|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | INTERNATIONAL TELECOMMUNICATION UNION  **TELECOMMUNICATION STANDARDIZATION SECTOR**  STUDY PERIOD 2022-2024 | | FG-AI4H-Q-015-A01 | |
| ITU-T Focus Group on AI for Health | |
| Original: English | |
| **WG(s):** | | Plenary | Douala, 6-9 December 2022 | |
| **DOCUMENT** | | | | |
| **Source:** | | TG-MCH Topic Driver | | |
| **Title:** | | Att.1 – TDD update (TG-MCH) | | |
| **Purpose:** | | Discussion | | |
| **Contact:** | | Alexandre Chiavegatto Filho TG-MCH University of Sao Paulo, Brazil | | Email: [alexdiasporto@usp.br](mailto:alexdiasporto@usp.br) |
| **Contact:** | | Raghu Dharmaraju TG-MCH Wadhwani AI, India | | Email: [raghu@wadhwaniai.org](mailto:raghu@wadhwaniai.org) |
| **Contact:** | | Hafsa Mwita TG-MCH University of Zanzibar, Tanzania | | Email: [hafsa.m.mwita@gmail.com](mailto:hafsa.m.mwita@gmail.com) |

|  |  |
| --- | --- |
| **Abstract:** | This topic description document (TDD) aims to specify a standardized benchmarking for AI-based maternal and child health. It covers all scientific, technical, and administrative aspects relevant for setting up this benchmarking (and follows the template structure defined in document FGAI4H-J-105). The creation of this TDD is an ongoing iterative process until it is approved by the Focus Group on AI for Health (FG-AI4H) as deliverable No. 10.07. This draft will be a continuous input- and output document. |

**CONTENTS**

| Page |
| --- |
| [1 Introduction 6](#_Toc55962717)  [2 About the FG-AI4H topic group on MCH 6](#_Toc55962718)  [2.1 Documentation 6](#_Toc55962719)  [2.2 Status of this topic group 7](#_Toc55962720)  [2.2.1 Status update for meeting L 7](#_Toc55962721)  [2.2.2 Status update for meeting L 7](#_Toc55962722)  [2.3 Topic group participation 7](#_Toc55962723)  [3 Topic description 8](#_Toc55962724)  [3.1 Maternal and child halth 8](#_Toc55962725)  [3.1.1 Definition of the AI task 8](#_Toc55962726)  [3.1.2 Current gold standard 9](#_Toc55962727)  [3.1.3 Relevance and impact of an AI solution 9](#_Toc55962728)  [3.1.4 Existing AI solutions 9](#_Toc55962729)  [4 Ethical considerations 9](#_Toc55962731)  [5 Existing work on benchmarking 10](#_Toc55962732)  [5.1 Maternal and Child Health 10](#_Toc55962733)  [5.1.1 Publications on benchmarking systems 10](#_Toc55962734)  [5.1.2 Benchmarking by AI developers 11](#_Toc55962735)  [5.1.3 Relevant existing benchmarking frameworks 11](#_Toc55962736)  [6 Benchmarking by the topic group 11](#_Toc55962738)  [6.1 Maternal and Child Health 12](#_Toc55962739)  [6.1.1 Benchmarking version [Y] 12](#_Toc55962740)  [6.1.2 Benchmarking version [X] 21](#_Toc55962741)  [7 Overall discussion of the benchmarking 22](#_Toc55962743)  [8 Regulatory considerations 22](#_Toc55962744)  [8.1 Existing applicable regulatory frameworks 23](#_Toc55962745)  [8.2 Regulatory features to be reported by benchmarking participants 23](#_Toc55962746)  [8.3 Regulatory requirements for the benchmarking systems 23](#_Toc55962747)  [8.4 Regulatory approach for the topic group 23](#_Toc55962748)  [9 References 24](#_Toc55962749)  [Annex A: Glossary 24](#_Toc55962750)  [Annex B: Declaration of conflict of interests 26](#_Toc55962751) |

**List of Tables**

| Page |
| --- |
| [Table 1: Topic group output documents 7](#_Toc55950345) |

**List of Figures**

| Page |
| --- |
| [Figure 1: Example of a figure 25](#_Toc55950354) |

FG-AI4H Topic Description Document

Topic group-MCH

# 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. Despite notable recent improvements for most countries, the Millenium Development Goal (MDG) target for 2015 of reducing child mortality globally by two thirds was not achieved.

Artificial intelligence methods have the potential of improving maternal and child health decisions, especially in low-resource settings. As advances are made in the collection and availability of data on maternal and child health, the possibility of using this data to improve health decisions increases, especially when access to specialized professionals is scarce. Several challenges in the maternal and child area can be overcome with adequate preventive methods, which in turn depend on the establishment of risk scores for the development of targeted public policies, especially in low income countries where the available resources for these policies are lower.

The aim of this document therefore is to develop a standardised benchmarking approach for AI for maternal and child health, with a focus on developing countries. This topic description document specifies the standardized benchmarking for maternal and child health (MCH) systems. It serves as deliverable No. 10.07 of the ITU/WHO Focus Group on AI for Health (FG-AI4H).

# About the FG-AI4H topic group on Maternal and Child Health

The introduction highlights the potential of a standardized benchmarking of AI systems for maternal and child health to help solving important health issues and provide decision-makers with the necessary insight to successfully address these challenges.

To develop this benchmarking framework, FG-AI4H decided to create the TGMCH at the meeting Brasília, Brazil.

FG-AI4H assigns a *topic driver* to each topic group (similar to a moderator) who coordinates the collaboration of all topic group members on the TDD.Alexandre Chiavegatto Filho from the Department of Epidemiology of the University of São Paulo, Brazil, was nominated as topic driver for the TG-MCH.

## Documentation

This document is the TDD for the TG-MCH. It introduces the health topic including the AI task, outlines its relevance and the potential impact that the benchmarking will have on the health system and patient outcome, and provides an overview of the existing AI solutions for maternal and child health (MCH). It describes the existing approaches for assessing the quality of maternal and child health systems and provides the details that are likely relevant for setting up a new standardized benchmarking. It specifies the actual benchmarking methods for all subtopics at a level of detail that includes technological and operational implementation. There are individual subsections for all versions of the benchmarking. Finally, it summarizes the results of the topic group’s benchmarking initiative and benchmarking runs. In addition, the TDD addresses ethical and regulatory aspects.

The TDD will be developed cooperatively by all members of the topic group over time and updated TDD iterations are expected to be presented at each FG-AI4H meeting.

The final version of this TDD will be released as deliverable “DEL 10.07 Maternal and Child Health (TG-MCH).” The topic group is expected to submit input documents reflecting updates to the work on this deliverable **(Table 1)** to each FG-AI4H meeting.

Table 1: Topic group output documents

| Number | Title |
| --- | --- |
| FGAI4H-L-07-A01 | Latest update of the Topic Description Document of the TG-MCH |
| FGAI4H-L-07-A02 | Latest update of the Call for Topic Group Participation (CfTGP) |
| FGAI4H-L-0y-A03 | The presentation summarizing the latest update of the Topic Description Document of the TG-MCH |

## Status of this topic group

The following subsections describe the update of the collaboration within the TG-MCH for the official focus group meetings.

### Status update for meeting

The topic group had two online meetings during the 2021-2022 cycle. Due to the serious consequences brought about by the COVID-19 pandemic in the countries of the topic leaders (India and Brazil), there was a difficulty in holding more frequent meetings during this period of worsening of the crisis. However, online meetings served to guide forward the topic group and the establishment of priorities for the next documents. From the last meeting held in 2020, professor Alexandre Chiavegatto Filho from the University of São Paulo, Brazil, assumed the responsibility of writing this document according to the current standardized format. The next steps of this working group include the addition of new international partners and the completion of some of the topics still under construction.

## Topic group participation

The participation in both, the Focus Group on AI for Health and in a TG is generally open to anyone (with a free ITU account). For this TG, the corresponding ‘Call for TG participation’ (CfTGP) can be found here:

<https://www.itu.int/en/ITU-T/focusgroups/ai4h/Documents/tg/CfP-TG-MCH.pdf>

Each topic group also has a corresponding subpage on the ITU collaboration site. The subpage for this topic group can be found here:

<https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/tg/SitePages/TG-MCH.aspx>

For participation in this topic group, interested parties can also join the regular online meetings. For all TGs, the link will be the standard ITU-TG ‘zoom’ link:

<https://itu.zoom.us/my/fgai4h>

All relevant administrative information about FG-AI4H—like upcoming meetings or document deadlines—will be announced via the general FG-AI4H mailing list [fgai4h@lists.itu.int](mailto:fgai4h@lists.itu.int).

All TG members should subscribe to this mailing list as part of the registration process for their ITU user account by following the instructions in the ‘Call for Topic Group participation’ and this link:

* <https://itu.int/go/fgai4h/join>

Regular FG-AI4H workshops and meetings proceed about every two months at changing locations around the globe or remotely. More information can be found on the official FG-AI4H website:

* <https://itu.int/go/fgai4h>

# 

# Topic description

This section contains a detailed description and background information of the specific health topic for the benchmarking of AI in maternal and child health and how this can help to solve a relevant ‘real-world’ problem.

Topic groups summarize related benchmarking AI subjects to reduce redundancy, leverage synergies, and streamline FG-AI4H meetings. However, in some cases different subtopic groups can be established within one topic group to pursue different topic-specific fields of expertise. The TG-MCH currently has no subtopics. Future subtopics for MCH might be introduced.

## Maternal and Child Health

### Definition of the AI task

This section provides a detailed description of the specific task the AI systems of this TG are expected to solve. It is *not* about the benchmarking process (this will be discussed more detailed in chapter 4). This section corresponds to [DEL03](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7B7997F2C1-5A1D-4409-B2A0-CBC4E9CE8CDA%7D&file=DEL03.docx&action=default) *“AI requirements specifications*,” which describes the functional, behavioural, and operational aspects of an AI system.

Improving the health and well-being of mothers, infants, and children is one of the most important public health goals worldwide, and progress has been slower than expected. From 2018 to 2030, there will be an estimated 27.8 million worldwide deaths in the first month of life in case every country maintains their current rate of reduction. 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.

Almost half of under-five deaths will be of newborns whose deaths could be effectively 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. Every year, around 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:

i. Predictions during pregnancy:

a. Fetal growth prediction.

b. Final gestational age prediction.

c. Incidence of relevant comorbities during pregnancy.

d. Mortality risk:

d1. Fetal mortality.

d2. Maternal mortality.

d3. Neonatal mortality.

d4. Infant mortality.

ii. Hospital warning systems for the necessity of:

1. Labour rooms.
2. Neonatal intensive care units.
3. Healthcare emergency specialists.

iii. Patient-centric health screening tools: IOT tools to screen for the risk of common diseases such as pneumonia, jaundice, anaemia, etc.

iv. Post-natal predictions:

b. Risk of neonatal and postneonatal mortality given the characteristics of the newborn.

c. Risk of rehospitalization after childbirth discharge.

In developing countries, the burden of delivery of health services pertaining to maternal and child is the responsibility of frontline health workers, who in some cases 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, automatic, accurate and tamper-proof screening.

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.

### Current gold standard

While there are already several research and commercial groups working on AI applications in maternal and child health, the lack of consistent standardization makes it difficult for organisations to adopt 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.

In recent years there has been a strong growth in AI solutions for maternal and child health issues. However, it is important to consider some growing and important challenges for the area. First, rigorous validation of these techniques is required to ensure that issues commonly present in machine learning projects are not present, such as data leakage. Second, it is important to ensure that all the data is available for AI solutions to be applied where they are most needed, i.e. in developing countries that have less data collection and a still incipient use of electronic medical records.

### Relevance and impact of an AI solution

This section addresses the relevance and impact of the AI solution (e.g., on the health system or the patient outcome) and describes how solving the task with AI improves a health issue.

Avoiding the occurrence of negative child outcomes often involves low-cost interventions, but which require an alert early enough to prevent its occurrence. Predictive machine learning algorithms can assist in this task by analyzing the patient's maternal and child characteristics and providing an early risk score for the occurrence of the adverse effect.

### Existing AI solutions

This section provides an overview of existing AI solutions for the same health topic that are already in operation. It should contain details of the operations, limitations, robustness, and the scope of the available AI solutions. The details on performance and existing benchmarking procedures will be covered in chapter 6.

A few scientific studies have been performed in recent years that apply machine learning algorithms to predict maternal and child health events. A 2019 study used machine learning to predict postpartum hospital admission in the first 12 weeks after delivery found a high predictive performance for hospitalization from hypertensive disorders (AUC = 0.879) (Betts et al., 2019). Another analysis from 2020 found that machine learning was able to predict height-for-age z-scores in children from a rural area of Pakistan (Harrison et al., 2020). A more recent study from 2021 fond that machine learning algorithms were able to predict with reasonable accuracy the risk of readmission for complications of hypertensive disorders of pregnancy (Hoffman et al., 2021).

Real-time fetal electrocardiogram recordings can be used by patients and clinicians to monitor fetal status. AI-based sensors can be used to monitor blood glucose and blood pressure, which are especially useful in low-resource settings. This has become even more important since the COVID-19 pandemic, where healthcare professionals are overwhelmed by the huge wave of patients because of which patients with pre-existing conditions are not being able to seek help and timely follow-up (Oprescu et al., 2020). This is also related to one of the most promising areas of AI applications, namely mobile health (mHealth). mHealth is extremely useful for prenatal care, especially in low-resource settings, where community health workers can facilitate care, monitor health, and enable patient self-management. Even in high-resource settings, mHealth allows for personalized monitoring to support pregnant women (Davidson et al., 2021).

# Ethical considerations

The rapidly evolving field of AI and digital technology in the fields of medicine and public health raises a number of ethical, legal, and social concerns that have to be considered in this context. They are discussed in deliverable [DEL01](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7B0505B020-362C-45B2-94BF-215D2EBBD8F5%7D&file=DEL01.docx&action=default) “*AI4H ethics considerations,”* which was developed by the working group on “Ethical considerations on AI4H” (WG-Ethics). This section refers to DEL01 and should reflect the ethical considerations of the TG-MCH.

Topics to consider:

* Technical robustness of the algorithms according to:
  + Racial differences.
  + Dataset shifts.
  + Patient income.
* Overall predictive performance of the algorithms.
* Differences in predictive performance according to vulnerable subgroups.
* Availability of data for training the algorithm in low and middle-income countries in terms of:
  + Data-collection quality.
  + Enough variables to perform an accurate prediction.
* Data governance (storage, access and security) and privacy.
* Bias and fairness of training datasets.
* Generalization ability:
  + From large urban areas to remote rural areas.
  + From high-income to low-income areas.
  + According to patient’s characteristics.
  + According to differences in local clinical protocols.
* Explainability.
* Accountability.

# Existing work on benchmarking

This section focuses on the existing benchmarking processes in the context of AI and maternal and child health for quality assessment. It addresses different aspects of the existing work on benchmarking of AI systems (e.g., relevant scientific publications, benchmarking frameworks, scores and metrics, and clinical evaluation attempts). The goal is to collect all relevant learnings from previous benchmarking that could help to implement the benchmarking process in this topic group.

## Maternal and child health

### Publications on benchmarking systems

While a representative comparable benchmarking for maternal and child health does not yet exist, some work has been done in the scientific community assessing the performance of such systems. This section summarizes insights from the most relevant publications on this topic. It covers parts of the deliverable DEL07 *“AI for health evaluation considerations,”* [DEL07\_1](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7B565EEC0A-D755-41C8-AC68-37B4C38C953F%7D&file=DEL07_1.docx&action=default) *“AI4H evaluation process description,”* [DEL07\_2](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7B58679341-C738-40F0-A822-3AC2B24DD09F%7D&file=DEL07_2.docx&action=default) *“AI technical test specification*,*”* [DEL07\_3](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7BA3088882-F82B-493B-B1C5-49CFF0EEEFA8%7D&file=DEL07_3.docx&action=default) *“Data and artificial intelligence assessment methods (DAISAM),”* and [DEL07\_4](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7BB846B260-373A-41FC-A892-EE5BBCFE3CF8%7D&file=DEL07_4.docx&action=default) *“Clinical Evaluation of AI for health”*.

### Benchmarking by AI developers

All developers of AI solutions for maternal and child health implemented internal benchmarking systems for assessing the performance. This section will outline the insights and learnings from this work of relevance for benchmarking in this topic group.

### Relevant existing benchmarking frameworks

Triggered by the hype around AI, recent years have seen the development of a variety of benchmarking platforms where AIs can compete for the best performance on a determined dataset. Given the high complexity of implementing a new benchmarking platform, the preferred solution is to use an established one. This section reflects on the different existing options that are relevant for this topic group and includes considerations of using the assessment platform that is currently developed by FG-AI4H and presented by deliverable [DEL07\_5](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7B8BFCFF21-3908-4BAD-AB9C-9814EB3F9B36%7D&file=DEL07_5.docx&action=default) *“FG-AI4H assessment platform”* (the deliverable explores options for implementing an assessment platform that can be used to evaluate AI for health for the different topic groups).

# Benchmarking by the topic group

This section describes all technical and operational details regarding the benchmarking process for the maternal and child health AI task including subsections for each version of the benchmarking that is iteratively improved over time.

It reflects the considerations of various deliverables: [DEL05](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7B2012357A-941E-44BD-B965-370D7829F52C%7D&file=DEL05.docx&action=default) *“Data specification”* (introduction to deliverables 5.1-5.6), [DEL05\_1](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7B19830259-F63B-42D4-A408-48C854D6C124%7D&file=DEL05_1.docx&action=default)*“Data requirements”* (which lists acceptance criteria for data submitted to FG-AI4H and states the governing principles and rules), [DEL05\_2](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7B25141F77-E59A-45F1-B081-185C2194FE67%7D&file=DEL05_2.docx&action=default) *“Data acquisition”*, [DEL05\_3](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7B05D8938E-BC2A-4A62-BCB0-1FD46AA72235%7D&file=DEL05_3.docx&action=default) *“Data annotation specification”*, [DEL05\_4](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7BF267A95C-4C5B-4D63-A135-58AF487C3AD3%7D&file=DEL05_4.docx&action=default) *“Training and test data specification”* (which provides a systematic way of preparing technical requirement specifications for datasets used in training and testing of AI models), [DEL05\_5](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7B71FE8B9D-ACB3-48CE-AA3F-136409B550A4%7D&file=DEL05_5.docx&action=default) *“Data handling”* (which outlines how data will be handled once they are accepted), [DEL05\_6](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7B5C95327E-96A5-4175-999E-3EDB3ED147C3%7D&file=DEL05_6.docx&action=default) *“Data sharing practices”* (which provides an overview of the existing best practices for sharing health-related data based on distributed and federated environments, including the requirement to enable secure data sharing and addressing issues of data governance), [DEL06](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7BF5967277-90C8-4252-A0B9-43A5692F35E2%7D&file=DEL06.docx&action=default) *“AI training best practices specification”* (which reviews best practices for proper AI model training and guidelines for model reporting), [DEL07](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7B47E77197-F87B-49F4-80B3-2DD949A5F185%7D&file=DEL07.docx&action=default)*“AI for health evaluation considerations”* (which discusses the validation and evaluation of AI for health models, and considers requirements for a benchmarking platform), [DEL07\_1](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7B565EEC0A-D755-41C8-AC68-37B4C38C953F%7D&file=DEL07_1.docx&action=default) *“AI4H evaluation process description”* (which provides an overview of the state of the art of AI evaluation principles and methods and serves as an initiator for the evaluation process of AI for health), [DEL07\_2](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7B58679341-C738-40F0-A822-3AC2B24DD09F%7D&file=DEL07_2.docx&action=default) *“AI technical test specification”* (which specifies how an AI can and should be tested *in silico*), [DEL07\_3](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7BA3088882-F82B-493B-B1C5-49CFF0EEEFA8%7D&file=DEL07_3.docx&action=default) *“Data and artificial intelligence assessment methods (DAISAM)”* (which provides the reference collection of WG-DAISAM on assessment methods of data and AI quality evaluation), [DEL07\_4](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7BB846B260-373A-41FC-A892-EE5BBCFE3CF8%7D&file=DEL07_4.docx&action=default)*“Clinical Evaluation of AI for health”* (which outlines the current best practices and outstanding issues related to clinical evaluation of AI models for health), [DEL07\_5](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7B8BFCFF21-3908-4BAD-AB9C-9814EB3F9B36%7D&file=DEL07_5.docx&action=default) *“FG-AI4H assessment platform”* (which explores assessment platform options that can be used to evaluate AI for health for the different topic groups), [DEL09](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7B3E940987-8D75-44B8-85E4-F0E475964F15%7D&file=DEL09.docx&action=default) *“AI for health applications and platforms”* (which introduces specific considerations of the benchmarking of mobile- and cloud-based AI applications in health), [DEL09\_1](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7B1A2EC8D5-53CA-4C8C-9B09-B61CA6F428C5%7D&file=DEL09_1.docx&action=default) *“Mobile based AI applications,”* and [DEL09\_2](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7B3B5A31DE-D3B1-4EC1-A261-2C2E19F73810%7D&file=DEL09_2.docx&action=default) *“Cloud-based AI applications”* (which describe specific requirements for the development, testing and benchmarking of mobile- and cloud-based AI applications).

## Subtopic

The benchmarking of maternal and child health is going to be developed and improved continuously to reflect new features of AI systems or changed requirements for benchmarking. This section outlines all benchmarking versions that have been implemented thus far and the rationale behind them. It serves as an introduction to the subsequent sections, where the actual benchmarking methodology for each version will be described.

### Benchmarking version 1.

This section includes all technological and operational details of the benchmarking process for the benchmarking version 1.

#### Overview

This section provides an overview of the key aspects of this benchmarking iteration, version 1.

#### Benchmarking methods

This section provides details about the methods of the benchmarking version 1. It contains detailed information about the benchmarking system architecture, the dataflow and the software for the benchmarking process (e.g., test scenarios, data sources, and legalities).

##### Benchmarking system architecture

This section covers the architecture of the benchmarking system. For well-known systems, an overview and reference to the manufacturer of the platform is sufficient. If the platform was developed by the topic group, a more detailed description of the system architecture is required.

##### Benchmarking system dataflow

This section describes the dataflow throughout the benchmarking architecture.

##### Safe and secure system operation and hosting

This section addresses security considerations about the storage and hosting of data (benchmarking results and reports) and safety precautions for data manipulation, data leakage, or data loss.

In the case of a manufactured data source (vs. self-generated data), it is possible to refer to the manufacturer’s prescriptions.

##### Benchmarking process

This section describes how the benchmarking looks from the registration of participants, through the execution and resolution of conflicts, to the final publication of the results.

#### AI input data structure for the benchmarking

This section describes the input data provided to the AI solutions as part of the benchmarking of maternal and child health. It covers the details of the data format and coding at the level of detail needed to submit an AI for benchmarking. This is the only TDD section addressing this topic. Therefore, the description needs to be complete and precise. This section does *not* contain the encoding of the labels for the expected outcomes. It is only about the data the AI system will see as part of the benchmarking.

#### AI output data structure

Similar to the input data structure for the benchmarking, this section describes the output data the AI systems are expected to generate in response to the input data. It covers the details of the data format, coding, and error handling at the level of detail needed for an AI to participate in the benchmarking.

#### Test data label/annotation structure

While the AI systems can only receive the input data described in the previous sections, the benchmarking system needs to know the expected correct answer (sometimes called ‘labels’) for each element of the input data so that it can compare the expected AI output with the actual one. Since this is only needed for benchmarking, it is encoded separately. The details are described in the following section.

#### Scores and metrics

Scores and metrics are at the core of the benchmarking. This section describes the scores and metrics used to measure the performance, robustness, and general characteristics of the submitted AI systems.

#### Test dataset acquisition

Test dataset acquisition includes a detailed description of the test dataset for the AI model and, in particular, its benchmarking procedure including quality control of the dataset, control mechanisms, data sources, and storage.

#### Data sharing policies

This section provides details about legalities in the context of benchmarking. Each dataset that is shared should be protected by special agreements or contracts that cover, for instance, the data sharing period, patient consent, and update procedure (see also [DEL05\_5](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7B71FE8B9D-ACB3-48CE-AA3F-136409B550A4%7D&file=DEL05_5.docx&action=default) on *data handling* and [DEL05\_6](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7B5C95327E-96A5-4175-999E-3EDB3ED147C3%7D&file=DEL05_6.docx&action=default) on *data sharing practices*).

#### Baseline acquisition

The main purpose of benchmarking is to provide stakeholders with the numbers they need to decide whether AI models provide a viable solution for a given health problem in a designated context. To achieve this, the performance of the AI models needs to be compared with available options achieving the same clinically meaningful endpoint. This, in turn, requires data on the performance of the alternatives, ideally using the same benchmarking data. As the current alternatives typically involve doctors, it might make sense to combine the test data acquisition and labelling with additional tasks that allow the performance of the different types of health workers to be assessed.

#### Reporting methodology

This section discusses how the results of the benchmarking runs will be shared with the participants, stakeholders, and general public.

#### Result

This section gives an overview of the results from runs of this benchmarking version of your topic. Even if your topic group prefers an interactive drill-down rather than a leader board, pick some context of common interest to give some examples.

#### Discussion of the benchmarking

This section discusses insights of this benchmarking iterations and provides details about the ‘outcome’ of the benchmarking process (e.g., giving an overview of the benchmark results and process).

#### Retirement

This section addresses what happens to the AI system and data after the benchmarking activity is completed. It might be desirable to keep the database for traceability and future use. Alternatively, there may be security or privacy reasons for deleting the data. Further details can be found in the reference document of this section [DEL04](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7BC68833D1-9B31-4E8E-8A4A-3939D7DEA56F%7D&file=DEL04.docx&action=default) “*AI software lifecycle specification”* (identification of standards and best practices that are relevant for the AI for health software life cycle).

# Overall discussion of the benchmarking

This section discusses the overall insights gained from benchmarking work in this topic group. This should not be confused with the discussion of the results of a concrete benchmarking run (e.g., in 4.2.11).

# Regulatory considerations

For AI-based technologies in healthcare, regulation is not only crucial to ensure the safety of patients and users, but also to accomplish market acceptance of these devices. This is challenging because there is a lack of universally accepted regulatory policies and guidelines for AI-based medical devices. To ensure that the benchmarking procedures and validation principles of FG-AI4H are secure and relevant for regulators and other stakeholders, the working group on *“*[*Regulatory considerations on AI for health”* *(WG-RC)*](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/wg/SitePages/WG-RC.aspx) compiled the requirements that consider these challenges.

The deliverables with relevance for regulatory considerations are [DEL02](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7BF2F46A99-7457-4BC8-81A3-0E1E63D6072A%7D&file=DEL02.docx&action=default) *“AI4H regulatory considerations”* (which provides an educational overview of some key regulatory considerations), [DEL02\_1](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7B6AF7C004-8BCE-4151-9F44-45F041A1EB1D%7D&file=DEL02_1.docx&action=default) *“Mapping of IMDRF essential principles to AI for health software”,* and[DEL02\_2](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7B1ED0D4D1-876C-4A0F-AEF7-06D3F445F5E6%7D&file=DEL02_2.docx&action=default) *“Guidelines for AI based medical device (AI-MD): Regulatory requirements”* (which provides a checklist to understand expectations of regulators, promotes step-by-step implementation of safety and effectiveness of AI-based medical devices, and compensates for the lack of a harmonized standard). [DEL04](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7BC68833D1-9B31-4E8E-8A4A-3939D7DEA56F%7D&file=DEL04.docx&action=default) identifies standards and best practices that are relevant for the “*AI software lifecycle specification*.*”* The following sections discuss how the different regulatory aspects relate to the TG-MCH.

## Existing applicable regulatory frameworks

Most of the AI systems that are part of the FG-AI4H benchmarking process can be classified as *software as medical device* (SaMD) and eligible for a multitude of regulatory frameworks that are already in place. In addition, these AI systems often process sensitive personal health information that is controlled by another set of regulatory frameworks. The following section summarizes the most important aspects that AI manufacturers need to address if they are developing AI systems for maternal and child health.

## Regulatory features to be reported by benchmarking participants

In most countries, benchmarked AI solutions can only be used legally if they comply with the respective regulatory frameworks for the application context. This section outlines the compliance features and certifications that the benchmarking participants need to provide as part of the metadata. It facilitates a screening of the AI benchmarking results for special requirements (e.g., the prediction of prediabetes in a certain subpopulation in a country compliant to the particular regional regulatory requirements).

## Regulatory requirements for the benchmarking systems

The benchmarking system itself needs to comply with regulatory frameworks (e.g., some regulatory frameworks explicitly require that all tools in the quality management are also implemented with a quality management system in place). This section outlines the regulatory requirements for software used for benchmarking in this topic group.

## Regulatory approach for the topic group

Building on the outlined regulatory requirements, this section describes how the topic group plans to address the relevant points in order to be compliant. The discussion here focuses on the guidance and best practice provided by the [DEL02](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7BF2F46A99-7457-4BC8-81A3-0E1E63D6072A%7D&file=DEL02.docx&action=default) *“AI4H regulatory considerations.”*

# References

Betts KS, Kisely S, Alati R. Predicting common maternal postpartum complications: leveraging health administrative data and machine learning. BJOG 2019;126(6):702-9.

Harrison E, Syed S, Ehsan L, Iqbal NT, Sadiq K, Umrani F, Ahmed S, Rahman N, Jakhro S, Ma JZ, Hughes M, Ali SA. Machine learning model demonstrates stunting at birth and systemic inflammatory biomarkers as predictors of subsequent infant growth - a four-year prospective study. BMC Pediatr 2020;20(1):498.

Hoffman MK, Ma N, Roberts A. A machine learning algorithm for predicting maternal readmission for hypertensive disorders of pregnancy. Am J Obstet Gynecol 2021;3:100250.

Annex A:  
Glossary

This section lists all the relevant abbreviations, acronyms and uncommon terms used in the document.

|  |  |  |
| --- | --- | --- |
| Acronym/Term | Expansion | Comment |
| TDD | Topic Description Document | Document specifying the standardized benchmarking for a topic on which the FG AI4H Topic Group works. This document is the TDD for the Topic Group MCH |
| TG | Topic Group |  |
| WG | Working Group |  |
| FGAI4H | Focus Group on AI for Health |  |
| AI | Artificial intelligence |  |
| ITU | International Telecommunication Union |  |
| WHO | World Health Organization |  |
| DEL | Deliverable |  |
| CfTGP | Call for topic group participation |  |
| AI4H | Artificial intelligence for health |  |
| IMDRF | International Medical Device Regulators Forum |  |
| MDR | Medical Device Regulation |  |
| ISO | International Standardization Organization |  |
| GDPR | General Data Protection Regulation |  |
| FDA | Food and Drug administration |  |
| SaMD | Software as a medical device |  |
| AI-MD | AI based medical device |  |
| LMIC | Low-and middle-income countries |  |
| GDP | Gross domestic product |  |
| API | Application programming interface |  |
| IP | Intellectual property |  |
| PII | Personal identifiable information |  |
| […] |  |  |

Annex B:  
Declaration of conflict of interests

In accordance with the ITU transparency rules, this section lists the conflict-of-interest declarations for everyone who contributed to this document. Please see the guidelines in FGAI4H-F-105 “ToRs for the WG-Experts and call for experts” and the respective forms (Application form & Conflict of interest form).

Text

Description automatically generated

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

Prof. Dr. Alexandre Chiavegatto Filho

School of Public Health

University of Sao Paulo