|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ITU Logo | INTERNATIONAL TELECOMMUNICATION UNION  **TELECOMMUNICATION STANDARDIZATION SECTOR**  STUDY PERIOD 2017-2020 | | FGAI4H-I-015-A01 | |
| **ITU-T Focus Group on AI for Health** | |
| **Original: English** | |
| **WG(s):** | | Plen | E-meeting, 7-8 May 2020 | |
| **DOCUMENT** | | | | |
| **Source:** | | TG-MCH Topic Driver | | |
| **Title:** | | Att.1 – TDD update (TG-MCH) [same as Meeting H] | | |
| **Purpose:** | | Discussion | | |
| **Contact:** | | Raghu Dharmaraju, Hafsa Mwita Wadhwani AI, India; Zanzibar University, Tanzania | | Email: rdharmaraju@gmail.com, hafsa.m.mwita@gmail.com |
| **Contact:** | | Hafsa Mwita University of Zanzibar Tanzania | | Email: [hafsa.m.mwita@gmail.com](mailto:hafsa.m.mwita@gmail.com) |

|  |  |
| --- | --- |
| **Abstract:** | This document serves as initial “skeleton” document (“DEL10.7”) for the forthcoming topic description document (TDD) of the topic group *Maternal and child health (TG-MCH)*, which is concerned with the standardized benchmarking of AI in the support of mothers and their children, e.g. through automated child growth monitoring. The outline will follow the template structure defined in FGAI4H-C-105. This TDD draft 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.  This version of the TDD is the same as seen in Meeting H (FGAI4H-H-015-A01), reproduced for easier reference as a Meeting I document. |

Table of Contents

[1 Introduction 3](#_Toc29900634)

[1.1 Document Structure 3](#_Toc29900635)

[1.2 Topic Description 3](#_Toc29900636)

[1.3 Ethical Considerations 4](#_Toc29900637)

[1.3.1 Ethical considerations on usage of AI 4](#_Toc29900638)

[1.3.2 Ethical consideration of and benchmarking including its data acquisition 4](#_Toc29900639)

[1.4 Existing AI Solutions 4](#_Toc29900640)

[1.5 Existing work on benchmarking 5](#_Toc29900641)

[2 AI4H Topic group 5](#_Toc29900642)

[3 Method 5](#_Toc29900643)

[3.1 AI Input Data Structure 5](#_Toc29900644)

[3.2 AI Output Data Structure 5](#_Toc29900645)

[3.3 Test Data Labels 5](#_Toc29900646)

[3.4 Scores & Metrics 6](#_Toc29900647)

[3.5 Undisclosed Test Data Set Collection 6](#_Toc29900648)

[3.6 Benchmarking Methodology and Architecture 6](#_Toc29900649)

[3.7 Reporting Methodology 6](#_Toc29900650)

[4 Results 7](#_Toc29900651)

[5 Discussion 7](#_Toc29900652)

[6 Declaration of Conflict of Interest 7](#_Toc29900653)

# Introduction

Improving the health and well-being of mothers, infants, and children is one of the most important public health goals worldwide. Every day, an estimated 810 women die from causes related to pregnancy or childbirth and over 15,000 children die from preventable diseases. AI as a tool in maternal and child health care will benefit individuals and communities across the world, especially in low-resource settings.

The aim of this document therefore is to develop a standardised benchmarking approach for AI for maternal & child health. The initial focus of this topic group will be growth monitoring (from birth to early adolescence), followed by other classes of algorithms, such as those for pregnancy risk stratification.

## Document Structure

* overview of the whole document

## Topic Description

This topic group is dedicated to AI for Maternal & Child Health (C-043).

Improving the health and well-being of mothers, infants, and children is one of the most important public health goals worldwide. Every day, an estimated 810 women die from causes related to pregnancy or childbirth and over 15,000 children die from preventable diseases. Malnourished children, particularly those with severe acute malnutrition, have a higher risk of death from common childhood illness such as diarrhoea, pneumonia, and malaria. Nutrition-related factors contribute to about 45% of deaths in children under-5 years of age. Current trends predict that 52 million children under 5 will die between 2019 and 2030. Almost half of these under-five deaths will be newborns whose deaths could be prevented by providing high quality antenatal care, skilled care at birth, postnatal care for mothers and their babies, and care of small and sick newborns. 94% of maternal deaths and over 80% of under-5 deaths occur in low and lower middle-income countries.

AI as a tool in maternal and child health care will benefit individuals and communities across the world, especially in low-resource settings. Some examples of potential AI-based applications in this area include:

1. Pregnancy risk estimation and stratification: Algorithms that predict a pregnant woman’s risk of developing complications or having adverse pregnancy outcomes in order to identify high-risk pregnancies
2. Early warning systems for labour rooms and neonatal intensive care units
3. Health screening tools: Smartphone-based tools to screen for common diseases such as pneumonia, jaundice, anaemia, etc
4. Growth monitoring: Smartphone-based tools to screen newborns, infants and children for low birth weight, inadequate growth and development, and malnutrition

In developing countries, the burden of service delivery of health services pertaining to maternal and child health falls on frontline health workers, who have limited skills and training and are often overworked and underpaid. In this scenario, AI can help close the expertise gap and lead to better monitoring and accountability by enabling easy, accurate and tamper-proof screening.

The initial focus of this topic group will be growth monitoring (from birth to early adolescence), followed by other classes of algorithms, such as those for pregnancy risk stratification. In growth monitoring, the gold standard is manual anthropometry carried out by trained medical experts.

While there are several research and commercial groups working on AI applications in this area, the lack of consistent standardization makes it difficult for organisations like the WHO, governments, and other key players to adopt symptom assessment systems as part of their solutions to address global health challenges. The implementation of a standardised benchmarking for these classes of applications as part of the WHO/ITU’s AI for Health Focus Group will therefore be an important step towards addressing this issue.

## Ethical Considerations

### 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
* Explainability
* 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

## Existing AI Solutions

Currently, there are at least three groups that are working on smartphone-based anthropometry and growth tracking tools. They target different age groups and have varying hardware requirements

* A smartphone-based anthropometry tool and visual weighing machine being developed by Wadhwani AI uses regular Android phones with no additional hardware to screen low birthweight newborns. They have conducted field experiments in controlled settings in India
* Child Growth Monitor, an initiative of the German aid organisation Welthungerhilfe, uses an AR-enabled smartphone (i.e., with a. depth sensor) to estimate child anthropometry. Child Growth Monitor has commenced data collection and technology validation efforts in three states in India
* AutoAnthro, an anthropometry tool developed by Body Surface Translations with a grant from the Bill & Melinda Gates Foundation currently requires a special infrared camera to be used with newer generations of the iPhone or iPad. AutoAnthro is currently being tested in Guatemala and Kenya.

All three are in the R&D phase and have not made performance data public.

## Existing work on benchmarking

There is no existing work on benchmarking AI-based solutions for growth monitoring. This group will have to start from scratch.

# AI4H Topic group

### Objectives & general mandate

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.

### Structure:

* The group will have a speaker, who is elected for 1 year by the group members. The speaker coordinates the group’s activities. A wider board may be implemented if needed.
* The group will consist of subgroups along various AI use case areas (e.g., pregnancy risk estimation and stratification, early warning systems, etc)
* The first subgroup to be established is AI for growth monitoring
* Further subgroups are subject to contributors, and will be agreed to by the group
* The group might seek financial support from sponsors to fund its activities and support it. Funding will be made transparent and must be unrelated to the group’s activities

### Participation

* The group will be open for participation for everyone.

### Tools/process of cooperation

* To be established (collaborative document repository, website, blogs, slack groups etc.)

### Interaction

* Group members will regularly interact via email.
* A regular quarterly report will be provided by email by the speaker to the group’s members.
* The group will convene in irregular intervals according to availability.

### Current status

* The group is currently being founded.

### Contributors

* Raghu Dharmaraju, Wadhwani Institute for Artificial Intelligence (Wadhwani AI)
* Hafsa Mwita, University of Zanzibar.

### Meetings

* 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

## AI Input Data Structure

* TBD
  + possible inputs for benchmarking: Labeled images / videos of children with ground truth measurements and other information, longitudinal datasets with images / videos and measurements tracked over time
  + ontologies, terminologies
  + data format

## AI Output Data Structure

* TBD
  + Possible outputs to benchmark: classification, confusion matrix, measurement value, percentile scores / z-scores, % change, etc.
  + ontologies, terminologies
  + data format

## Test Data Labels

* TBD
  + label types
  + ontologies, terminologies
  + data format

## Scores & Metrics

* Potential metrics:
  + Accuracy
  + Precision & recall / sensitivity & specificity
  + AUC
    - ROC (Receiver operating characteristic curve)
    - PRC (Precision Recall curve)
  + Absolute error
* Open questions:
  + considering relation to parameters stakeholders need for decision making
  + considering scores that providers use
  + considering the scope providers designed their solutions for
  + considering the state of the art in RCT, statistics, AI benchmarking etc.
  + considering bias transparency

## Undisclosed Test Data Set Collection

* 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

* 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

* 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

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