



WG(s): N/A Zanzibar, 3-5 September 2019

DOCUMENT

Source: TG-Cogni Driver

Title: Call for Topic Group Participation: Standardized benchmarking of AI against neuro-cognitive diseases

Purpose: Engagement

Contact: Topic driver: Marc Lecoultre Email: ml@mllab.ai

Abstract: Calling on members of the medical and artificial intelligence communities with a vested interest in AI against neuro-cognitive diseases! Become engaged in the group dedicated to establishing a standardized benchmarking platform for AI against neuro-cognitive diseases within the International Telecommunication Union (ITU)/World Health Organization (WHO) Focus Group on “Artificial Intelligence for Health” (FG-AI4H). This document is the same as seen in meeting E, reproduced for meeting F for easier reference.

Call for Topic Group Participation: AI against neuro-cognitive diseases

The International Telecommunication Union (ITU)/World Health Organization (WHO) Focus Group on “Artificial Intelligence for Health” (FG-AI4H; <https://www.itu.int/go/fgai4h>) seeks engagement from members of the medical and artificial intelligence (AI) communities (including clinicians, technologists, entrepreneurs, potential benchmarking data providers, machine learning experts, software developers, researchers, regulators, policy-makers, companies/institutions, and field experts) with a vested interest in shaping the benchmarking process of AI against neuro-cognitive diseases.

About FG-AI4H

Over the past decade, considerable resources have been allocated to exploring the use of AI for health, which has revealed an immense potential. Yet, due to the complexity of AI models, it is difficult to understand their strengths, weaknesses, and limitations. If the technology is poorly designed or the underlying training data are biased or incomplete, errors or problematic results can occur. AI technology can only be used with complete confidence if it has been quality controlled through a rigorous evaluation in a standardized way. Towards developing this standard assessment framework of AI for health, the ITU has established FG-AI4H in partnership with the WHO.

Thus far, FG-AI4H has established thirteen topic groups. These are concerned with: AI and cardiovascular disease risk prediction, child growth monitoring, dermatology, falls among the elderly, histopathology, neuro-cognitive diseases, ophthalmology (retinal imaging diagnostics), psychiatry, radiotherapy, snakebite and snake identification, symptom checkers, tuberculosis, and volumetric chest computed tomography.

Each topic group agrees upon representative benchmarking tasks in a pragmatic, best-practice approach, which can later be scaled and expanded to similar tasks. Every benchmarking task should address a health problem of relevance (e.g. impacting a large and diverse part of the global

population or challenging to treat) and for which AI technology would provide a tangible improvement relative to the current practice (e.g. better care, results, and/or cost/time effectiveness).

For a rigorous and sound evaluation, undisclosed test data sets must be available (or have to be collected) for each task. All data must be of high quality and compliant with ethical and legal standards. In addition, the data must originate from a variety of sources so that it can be determined whether an AI algorithm can generalize across different conditions, locations, or settings (e.g. across different people, hospitals, and/or measurement devices). The format/properties of the data serving as input to the AI and of the output expected from the AI, as well as the benchmarking metrics are agreed upon and specified by the topic group.

Finally, the AI-to-be-evaluated will be benchmarked with the undisclosed test data on FG-AI4H computing infrastructure. Here, the AI will process single samples of the undisclosed test data set and predict output variables, which will be compared with the "ground truth." The results of the benchmarking will be provided to the AI developers and will appear on a (potentially anonymized) leaderboard.

Topic group: AI against neuro-cognitive diseases

A topic group is a community of stakeholders from the medical and AI communities with a shared interest in a topic. The objectives of the topic groups are manifold:

1. to provide a forum for open communication among various stakeholders,
2. to agree upon the benchmarking tasks of this topic and scoring metrics,
3. to facilitate the collection of high-quality labelled test data from different sources,
4. to clarify the input and output format of the test data,
5. to define and set-up the technical benchmarking infrastructure, and
6. to coordinate the benchmarking process in collaboration with the Focus Group management and working groups.

The primary output of a topic group is one document that describes all aspects of how to perform the benchmarking for this topic. (The document will be developed in a cooperative way by suggesting changes as input documents for the next FG-AI4H meeting that will then be discussed and integrated into an official output document of this meeting. The process will continue over several meetings until the topic description document is ready for performing the first benchmarking.)

This topic group is dedicated to AI against neuro-cognitive diseases. We provide an empirical basis for testing the clinical validity of machine learning-based diagnostics for Alzheimer's disease (AD) and related dementia syndromes (defined by DSM V as 'Neurocognitive disorders') using real world brain imaging and genetic data. With increased life expectancy in modern society, the number of individuals who will potentially become demented is growing proportionally. Current estimates count world-wide over 48 million people suffering from dementia bringing the social cost of care to 1% of world's gross domestic product – GDP. These numbers led the World Health Organisation to classify neurocognitive disorders as a global public health priority.

Compared to visual assessment, automated diagnostic methods based on brain imaging are more reproducible and have demonstrated a high accuracy in separating AD from healthy aging, but also the clinically more challenging separations between different types of neurocognitive disorders. Similarly, although ApoE genotypes carrying higher risk for AD are easily obtainable, this information is rarely integrated in machine learning-based diagnostics for AD. Although encouraging, implementations into clinical routine have been challenging.

A large representative sample will be created and will be use for the creation of the models. The models will be then validated (see benchmarking methods below) on the real-world undisclosed patient's data.

The benchmarking process will be based on the most modern methods used by the ML community, but also on the recommended methodology for clinical trials. Thus, assessment of clinical validity involves measurement of the following metrics derived from the confusion matrix:

- Test accuracy: F1 score
- Clinical sensitivity: ability to identify those who have or will get the disease = $TP/(TP+FN)$
- Clinical specificity ability to identify those who do not have or will not get the disease = $TN/(FP+FN)$
- Clinical precision the probability that the disease is present when the test is positive = $\text{sensitivity} \times \text{prevalence} / (\text{sensitivity} \times \text{prevalence} + (1-\text{specificity}) \times (1-\text{sensitivity}))$

In addition, we propose to integrate clinician feedback by measuring the Clinical utility. This measure assesses the impact of the automated decision in term of impact on the clinical path of the patients, impact on the treatment and impact on the relatives ...).

The primary data are already available and growing in volume. Data will include both real world patient's data and data collected from research cohorts. The data will include clinical scores, diagnostic, cognitive measures and biological measures (PET, MRI, fMRI, lab results).

The data include patients on more than 6 000 patients on dementia (one of the largest patients' cohort) different stages of the disease (subjective complains, mild impairments or demented)

More details about the activities of the topic group can be found in the documents:

[FGAI4H-C-020-R1: Status report for Alzheimer's disease use case](#)

[FGAI4H-B-013-R1: Proposal: Using machine learning and AI for validation of Alzheimer's disease biomarkers for use in the clinical practice](#)

These can be accessed with a free ITU account (cf. "Get involved").

Current members of the topic group on AI against neuro-cognitive diseases include:

Kherif Ferah, vice-director LREN, CHUV - Switzerland

Senior Lecturer at the University of Lausanne and vice director of the Laboratoire de Recherche en Neuroimagerie (LREN) of Departement des Neurosciences Cliniques (DNC) at the University Hospital of Lausanne (CHUV). He obtained his PhD in neuroscience at Pierre and Marie Curie University, Paris. He was research fellow at MRC-CBSU in Cambridge and then at the Wellcome Trust Centre for Neuroimaging in London before his arrival in Lausanne in 2010. He used functional imaging to probe cognitive function and used my mathematical background to test new hypotheses pertaining the explanation of individual differences.

Marc Lecoultre, MLLab.ai – Switzerland

Expert in AI & Data Science, strong entrepreneurship professional with a master's degree from the Swiss Federal Institute of Technology, a Graduate Certificate from Stanford and multiple certifications in Lean Management and AI domains. He founded several companies in these fields. He practiced AI and Machine Learning for over 15 years. He has worked on dozens of projects in various companies and industries. He is an editor and actively participates to the WHO/ITU focus group on AI for health.

The topic group would benefit from further expertise of the medical and AI communities and from additional data.

Get involved

To join this topic group, please send an e-mail to the focus group secretariat (tsbfgai4h@itu.int) and the topic driver (Marc Lecoulre, ml@mllab.ai). Please use a descriptive e-mail subject (e.g. "Participation topic group AI against neuro-cognitive diseases"), briefly introduce yourself and your organization, concisely describe your relevant experience and expertise, and explain your interest in the topic group.

Participation in FG-AI4H is free of charge and open to all. To attend the workshops and meetings, please visit the Focus Group website (<https://www.itu.int/go/fgai4h>), where you can also find the whitepaper, get access to the documentation, and sign up to the mailing list.
