case the metrics most commonly used are:

1. **Accuracy:** the proportion of instances that are correctly classified (the accuracy of each instance class is summed across all instance classes and divided by the number of all instance classes)

**Cohen's Kappa and QuaDiabetic retinopathyatic Weighted Kappa:** This metric measures the degree of agreement between two different raters - for example between an AI model's predictions and the corresponding human verified values. This metric typically varies from 0 (in case of random agreement between raters) to 1 (complete agreement between raters). It has been used in the past in Kaggle competitions to measure the effectiveness of algorithms for detecting Diabetic retinopathy. ***(Source TG-Ophthalmo Topic Diabetic retinopathy: FG-AI4H-H-017-A01)***

1. **Diabetic Foot TBC**
2. **Patient Self-Management Tools TBC**

### Segmentation Tasks:

1. **Predictive Population Risk Stratification and Clinical decision Support: TBD**
2. **Diabetic retinopathy**

Segmentation Tasks: Intersection over Union (IOU): This metric is used only for segmentation tasks. IOU is defined as follows: IOU = Area (A ∩ G) / Area ( A ∪ G) where A indicates the segmentation from the algorithm and G indicates the manual ground truth segmentation of an image.

In all the above cases – higher values are better, and algorithms would be ranked in descending order of these metrics. ***(Source TG-Ophthalmo Topic Diabetic retinopathy: FG-AI4H-H-017-A01)***

1. **Diabetic Foot TBC**
2. **Patient Self-Management Tools TBC**

## Undisclosed test data set collection (TBC)

* raw data acquisition / acceptance
* test data source(s): availability, reliability,
* labelling process / acceptance
* bias documentation process
* quality control mechanisms
* discussion of the necessary size of the test data set for relevant benchmarking results
* specific data governance derived by general data governance document (currently C-004)

## Benchmarking methodology and architecture (TBC)

* technical architecture
* hosting (IIC, etc.)
* possibility of an online benchmarking on a public test dataset
* protocol for performing the benchmarking (who does what when etc.)
* AI submission procedure including contracts, rights, IP etc. considerations

# Reporting methodology (TBC)

* Report publication in papers or as part of ITU documents
* Online reporting
* public leaderboards vs. private leaderboards
* Credit-Check like on approved sharing with selected stakeholders
* Report structure including an example

Frequency of benchmarking

# Results

Reports of the different benchmarking runs will be inserted here

# Discussion

* Discussion of the insights from executing the benchmarking on
* external feedback on the whole topic and its benchmarking
* technical architecture
* data acquisition
* benchmarking process
* benchmarking results
* field implementation success stories

# Declaration of conflict of interest

by each contributor to this document

# References

International Diabetes Federation (IDF). IDF diabetes atlas, 7th edition. Brussels, Belgium: International Diabetes Federation, 2015. [Google Scholar](http://scholar.google.com/scholar?hl=en&q=+International+Diabetes+Federation+%28IDF%29.+IDF+diabetes+atlas%2C+7th+edition.+Brussels%2C+Belgium%3A+International+Diabetes+Federation%2C+2015.)

Liu J, Wang R, Ganz ML, Paprocki Y, Schneider D, Weatherall J. The burden of severe hypoglycemia in type 2 diabetes. Curr Med Res Opin 2018;34:179–186. [Crossref](https://www.liebertpub.com/servlet/linkout?suffix=B2&dbid=16&doi=10.1089%2Fpop.2018.0129&key=10.1080%2F03007995.2017.1391080), [Medline](https://www.liebertpub.com/servlet/linkout?suffix=B2&dbid=8&doi=10.1089%2Fpop.2018.0129&key=29017368), [Google Scholar](http://scholar.google.com/scholar_lookup?hl=en&volume=34&publication_year=2018&pages=179-186&journal=Curr+Med+Res+Opin&author=J+Liuauthor=R+Wangauthor=ML+Ganzauthor=Y+Paprockiauthor=D+Schneiderauthor=J+Weatherall&title=The+burden+of+severe+hypoglycemia+in+type+2+diabetes)

Centers for Disease Control and Prevention. Estimates of diabetes and its burden in the United States. Washington, DC: US Department of Health and Human Services, 2017:1–20. [Google Scholar](http://scholar.google.com/scholar?hl=en&q=+Centers+for+Disease+Control+and+Prevention.+Estimates+of+diabetes+and+its+burden+in+the+United+States.+Washington%2C+DC%3A+US+Department+of+Health+and+Human+Services%2C+2017%3A1%E2%80%9320.)

Irene Dankwa-Mullan, Marc Rivo, Marisol Sepulveda, Yoonyoung Park, Jane Snowdon, and Kyu Rhee.Population Health Management.Jun 2019.229-242.

Lewis K. Humans and systems: creating natural interfaces to augment human ability. [https://www.ibm.com/blogs/internet-of-things/creating-natural-interfaces/](http://www.ibm.com/blogs/internet-of-things/creating-natural-interfaces/) Accessed July 18, 2018. [Google Scholar](http://scholar.google.com/scholar?hl=en&q=+LewisK.+Humans+and+systems%3A+creating+natural+interfaces+to+augment+human+ability.+https%3A%2F%2Fwww.ibm.com%2Fblogs%2Finternet-of-things%2Fcreating-natural-interfaces%2F+Accessed+July+18%2C+2018.)

Aggarwal M, Madhukar M. IBM's Watson analytics for health care: a miracle made true. Hershey, PA: IGI Global, 2016:117–134. [Google Scholar](http://scholar.google.com/scholar?hl=en&q=+AggarwalM%2C+MadhukarM.+IBM%27s+Watson+analytics+for+health+care%3A+a+miracle+made+true.+Hershey%2C+PA%3A+IGI+Global%2C+2016%3A117%E2%80%93134.)

Turing Archive for the History ofComputing. [www.alanturing.net/turing\_archive/archive/index/archiveindex.html](http://www.alanturing.net/turing_archive/archive/index/archiveindex.html) Accessed July 18, 2018. [Google Scholar](http://scholar.google.com/scholar?hl=en&q=+Turing+Archive+for+the+History+of+Computing.+www.alanturing.net%2Fturing_archive%2Farchive%2Findex%2Farchiveindex.html+Accessed+July+18%2C+2018.)

Research 2 Guidance. Top 3 therapy fields with the best market potential for digital health apps. [https://research2guidance.com/top-3-therapy-fields-with-the-best-market-potential-for-digital-health-apps/](http://research2guidance.com/top-3-therapy-fields-with-the-best-market-potential-for-digital-health-apps/)Accessed July 18, 2018. [Google Scholar](http://scholar.google.com/scholar?hl=en&q=+Research+2+Guidance.+Top+3+therapy+fields+with+the+best+market+potential+for+digital+health+apps.+https%3A%2F%2Fresearch2guidance.com%2Ftop-3-therapy-fields-with-the-best-market-potential-for-digital-health-apps%2F+Accessed+July+18%2C+2018.)

Yeung S, Downing NL, Fei-Fei L, Milstein A. Bedside computer vision—moving artificial intelligence from Diabetic retinopathyr assistance to patient safety. N Engl J Med 2018;378:1271–1273. [Crossref](https://www.liebertpub.com/servlet/linkout?suffix=B11&dbid=16&doi=10.1089%2Fpop.2018.0129&key=10.1056%2FNEJMp1716891), [Medline](https://www.liebertpub.com/servlet/linkout?suffix=B11&dbid=8&doi=10.1089%2Fpop.2018.0129&key=29617592), [Google Scholar](http://scholar.google.com/scholar_lookup?hl=en&volume=378&publication_year=2018&pages=1271-1273&journal=N+Engl+J+Med&author=S+Yeungauthor=NL+Downingauthor=L+Fei-Feiauthor=A+Milstein&title=Bedside+computer+vision%E2%80%94moving+artificial+intelligence+from+driver+assistance+to+patient+safety)

Young K. Newly approved software uses AI to improve Diabetic retinopathy. [https://www.jwatch.org/fw114063/2018/04/12/newly-approved-software-uses-ai-detect-diabetic](http://www.jwatch.org/fw114063/2018/04/12/newly-approved-software-uses-ai-detect-diabetic)Accessed July 18, 2018. [Google Scholar](http://scholar.google.com/scholar?hl=en&q=+YoungK.+Newly+approved+software+uses+AI+to+improve+diabetic+retinopathy.+https%3A%2F%2Fwww.jwatch.org%2Ffw114063%2F2018%2F04%2F12%2Fnewly-approved-software-uses-ai-detect-diabetic+Accessed+July+18%2C+2018.)

Yeung S, Downing NL, Fei-Fei L, Milstein A. Bedside computer vision—moving artificial intelligence from Diabetic retinopathyr assistance to patient safety. N Engl J Med 2018;378:1271–1273. [Crossref](https://www.liebertpub.com/servlet/linkout?suffix=B11&dbid=16&doi=10.1089%2Fpop.2018.0129&key=10.1056%2FNEJMp1716891), [Medline](https://www.liebertpub.com/servlet/linkout?suffix=B11&dbid=8&doi=10.1089%2Fpop.2018.0129&key=29617592), [Google Scholar](http://scholar.google.com/scholar_lookup?hl=en&volume=378&publication_year=2018&pages=1271-1273&journal=N+Engl+J+Med&author=S+Yeungauthor=NL+Downingauthor=L+Fei-Feiauthor=A+Milstein&title=Bedside+computer+vision%E2%80%94moving+artificial+intelligence+from+driver+assistance+to+patient+safety)

Young K. Newly approved software uses AI to improve Diabetic retinopathy. [https://www.jwatch.org/fw114063/2018/04/12/newly-approved-software-uses-ai-detect-diabetic](http://www.jwatch.org/fw114063/2018/04/12/newly-approved-software-uses-ai-detect-diabetic)Accessed July 18, 2018. [Google Scholar](http://scholar.google.com/scholar?hl=en&q=+YoungK.+Newly+approved+software+uses+AI+to+improve+diabetic+retinopathy.+https%3A%2F%2Fwww.jwatch.org%2Ffw114063%2F2018%2F04%2F12%2Fnewly-approved-software-uses-ai-detect-diabetic+Accessed+July+18%2C+2018.)

https://towardsdatascience.com/machine-learning-for-diabetes-562dd7df4d42 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5257026/

https://www.kaggle.com/uciml/pima-indians-diabetes-database https://hackernoon.com/ml-for-diabetes-from-bangladesh-d99d1d058d82

https://www.fundaciondiabetes.org/general/noticia/14091/primer-proyecto-que-usa-la-inteligencia-artificial-para-tratar-la-diabetes

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3810070/ https://arxiv.org/abs/1801.04062 https://arxiv.org/abs/1808.06670 a Libros y Articulos The diabetes challenge in Chile Developing Knowledge-enhanced Chronic Disease Risk Prediction Models from Regional EHR Repositories (Jing-Mei, Ph.D.)

Machine learning for chronic disease (Katherine E. Niehaus and David A. Clifton) PEPPER: Patient Empowerment Through Predictive Personalised Decision Support (Pau Herrero, Beatriz Lopez y Clare Martin)

1. National Institute for Clinical Excellence. Management of Type 2 Diabetes: Retinopathy—Screening and Early Management. London, UK: Inherited Clinical Guideline; 2002 [↑](#endnote-ref-1)
2. International Council of Ophthalmology. Number of Ophthalmologists in Practice and Training Worldwide. <http://www.icoph.org/ophthalmologists-worldwide.html> [↑](#endnote-ref-2)
3. <http://www.eyepacs.com/data-analysis> [↑](#endnote-ref-3)
4. <http://www.adcis.net/en/third-party/messidor/> [↑](#endnote-ref-4)
5. <http://www.it.lut.fi/project/imageret/diaretdb1/> [↑](#endnote-ref-5)
6. Varun Gulshan, PhD1; Lily Peng, MD, PhD1; Marc Coram, PhD; et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. <https://jamanetwork.com/journals/jama/fullarticle/2588763> [↑](#endnote-ref-6)
7. M. Abràmoff, Y. Lou, A. Erginay, W. Clarida, R. Amelon, J. Folk, M. Niemeijer, Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning, Investigative ophthalmology & visual science 57 (13) (2016) 5200–5206. [↑](#endnote-ref-7)
8. 1. R. Gargeya, T. Leng, Automated identification of diabetic retinopathy using deep learning, Ophthalmology 124 (7) (2017) 962–969.

   [↑](#endnote-ref-8)
9. 1. H. Takahashi, H. Tampo, Y. Arai, Y. Inoue, H. Kawashima, Applying artificial intelligence to disease staging: Deep learning for improved staging of diabetic retinopathy, PloS one 12 (6) (2017) e0179790.

   [↑](#endnote-ref-9)
10. 1. G. García, J. Gallardo, A. Mauricio, J. López, C. Del Carpio, Detection of diabetic retinopathy based on a convolutional neural network using retinal fundus images, in: International Conference on Artificial Neural Networks, Springer, 2017, pp. 635–642.

    [↑](#endnote-ref-10)
11. V. Gulshan, L. Peng, M. Coram, M. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, et al., Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs, Jama 316 (22) (2016) 2402–2410. [↑](#endnote-ref-11)