International Telecommunication Union

ITU-T FG-AI4H Deliverable

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

15 September 2023

PRE-PUBLISHED VERSION

DEL10.6

FG-AI4H Topic Description Document for the Topic Group on malaria (TG-Malaria)

ITU-T

International Telecommunication Union

Summary

This topic description document (TDD) specifies a standardized benchmarking for AI-based malaria detection. It covers all scientific, technical, and administrative aspects relevant for setting up this benchmarking.

Keywords

Artificial intelligence; benchmarking; health; topic groups; overview; ethics; regulations; data quality; data audit; clinical relevance; topic description; malaria detection; microscopy

Change Log

This document contains Version 1 of the Deliverable DEL10.6 on "*FG-AI4H Topic Description Document for the Topic Group on malaria (TG-Malaria)*" approved on 15 September 2023 via the online approval process for the ITU-T Focus Group on AI for Health (FG-AI4H).

Editor: Rose Nakasi
Makerere University
Uganda
E-mail: g.nakasirose@gmail.com

Contributors: (in alphabetical order)

Herilalaina Rokototarison
Université Paris-Saclay
France
Email: heri@lri.fr

CONTENTS

Page

List of Tables

List of Figures

ITU-T FG-AI4H Deliverable 10.6

FG-AI4H Topic Description Document for the Topic Group on malaria (TG-Malaria)

1 Introduction

Malaria is one of the largest endemic diseases in Sub Saharan Africa [5]. In low developed countries (LDCs), the scourge is further buttressed by the lack of enough skilled lab technologists in health centres to detect the disease using the widely accepted gold standard Microscopy method. Thus, the need for reliable detection interventions. This explains the birth of Automated malaria detection using Artificial Intelligence (AI). The aim is to harness AI to automate the detection of malaria in a more fast, accurate and cost-effective manner. Recently AI and machine learning techniques have been successful in different medical image analysis tasks and have a capability to improve public health.

The document therefore aims at developing a standardised benchmarking approach for AI based detection of malaria.

This topic description document specifies the standardized benchmarking for Topic Group Malaria systems. It serves as deliverable No.06 of the ITU/WHO Focus Group on AI for Health (FG-AI4H).

2 About the FG-AI4H topic group on TG-Malaria

The introduction highlights the potential of a standardized benchmarking of AI systems for Topic Group Malaria 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 TG-Malaria at the meeting H in Zanzibar, 3-5 September 2019.

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. During FG-AI4H meeting H in Zanzibar, 3-5 September 2019, Rose Nakasi from Makerere University was nominated as topic driver for the TG-Malaria.

2.1 Documentation

This document is the TDD for the TG-Malaria. 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 TG-Malaria. It describes the existing approaches for assessing the quality of TG-Malaria 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 014. AI based detection of Malaria (TG-Malaria)." The topic group is expected to submit input documents reflecting updates to the work on this deliverable **(Table 1)** to each FG-AI4H meeting.

The final version of this TDD will be released as deliverable "DEL 014. AI based detection of Malaria (TG-Malaria)." 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-S-014-A01	Latest update of the Topic Description Document of the TG-Malaria
FGAI4H-O-014-A02	Latest update of the Call for Topic Group Participation (CfTGP)
FGAI4H-S-014-A03	The presentation summarizing the latest update of the Topic Description Document of the TG-Malaria

2.2 Status of this topic group

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

2.2.1 Status update for meeting [Discussions arising out of e-meetings]

– Discuss updates on benchmarking platform improvements (data, AI models, Interface)

– Discuss technical implementation details that come with improvements

– Develop simple models for testing the updated benchmarking platform

– Platform beta testing

– *Launching the challenge

2.2.2 Status Update [Members]

Response to call for contribution to the TG-Malaria;

– Laura Moro, PhD. Researcher, science & medical writer. Co-founder of AI Scope. AI Scope a non-profit organization working in AI for improved diagnosis (mostly malaria for now) in low-resource settings.

– Dr. Helmi Zakariah. AIME inc.

– Is a cofounder of AIME company and they work primarily in Forecasting Vector-Borne Disease Outbreak by using AI & ML. While their focus is in Dengue and West Nile Virus, they have begun work in Malaria through collaboration with APMEN members in Malaysia.

– Martha Shaka, a researcher at University of Dodoma and leads a team focusing on the automation of malaria diagnosis using deep learning. They are made up of 2 organization with medical and computer science experts. The team has collaboration with local researchers in the field of malaria diagnosis and their next step is on creating ground truth data sets in Tanzania.

– Phil Verstraete. Co-Managing Director, Milan & Associates

– Ana Rivière Cinnamond, Advisor and Pubic Health Expert in disease surveillance and prevention, DMAP under Health Emergency Information & Risk Assessment Department with PAHO/WHO.

– Herilalaina RAKOTOARISON, PhD student of Machine learning from the Université Paris-Saclay). Herilalaina has been pivotal in implementing the benchmarking platform.

– Fetulhak Abdurahman, is a Lecturer in Jimma University of Electrical and Computer Engineering Faculty, Ethiopia.

– William Mangion, Healthcare and Robotics Computer vision consultant, V7 Labs, London.

2.2.3 Status Update [Next steps]

– The experts aim to extend the topic of Malaria detection to all Malaria endemic Countries, while bringing together AI solutions and data from different countries. Next steps for the group can be of different forms:

– The group equally intends to undertake supervision of retraining and retooling of microscopists in endemic countries on AI based detection of Malaria.

– The group further intends to seek for the creation of data centres for annotated data of thick blood smears from the different medical centres in endemic countries. This will help in creation of data repository centre hence access to big data for further research.

– The group also intends to develop robust, reliable and low-cost AI solutions which can be deployed in a real-time environment. To this end, the group aims at providing Malaria parasite detection AI algorithms that can represent the real-time clinical setting. Our platform enables users to access benchmarking solutions and can integrate into their organization.

– The group will also participate at conferences to present the progress of our achievements and publish results of challenge-based benchmarking in reputable conferences and journals.

– Launch of challenge to a wider community.

– Iterate with the Focus Group benchmarking platform.

2.3 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-Malaria.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-Malaria.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.

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

In addition to the general FG-AI4H mailing list, each topic group can create an *individual mailing list:*

– fgai4htgmalaria@lists.itu.int

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

3 Topic description

This section contains a detailed description and background information of the specific health topic for the benchmarking of AI in TG Malaria 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-Malaria currently has no subtopics. Future subtopics for AI based Malaria Surveillance might be introduced.

According to the World Health Organization report of 2016, nearly half of the world population is at risk of malaria [5]. Records from the WHO report of 2015 indicates that in 2015, 212 million cases reported, Malaria accounted for over 480,000 deaths, 90% of which were from Africa, 7% from S.E Asia and 2% from Eastern Mediterranean region [6]. Although there were fewer Malaria cases in 2017 than in 2010 according to the WHO report of 2017, data for the period 2015-2017 highlighted that no significant progress in reducing global Malaria cases was made in this timeframe [7]. Malaria is thus of major concern to public health and therefore the need for early, fast and accurate diagnosis.

The gold standard method for detection of Malaria is microscopy of blood smear slides. Unlike Rapid Diagnostic Tests (RDTs), microscopy supports direct parasite detection and identification and provides monitoring of systemic inflammation and its response to therapy [9]. Detection of malaria requires examination of thin and thick blood smear images through conventional light microscopy. In general, Malaria parasite detection, species identification, and parasitemia determination requires expertise from trained Microscopists (lab technicians).

Effective Malaria control can be achieved by a fast, consistent and accurate diagnosis. However, this requires the expertise of Microscopists to operate the gold standard method of microscopy screening of Malaria. Unfortunately, highly Malaria endemic Countries have very few expert Microscopists to diagnose and interpret the results of the huge numbers of malaria patients.

A nationwide study in Ghana, for example, found 1.72 microscopes per 100,000 population, but only 0.85 trained laboratory technicians per 100,000 population [1] which is grossly inadequate. As a result, diagnoses are often made on the basis of clinical signs and symptoms alone, which are error-prone and leads to higher mortality, drug resistance, and the economic burden of buying unnecessary drugs [2].

Computational Microscopy using Artificial Intelligence technologies which is the backbone of this TDD aims to reduce the need for many human Microscopists by providing a fast, consistent and accurate diagnosis with minimum human intervention. AI models have the capability to learn good representations of image data with reduced turnaround time bridging the gap for lack of enough skilled Microscopists and significantly improving diagnostic performance and reducing health costs associated with patient care and treatment.

3.1 Subtopic on AI based detection of malaria

3.1.1 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 DEL3 *"AI requirements specifications*," which describes the functional, behavioural, and operational aspects of an AI system.

A use case on Artificial Intelligence-based Diagnosis of Malaria is presented here. Malaria is one of the major diseases causing death in sub–Saharan Africa according to WHO report. Part of the reason for endemicity is poor diagnosis at the laboratory level which may lead to misdiagnosis of the disease as well as drug resistance. The burden is further increased because of lack of enough skilled lab

technologists in health facilities to diagnose the disease through the gold standard method of conventional microscopy. However conventional microscopy is subjective and results vary significantly by different Microscopists thus inaccurate and low throughput screening. Therefore, timely and accurate diagnostic interventions are necessary to reduce cases of misdiagnosis, drug residence burden.

Advances in technology help to push forward in the provision of health care facilities in the form of automated diagnosis of diseases, telemedicine, 3D-printing of medical devices, and mobile health. AI-based Detection of Malaria therefore focuses on use of artificial intelligence techniques to detect plasmodium pathogens in blood smear images in a timely and more accurate manner. Here we propose machine learning methods that deal with all aspects related to improving the conventional malaria diagnosis on blood films. Machine learning methodologies learn good representations of data directly from the pixel data thus providing a more reliable, fast and accurate diagnosis helping to provide confidence of a diagnosis to the lab technicians.

3.1.2 Current gold standard

This section provides a description of the established gold standard of the addressed health topic.

Conventional light microscopy remains the gold standard method of diagnosis of malaria. Microscopy is particularly well adapted to low-resource, high disease burden areas, being both simple and versatile. In contrast to alternatives such as rapid diagnostic tests, however, microscopy-based diagnosis does depend on the availability of skilled technicians, of which there is a critical shortage. As a result, diagnoses are often made on the basis of clinical signs and symptoms alone, which is error-prone and leads to higher mortality, drug resistance, and the economic burden of buying unnecessary drugs. There is therefore a need for alternatives which help to provide access to fast and quality diagnosis.

3.1.3 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.

AI based diagnosis of Malaria aims to reduce the need for many human Microscopists by providing a consistent and accurate diagnosis with minimum human intervention. This is because AI algorithms can accurately learn a good representation of data directly from the annotated datasets. Automation presents a significant advantage over a human Microscopists by potentially increasing the speed and accuracy for blood film analysis, reducing the turnaround time, and significantly improving diagnostic performance.

Currently malaria diagnosis is by use of symptoms and signs, Rapid Diagnostic Tests (RDTs) and conventional microscopy which methods are prone to human error, slow and lack specificity details and thus accuracy is based on human judgment which is usually biased. Therefore, the need to benchmark AI algorithms for malaria diagnosis. Automated AI based microscopy for malaria diagnosis maintains the benefits of manual microscopy (gold standard) by incorporating them in a machine vision platform which helps to provide the access to fast and quality diagnosis that is currently routinely unavailable.

In principle, new benchmarking AI Methods should focus on providing a robust, fast, low cost and more accurate malaria diagnosis that reduces biases through capturing all the necessary implementation parameters that provide a good representation of a dataset, and implementation platform.

The added advantage is that the solution can be applied to any microscopical assessment and in different implementation environments.

3.1.4 Definition of AI Tasks

Microscopy of malaria diagnosis is a diagnostic procedure in which microscopy is used to view images of blood smears for eventual examination by a microscopist.

In general, the AI task with mobile microscopy of malaria is mainly divided into classification, detection, and segmentation.

3.1.4.1 Classification tasks

Classification is a machine learning task for determining which classes are in an image, video or other types of data. It refers to training machine learning models with the intent of finding out which classes are present.

In clinical applications, it is possible to classify positive patches (those containing malaria parasites) and negative patches (without malaria parasites) in thick blood smear microscopic images. In thin blood smear images, parasitised and non-parasitised cells can be classified.

3.1.4.2 Detection

Object detection combines classification and localization to determine what objects are in the image and specify where they are in the image. Generally, bounding boxes are used to distinct objects in images. In clinical applications, it is possible to detect different findings in microscopy images for different object detection tasks, such as trophozoites, gametocytes, schizoites, White Blood Cells, Red Blood Cells. Specifically, trophozoites detection is the most usual Al application in microscopy for parasite identification. Accurate detection of trophozoites can effectively reduce the malaria parasite detection false positives and false negatives for effective malaria diagnosis.

3.1.4.3 Segmentation

Image segmentation separates an image into regions on pixel level, with particular shape and border, delineating potentially meaningful areas for further processing, such as measurement, classification and object detection. The regions may not take up the entire image, but the goal of segmentation is to highlight foreground elements and make it easier to be evaluated. Image segmentation provides pixel-by-pixel details of an object, distinguishing it from classification and object detection. WBC and Red Blood Cells especially in thin blood smear images can be segmented.

3.1.5 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. At the AI and Data Science lab of Makerere University, we have deployed both traditional machine learning and deep learning algorithms for pathogen detection in thick blood smear samples and improvements in detection accuracies have been registered. We have also extended this to other related microscopy diagnosis challenges for example in the detection of tuberculosis and intestinal parasites [4].

An extensive study by Rosado et al [3] reviewed the various image processing and analysis approaches for the automated detection of Malaria with the conclusion that improvements in accuracy are still needed. Some AI tools tend to fail on the undisclosed data sets due to false alarms. There is currently no certified AI based solution for Malaria diagnosis. A major factor contributing to this is the lack of availability of a bigger and diverse standardised dataset from which to infer and draw comparison from the different AI Solutions. Existing AI solutions have also focussed on single detection goals rather than learning complex relationships between different datasets that could provide a more representative diagnosis approach for better realistic results.

There is thus need to collect a large sample dataset that captures different settings of the image to create a wide array of data complexities that depict real life implementations. Datasets such as demographics, environment and any other contributory factor to Malaria prevalence could be captured to assure a more dependable analysis.

4 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 DEL1 "*AI4H ethics considerations,"* which was developed by the working group on "Ethical considerations on AI4H" (WG-Ethics). This section refers to DEL1 and should reflect the ethical considerations of the TG-Malaria.

Collecting massive data is necessary for AI solution development. However, ethical considerations such as patient privacy concerns should be taken into careful consideration and relevant regulations should be followed. Otherwise, the privacy of patients must be protected in the process of data collection, transmission, and utility. If the data contains patient private information or identified codes, data desensitization must be performed. Generally, it is better for data sources, such as hospitals and other clinical institutions, to be responsible for handling the ethical, legal and privacy.

The following procedures is executed in our practice and recommended to other practice of AI for Microscopy diagnosis of malaria;

– Acquisition of the Institutional Review Board approval from the medical regulatory body
– Patients consent procedure at each individual institution.
– Review of the data collection plan by a local medical ethics committee or an institutional review board.
– Anonymization of the image datasets (including demographic information) by clinical institution prior to sending to AI developer.
– Anonymization of the image datasets (including demographic information) by AI developer prior to utility (optional).

5 Existing work on benchmarking

This section focuses on the existing benchmarking processes in the context of AI and TG-Malaria 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.

5.1 Subtopic [AI based detection of malaria]

5.1.1 Publications on benchmarking systems

While a representative comparable benchmarking for TG-Malaria 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 DEL7 *"AI for health evaluation considerations,"* DEL7.1 *"AI4H evaluation process description,"* DEL7.2 *"AI technical test specification,"* DEL7.3 *"Data and artificial intelligence assessment methods (DAISAM),"* and DEL7.4 *"Clinical Evaluation of AI for health"*.

Although research on AI for malaria detection application is increasing rapidly in the recent past, public accessible dataset and benchmarking system does not exist. Several review papers have been published to summarize latest research in the field, but none of those can provide comparable benchmarking for different studies. It is therefore recommended that the research community

establishes a public accessible digital microscopy image database and benchmarking system to advance AI-based detection of malaria research.

5.1.2 Benchmarking by AI developers

All developers of AI solutions for TG-Malaria 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.

Putting into consideration the different AI solutions in respect to image analysis including (detection, classification, segmentation etc.), different metrics will be used in order to enable performance comparison. These metrics are not much different from those used in medical image analysis and computer vision such as mean average precision (mAP), intersection over union (IoU), Dice coefficient (DICE), positive predictive value (PPV), precision, recall, specificity, F1-measure, and Area Under Curve (AUC) of Receiver Operating Characteristic (ROC) curve.

5.1.3 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 DEL7.5 *"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). The Topic Group AI based detection of malaria has also built evaluation assessment platforms based on CodaLab and Codabench to specifically evaluate AI solutions for detection of malaria.

5.2 Subtopic [AI-based surveillance of malaria]

For further study.

6 Benchmarking by the topic group

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

It reflects the considerations of various deliverables: DEL5 *"Data specification"* (introduction to deliverables 5.1-5.6), DEL5.1 *"Data requirements"* (which lists acceptance criteria for data submitted to FG-AI4H and states the governing principles and rules), DEL5.2 *"Data acquisition"*, DEL5.3 *"Data annotation specification"*, DEL5.4 *"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), DEL5.5 *"Data handling"* (which outlines how data will be handled once they are accepted), DEL5.6 *"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), DEL6 *"AI training best practices specification"* (which reviews best practices for proper AI model training and guidelines for model reporting), DEL7 *"AI for health evaluation considerations"* (which discusses the validation and evaluation of AI for health models, and considers requirements for a benchmarking platform), DEL7.1 *"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), DEL7.2 *"AI technical test specification"* (which specifies how an AI can and should be tested *in silico*), DEL7.3 *"Data and artificial intelligence assessment methods (DAISAM)"* (which provides the reference collection of WG-DAISAM on assessment methods of data and AI quality

evaluation), DEL7.4 *"Clinical Evaluation of AI for health"* (which outlines the current best practices and outstanding issues related to clinical evaluation of AI models for health), DEL7.5 *"FG-AI4H assessment platform"* (which explores assessment platform options that can be used to evaluate AI for health for the different topic groups), DEL9 *"AI for health applications and platforms"* (which introduces specific considerations of the benchmarking of mobile- and cloud-based AI applications in health), DEL9.1 *"Mobile based AI applications,"* and DEL9.2 *"Cloud-based AI applications"* (which describe specific requirements for the development, testing and benchmarking of mobile- and cloud-based AI applications).

6.1 Subtopic [AI based detection of malaria]

The benchmarking of TG-Malaria 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.

6.2 Benchmarking versions

This section includes all technological and operational details of the benchmarking process for the latest benchmarking versions.

6.2.1 Overview

This section provides an overview of the key aspects of this benchmarking iterations.

6.2.1.1 Benchmarking methods

This section provides details about the methods of the different benchmarking versions. 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).

The benchmarking method will consider all aspects of Input data requirements, how data will be annotated and annotation formats, AI analysis engine requirements, output and test data formats and scoring metric requirements.

Blood smear Images of both thick and thin blood smear slides that have been annotated by laboratory experts from different Health facilities in different Malaria endemic countries would be required and an undisclosed test data for evaluation of the tool.

– The labels will depend on the specific attribute to be investigated.

– All data will be subject to permissions from the different country authorities.

6.2.1.2 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.

6.2.1.3 Technical architecture

The general benchmarking architecture (see figure 1) advanced by FGAI4H [8] will provide guidance to TG Malaria in the development of the benchmarking platform.

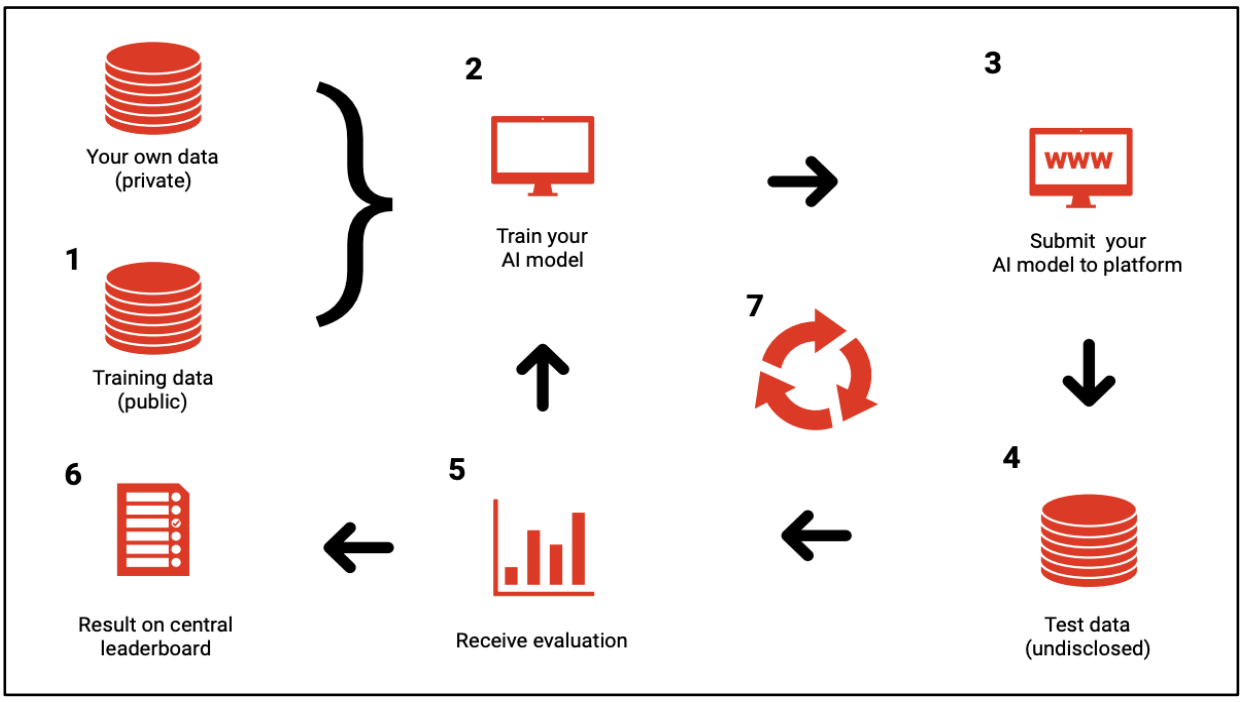

Figure 1 –General Benchmarking pipeline framework for implementation of AI based health solution [8].

6.2.2 TG Malaria Benchmarking methodology and architecture Version 1(V1)

To implement our first benchmarking task on detection of malaria in thick blood smear images, the benchmarking platform used is Codalab. It is an open-source framework designed for enhancing reproducibility of machine learning algorithms. We adopt this for benchmarking malaria detection modeling (see figure 2).

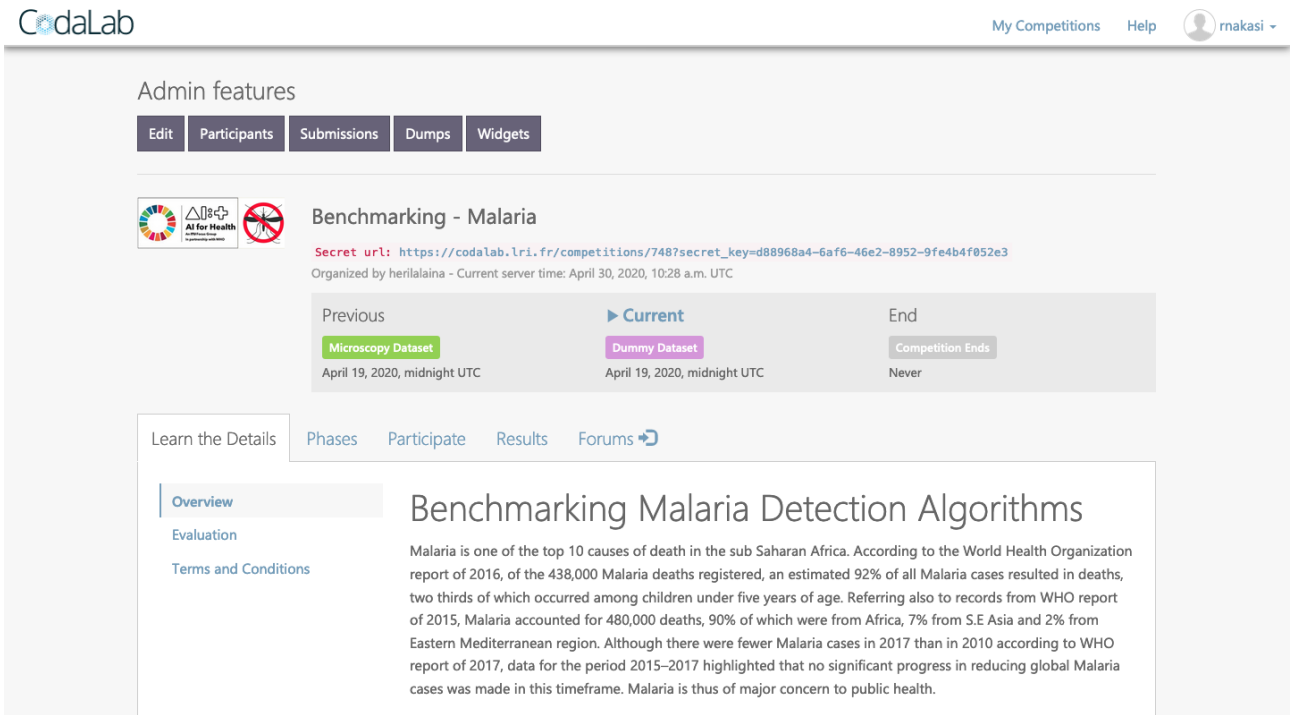

Figure 2 –Benchmarking-Malaria platform implemented using Codalab

The overall process of benchmarking is handled on the server side. Codalab allows organizers to define:

1) **How submission files will be handled, processed and scored;**
The benchmarking system in its current state has a prototype dataset stored at the site of the benchmarking system. Participants are required to use the available dataset send in their AI model by fine tuning a variant sample code on the leaderboard. A submission fails once it doesn't meet the submission criteria defined. At the organizers' site(s), derived detection accuracies of different models are shown (see figure 3).

#	SUBMITTED	SUBMITTED BY	SUBMISSION ID	FILENAME	STATUS	LEADERBOARD	RESULTS					
1	April 20, 2020, 7:58 p.m.	herilalaina	13249	sample_code_submission.zip	Failed	False	---	+	DEL	SHOW	FAILED	RE-RUN
2	April 20, 2020, 7:59 p.m.	herilalaina	13250	sample_code_submission.zip	Finished	True	0.2784	+	DEL	HIDE	FAILED	RE-RUN
3	April 24, 2020, 1:32 p.m.	pavao	13263	sample_code_submission.zip	Finished	False	0.2784	+	DEL	SHOW	FAILED	RE-RUN
4	April 24, 2020, 1:33 p.m.	pavao	13264	sample_code_submission(1).zip	Finished	False	0.2784	+	DEL	SHOW	FAILED	RE-RUN
5	April 24, 2020, 2:33 p.m.	rnakasi	13267	code_submission.zip	Finished	True	0.0599	+	DEL	HIDE	FAILED	RE-RUN
6	April 24, 2020, 3:50 p.m.	pavao	13270	model_with_error.zip	Failed	False	---	+	DEL	SHOW	FAILED	RE-RUN
7	April 24, 2020, 3:56 p.m.	pavao	13271	tf_submission.zip	Failed	False	---	+	DEL	SHOW	FAILED	RE-RUN

Figure 3 – Derived detection accuracies of different models

2) In which environment (programming language, time constraint, memory constraint) are submission files run?

For our first prototype, participants will need to set up their local environment by following the prerequisites below;

Install Anaconda Python 3.6.6, opencv-python (4.0.1), scikit-image (0.15.0). Download the starting kit. Usage: - modify sample_code_submission/model.py to provide a better model - zip the contents of sample_code_submission (without the directory, but with metadata)
The utility of Codalab is then to;
1) get submitted algorithm
2) score algorithm with predefined metric and environment constraint
3) update leaderboard.

- hosting (IIC, etc.)
  Since Codalab is an open-source framework, it can be deployed to any server. In the early stage of this project, we will use codalab.lri.fr (server maintained by Paris-Sud University) for testing and prototyping.
- possibility of an online benchmarking on a public test dataset
  At the moment, the platform does not allow public users to submit their own dataset to the benchmark. Otherwise, they can contact platform maintainers (TG-Malaria) to do so. In its current state, the platform has a sample public dataset for participants to prototype their ML solutions.
- protocol for performing the benchmarking (who does what when etc.)
  At the moment, minimal benchmarking system is created by the organizers for prototyping. The latter system will process submitted files in Python 3 environment with time budget of 10min. TG-Malaria benchmarking Organizers made available a starting kit (sample code submission) to ease the task of participants.

Participants on their side need to adapt their algorithm to fit the structure of the starting kit. The system allows a participant to submit up to 100 times but only 5 times in a day. This will enable each participant to fine tune their detection models.

- AI submission procedure including contracts, rights, IP etc. considerations
  Copyright of submitted source code will remain to the participants. Codalab allow participants to decide whether they want to make submissions publicly available or not.

6.2.3 Benchmarking V2 platform for malaria detection

Building from the first version 1 of the benchmarking platform, the new updated version 2 of the platform for benchmarking malaria detection with improvements is ready for an alpha test phase. To this end, a call for participation was drafted in our previous CfTGP document we are sought for active participation from people with background not only in computer vision, machine learning and artificial intelligence, but also data submission from microscopists to take part in our malaria detection challenge.

Some of the changes and updates are highlighted in the benchmarking Interface as shown in the Figure 4 below.

Figure 4: Updated user Interface for the benchmarking platform for malaria detection

The updates to the platform that have been affected are summarised below;

1. Adding support for uploading datasets;

   Unlike the first version of the platform which never had provision for data submission, Participants are now able to submit their own datasets through the "Upload dataset" to enrich current benchmark datasets (see figure 5). Submission contains several files: dataset information (name of features, target variable, copyright), dataset file (in Codalab format) and train-test split indices. To ensure data quality, the platform verifies uploaded dataset and rejects incorrect submission. The latter validation script is available for download by participants (reference is given in the platform). Note that at the moment, the uploading module only works for classification tasks.

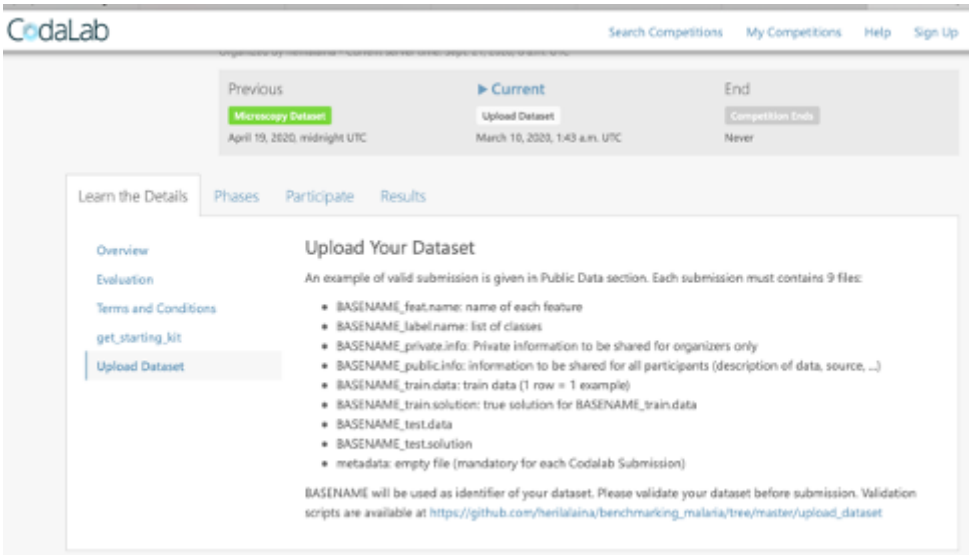

Figure 5: Update for data upload

2. Add New public dataset of thin blood smear dataset to the benchmark

A second aspect that was updated on the benchmarking platform is the provision for a new dataset. factors. In the initial version of the platform, only thick blood smear dataset was used. For test purposes, another dataset that comprises both infected and uninfected cell images obtained from [13] is added to have multiple tasks on the platform. The latter dataset is a classification task.

3. Adding support for a deep learning library (pytorch and tensorflow) and setting up time budget for 1 hour/submission.

   One of the limitations for the implementation of the initial benchmarking platform was on the implementation environment and time which allowed submission for only traditional machine learning models. With the update, comes an improvement with support for deep learning libraries and a submission time budget of up to 1 hour to enable bulky models.

6.2.3.1 Benchmarking system dataflow

This section describes the dataflow throughout the benchmarking architecture. The benchmarking platform has three components: data storage server, compute worker and frontend web application. Participants / organizers interact via the web application. It includes uploading data, configuring competition, manager list of participants, submitting AI models. On Codalab, data formats are flexible and only depend on the ingestion program. We may recall that the overall processing step is executed by the ingestion program, termed IP. In our competition, we stored data in libsvm format since it is the default format of Codalab datasets. Versioning of datasets are also handled by Codalab. Submissions (AI models) are processed as follows: first, they are handled by the ingestion program for an initial verification of its format and create the AI model (Python object). Then, the ingestion program reads data from storage (according to the path specified in the configuration files). The latter extracted dataset is fed to the AI model for training. The ingestion program monitors memory consumption and time budget during the training. If successful, IP fetches predictions and calls the script to compute the metrics. In our challenge, we computed predictive accuracy, ROC-AUC and Recall. Finally, IP pushed the scores on the leaderboard.

6.2.3.2 Safe and secure system operation and hosting

Dataset can be stored on Codalab or within another cloud server. Since the initial version of our challenge is hosted at the official platform Codalab server (codalab.org), we also opt to store our data on the same server. Codalab.org also hosts many AI challenges with well-known machine learning

conferences suggesting that the platform has some level of security. All the security aspects of data and source code are managed by the platform and the administrators of Codalab.org.

6.2.4 Updates on the benchmarking platform V3 for malaria detection

The initial version of the competition has been migrated to Codabench (https://www.codabench.org) - a successor version of Codalab focusing on benchmarking instead of competition (see figure 6). This choice is motivated for several reasons. Firstly, the improvement of Codabench over Codalab eases benchmark management tasks such as incorporating new metrics in the leaderboard or proposing new problems. Secondly, the functionality enabling participants to submit a new dataset is straightforward to implement in Codabench. And lastly, participants are given the flexibility to choose the set of datasets on which their submission will be benchmarked on. The figure below depicts the submission page on Codabench.

Note that our benchmarking is not published yet as it is still under development.

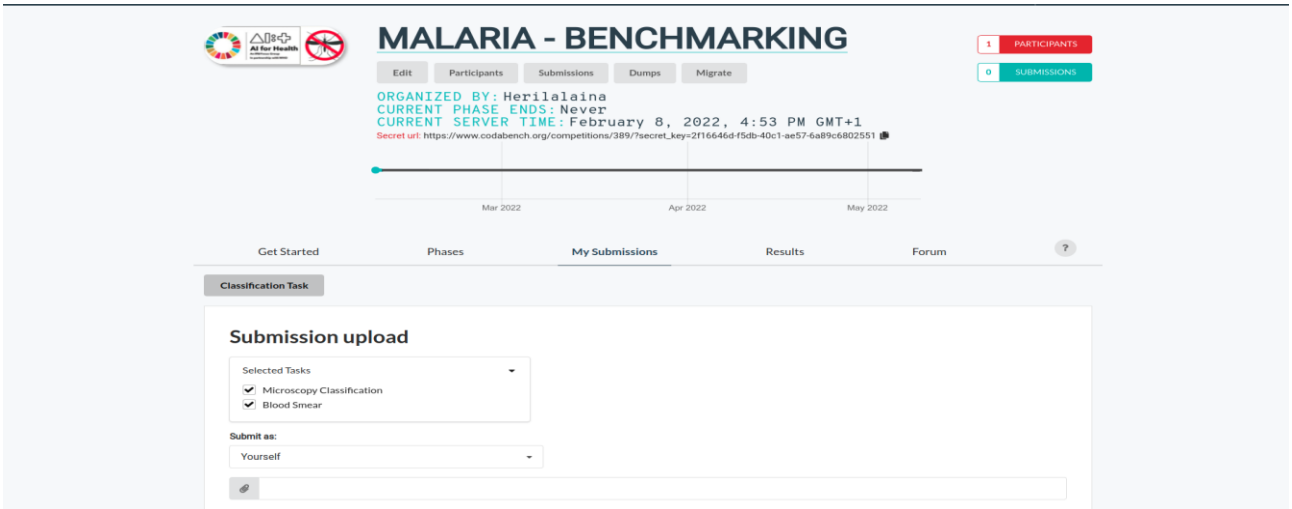

Figure 6: Updated user Interface for the benchmarking platform for malaria detection using Codabench

The next milestone is to incorporate object detection tasks to extend the available classification tasks. Since the processing pipeline for object detection is crucial, we plan to start with the VIPriors challenge template, which provides an end-to-end training of the model. The latter template thus enables participants to focus their effort on the model design part. However, some difficulties arise, especially in adapting the Microscopy dataset format into the VIPriors template and integrating the overall pipeline into Codabench. Therefore, future works will focus on addressing these issues.

6.3 Benchmarking process

The current version of the benchmarking system V3 will be a standalone system. The prediction of test dataset by AI systems, definition of AI tasks and benchmarking metrics in benchmarking, and execution of benchmarking calculation will be handled under the system implementation. A challenge-based implementation will be circulated to allow evaluation of performance of the different ML solutions.

6.3.1 AI input data structure for the benchmarking

This section describes the input data provided to the AI solutions as part of the benchmarking of TG-Malaria. 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.

For our first attempt on prototyping a benchmarking platform, the TG has leveraged on the existing dataset available (1182 images of thick blood smear slides that have been annotated by laboratory experts from Mulago referral hospital.

To this end, only image data of thick blood smears of image format .jpg is sufficient to build malaria detection models. This is currently because we do not have any other dataset at hand. We believe that future iterations will allow multiple datasets (thin blood smear images, demographic data, environment data) to allow derivation of more accurate predictive models.

Our first benchmarking task is built in the form of a codalab competition challenge in which we provide input dataset (thick blood smear images) to participants.

Image data together with corresponding labels (specifying presence or absence of malaria parasites) is provided. The TG envisions to attract machine learning experts who are particularly interested in automated malaria diagnosis to tune their models on the prototype dataset provided.

However, for a feasible and reliable solution, large amounts of data of both thick blood smear and thin blood smear images from different Health facilities in different malaria endemic countries would be required for machine learning models and an undisclosed test data for evaluation of the tool.

On the side of participants therefore, the input to our first benchmarking platform is a model to train on the available protype dataset available.

This section describes the recommended structure of input data provided to the AI solutions as part of the benchmarking of AI based detection of malaria'

– Image file format: JPEG format, PNG format or BMP format.
– Image file names: be unique in the dataset and anonymize the personal information of the patient.
– Image resolution: original resolution as captured with our microscopy data collection set-up using a mobile smart phone.

6.3.2 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.

The output should be documented in an arranged and clear way, like a CSV, XML or JSON file with the following information.

– Information of data (name, format, etc).
– Result of the data. It would depend upon the specific condition and the type of task that is being benchmarked.

6.3.2.1 Detection

– Data Information: data name, data format, etc.
– Result Information:
– Category Information: the types would depend on the task.
– Location Information: coordinates of a specific point (left-top or centre of the bounding box) in the image.
– Size Information: height and width in pixels.
– Task info(optional): task ID, task name, task type, etc.

6.3.2.2 Classification

– Data Information: data name, data format.

– Result Information

– Category Information: the types would depend on the task.

– Task Information (optional): task ID, task name, task type, etc.

6.3.2.3 Segmentation

– Data Information: data name, data format, etc.

– Result Information

– Category Information: the types would depend on the task.

– Path of segmentation file: the stored path of the segmentation file.

– Segmentation border Information (optional): coordinates of points of the segmentation mask.

– Task Information (optional): task ID, task name, task type, etc.

6.3.3 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.

A label/ annotation will be given of the blood smear Image that contains the malaria parasites. The labels will depend upon the specific condition that is being benchmarked and also the type of AI task.

For our first iteration, a binary task is considered with positive (parasite) and negative (no parasite) patches from an image are used.

6.3.4 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.

To evaluate AI tool's performance, labelled Dataset of blood smear images would be taken and tested against the performance of AI. The algorithm evaluation mechanism should include metrics like ROC accuracy, precision, recall, specificity F1 scores, specificity, sensitivity, mean Average Precision (mAP), average precision and the choice will be based on the algorithm used and purpose of the task.

6.3.5 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.

1) In order to assess algorithm robustness, sufficient undisclosed image data would be collected. This is envisioned to come from different health facilities both public and private.

2) There is need for examination of the quality of undisclosed dataset by a panel of experienced and skilled lab technicians. Bias in data will be considered.

3) An agreeable number of test data for a benchmarking task will be specified.

4) The annotation process of detection for example will include localizing the object inside the data and categorizing it. The bounding box is usually used to localize the object with a rectangular box which is called a bounding box.

The test dataset acquisition has not yet commenced.

6.3.6 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 DEL5.5 on *data handling* and DEL5.6 on *data sharing practices*).

6.3.7 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.

6.3.8 Reporting methodology

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

Reporting would be based on the accuracy of an AI tool's ability to detect the presence of malaria parasites,

– Public benchmarking leaderboard developed

– Making publication of the deliverables of the TG.

6.3.9 Some recent publications

With the growing interest in research around automated detection of malaria, some research has been conducted around improvement of malaria detection using AI with respect to assessing of data quality and use of new models and platforms for detection of malaria. Some publications are discussed below;

1) An approach for assessing quality of labelled Data for a machine learning task in Malaria detection [12]. While microscopy diagnosis through supervised learning for image analysis notably contributes to malaria detection, it has limitations. Among its principal challenges is the manual and tiresome process of data annotation for the classification task. The manual annotation of data is prone to inaccuracy defects due to bias, subjectivity and unclear images resulting in many false positives. This is normally due to personal independent judgements that vary from individual microscopists hence summatively affecting the accuracy of the model. In this study, we sought to investigate the possibility of classifying the negative far examples and the positive near examples from the positives in thick blood smear images for malaria detection. Assessing the classification performance could potentially inform us of the quality of training dataset and guide on selecting the best training dataset for a malaria parasite detection task. We employed the Mean Squared Error (MSE) to distinguish between positive and negative images. We later investigate the performance of the VGG-16 classification model based on how close or far negative examples are from positives. Experimental results showed that negative examples far from the positives produce better results than those near and that the proposed method could potentially be used to reduce false positives and bias in the training data.

2) A new approach for microscopic diagnosis of malaria parasites in thick blood smears using pre-trained deep learning models [11]. This research was motivated by the emerging technologies of machine learning that can learn complex image patterns and have accelerated research in medical image analysis. In this study, on a dataset of thick blood smear images, we evaluate and compare performance of three pre-trained deep learning architectures namely; faster regional convolutional neural network (faster R-CNN), Single-Shot multibox Detector (SSD) and RetinaNet through a Tensorflow object detection API. Data augmentation method was applied to optimise performance of the meta architectures. The possibility for mobile phone detector deployment was also investigated. The results revealed that faster R-CNN was the best trained model with a mean average precision of over 0.94 and SSD, was the best model for mobile deployment. We therefore deduce that faster R-CNN is best suited for obtaining high rates of accuracy in malaria detection while SDD is best suited for mobile deployment.

3) A web-based intelligence platform for diagnosis of malaria in thick blood smear images: A case for a developing Country [10]. The study was motivated by the need for development of remote systems that can provide fast, accurate and timely diagnosis of Malaria. With availability of internet, mobile phones and computers, rapid dissemination and timely reporting of medical image analytics is possible. This research aimed at developing and implementing an automated web-based Malaria diagnostic system for thick blood smear images under light microscopy to identify parasites. We implemented an image processing algorithm based on a pre-trained model of Faster Convolutional Neural Network (Faster R-CNN) and then integrated it with web-based technology to allow easy and convenient online identification of parasites by medical practitioners. The developed system holds the potential to improve the efficiency and accuracy in malaria diagnosis, especially in remote areas of developing countries that lack adequate skilled labour.

4) Mobile-Aware Deep Learning Algorithms for Malaria Parasites and White Blood Cells Localization in Thick Blood Smears [14]. The research was motivated by the need for effective determination of malaria parasitemia is paramount in aiding clinicians to accurately estimate the severity of malaria and guide the response for quality treatment. This study presents an end-to-end deep learning approach to automate the localization and count of P.falciparum parasites and White Blood Cells (WBCs) for effective parasitemia determination. The method involved building computer vision models on a dataset of annotated thick blood smear images. These computer vision models were built based on pre-trained deep learning models including Faster Regional Convolutional Neural Network (Faster R-CNN) and Single Shot multibox Detector (SSD) models that help process the obtained digital images. A mobile smartphone-based inference app to detect malaria parasites and WBCs in situ was developed. The proposed method can be applied to support malaria diagnostics in settings with few trained Microscopy Experts yet constrained with large volumes of patients to diagnose.

6.3.10 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.

Results are based on agreed upon evaluation metrics for an AI tool's ability to detect the presence of malaria parasites. We have so far developed our first and second version of the TG benchmarking system prototypes, TG-Malaria hinged on the following evaluation metrics; ROC AUC, precision, recall, Average precision.

Evaluation report for each AI solution submitted for in our trial versions.

The benchmarking platform computes the evaluation metrics and scores based on the public available dataset and the AI models used. Results of the different AI models in terms of evaluation metrics are finally shown on Codalab leaderboard. The system is time stamped and keeps track each time a participant submits a new entry.

Preliminary results of our prototype benchmarking project are as shown in the Figure 7 below;

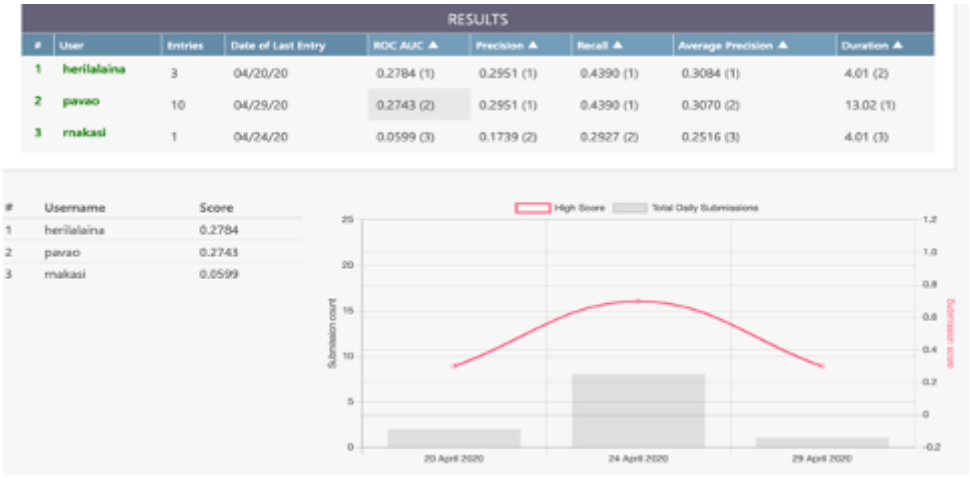

Figure 7: Result report for the AI models submitted

6.3.11 Overall 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). The benchmark platform is still under development and will be publicised for a concrete discussion of outcomes.

6.3.12 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 DEL4 "*AI software lifecycle specification"* (identification of standards and best practices that are relevant for the AI for health software life cycle).

7 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)* compiled the requirements that consider these challenges.

The deliverables with relevance for regulatory considerations are DEL2 *"AI4H regulatory considerations"* (which provides an educational overview of some key regulatory considerations), DEL2.1 *"Mapping of IMDRF essential principles to AI for health software",* and DEL2.2 *"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). DEL4 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-Malaria.

7.1 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 TG-Malaria.

7.2 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).

7.3 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.

7.4 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 DEL2 *"AI4H regulatory considerations."*

References

[1] I. Bates, V. Bekoe, and A. Asamoa-Adu. Improving the accuracy of malaria-related laboratory tests in Ghana. Malar. J. Vol.3. No. 38, 2004.

[2] C.A. Petti, C.R. Polage, T.C. Quinn, A.R. Ronald, and M.A. Sande. Laboratory medicine in Africa: A barrier to effective health care. Clinical Infectious Diseases, Vol. 42, N0. 3, PP. 377-382, 2006.

[3] L. Rosado, J.M. Correia da Costa, D. Elias, and J.S. Cardoso. A review of automatic malaria parasites detection and segmentation in microscopic images. Anti-Infective Agents, Vol 14, No. 1, pp 11-22, 2016.

[4] J. A. Quinn, R. Nakasi, P. K. B. Mugagga, P. Byanyima, W. Lubega, and A. Andama. Deep convolutional neural networks for microscopy-based point of care diagnosis. In proceedings of International Conference on Machine Learning for Health Care, Volume 50., 2016.

[5] WHO. World malaria report. Geneva, Switzerland, World Health Organization., 2016.

[6] WHO. World malaria report. Geneva, Switzerland, World Health Organization., 2015.

[7] WHO. World malaria report. Geneva, Switzerland, World Health Organization., 2017.

[8] M. Salathé, T. Wiegand, M. Wenz. Focus Group on Artificial Intelligence for Health. Available at; https://arxiv.org/ftp/arxiv/papers/1809/1809.04797.pdf

[9] D. Bhattacharya. Relevance of economic field microscope in remote rural regions for concurrent observation of malaria and inflammation. Advances In Infectious Diseases (AID), Vol.2. No.1, PP. 13-18., 2012.

[10] R. Nakasi, J.F. Tusubira, A. Zawedde, A. Mansourian, E. Mwebaze. A web-based intelligence platform for diagnosis of malaria in thick blood smear images: A case for a developing Country. Proceedings of the IEEE/ CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pp. 984-985, 2020.

[11] R. Nakasi, E. Mwebaze, A. Zawedde, J.F. Tusubira, B. Akera, G. Maiga. A new approach for microscopic diagnosis of malaria parasites in thick blood smears using pre-trained deep learning models. Springer SN Applied Sciences 2, 1255, 2020.

[12] R. Nakasi, E. Mwebaze, A. Zawedde, J.F. Tusubira, G. Maiga. An approach for Assessing quality of labeled Data for a machine learning task in Malaria detection. COMPOSS' 20: Proceedings of the 3rd ACM SIGCAS Conference on Computing and Sustainable Societies: PP. 301- 304, 2020.

[13] Arunava. Malaria cell images dataset. Available at : https://www.kaggle.com/iarunava/cell-images-for-detecting-malaria

[14] Nakasi, R.; Mwebaze, E.; Zawedde, A. Mobile-Aware Deep Learning Algorithms for Malaria Parasites and White Blood Cells Localization in Thick Blood Smears. *Algorithms* **2021**, *14*, 17. https://doi.org/10.3390/a14010017

Annex A:
Glossary

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

Acronym/Term	Expansion	Comment
AI	Artificial intelligence	
AI4H	Artificial intelligence for health	
AI-MD	AI based medical device	
API	Application programming interface	
CfTGP	Call for topic group participation	
DEL	Deliverable	
FDA	Food and Drug administration	
FGAI4H	Focus Group on AI for Health	
GDP	Gross domestic product	
GDPR	General Data Protection Regulation	
IMDRF	International Medical Device Regulators Forum	
IP	Intellectual property	
ISO	International Standardization Organization	
ITU	International Telecommunication Union	
LMIC	Low-and middle-income countries	
MDR	Medical Device Regulation	
PII	Personal identifiable information	
SaMD	Software as a medical device	
TBC	To Be Communicated	
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-Malaria
TG	Topic Group	
WG	Working Group	
WHO	World Health Organization	

Annex B:
Declaration of conflict of interests

The contributors declared that they have no conflict of interest.