



INTERNATIONAL TELECOMMUNICATION UNION

TELECOMMUNICATION
STANDARDIZATION SECTOR

STUDY PERIOD 2022-2024

FGAI4H-P-019-A02

ITU-T Focus Group on AI for Health

Original: English

WG(s): Plen Helsinki, 20-22 September 2022

DOCUMENT

Source: TG-Psy Topic Driver

Title: Att.2 – CfTGP Update (TG-Psy) [same as Meeting H]

Purpose: Engagement

Contact: Nicolas Langer Email: n.langer@psychologie.uzh.ch
ETH Zurich, Switzerland

Abstract: Calling on members of the medical and artificial intelligence communities with a vested interest in Psychiatry! Become engaged in the group dedicated to establishing a standardized benchmarking platform for AI in psychiatry within the International Telecommunication Union (ITU)/World Health Organization (WHO) Focus Group on “Artificial Intelligence for Health” (FG-AI4H).

This version of the CfTGP is the same as seen in Meeting H (FGAI4H-H-019-A02), reproduced for easier reference as a Meeting N document.

NOTE – For public distribution of this call for participation, remove this cover page.

Call for Topic Group Participation: AI for Psychiatry

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 in psychiatry.

1 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.

2 Topic group: AI for Psychiatry

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 for psychiatry. Psychiatric disorders are among the most common and debilitating illnesses across the lifespan and begin usually prior to age 24, which emphasizes the need for increased focus on studies of the developing brain. But the topic group “AI for Psychiatry” does also welcome stakeholders, which are interested in adult and geriatric psychiatry. The majority of existing studies have focused on differentiating between patients with an isolated psychiatric disorder and healthy controls. However, this line of research does not reflect the real-life situation (over 75% of patients with a clinical diagnosis have multiple psychiatric disorders), in which a clinician has the task to choose between different diagnoses and/or the combination of multiple diagnoses (multimorbidity). A key challenge in this topic group “AI for Psychiatry” is the annotation of the labels (psychiatric disorders) for machine learning. Mapping the diagnostic label from a clinically defined nosology (e.g., the Diagnostic and Statistical Manual of Mental Disorders (DSM) or the International Classification of Diseases (ICD)) to varying biological measures has proven to be problematic. It assumes consistent biological relationships with broad constellations of symptoms that in practice do not show [1, 2]. As a consequence, two research approaches emerged. First is the implementation of transdiagnostic models structured around behavioral and neurobiological dimensions that transcend current diagnostic boundaries [3]. Second is the identification of diagnostic subgroups to explain variation within diagnostic categories through the detection of behaviorally or biologically homogeneous subgroups [4, 5, 6]. The approach of integrating different (e.g. behavioral and neurobiological) measures of a comprehensive set of cognitive domains has become increasingly influential in the field of psychiatry. In fact, the multimodal and multi-dimensional integration of neuroscientific measures is central to one of the flagship projects of the National Institute of Mental Health (NIMH), the Research Domain Criteria Project (RDoC), initiated in their Strategic Plan in 2008. In a nutshell, RDoC advocates that mental illness may be better reflected in considering the continuously scaled profile of function and dysfunction on a broad range of cognitive domains rather than by a list of predefined “core symptoms”. In line with this, it is argued that in the long run genuine advances in the treatment of mental disorders may be reached by developing reliable and valid measures across a range of measurement modalities.

Furthermore, most of the previous studies employ traditional univariate statistics on relatively small samples. Multivariate machine learning/AI approaches have a great potential to overcome the limitations of this approach. The topic group “AI for Psychiatry” offers a unique large-sample pediatric dataset that provides a wide array of different psychiatric developmental disorders. We will leverage existing data from the biobank of the Healthy Brain Network (HBN) initiative (<https://healthybrainnetwork.org/>). The data acquisition included multimodal brain imaging (Diffusion Tensor Imaging, structural T1-weighted and functional MRI), electroencephalography (EEG), and an extensive phenotyping protocol of comprehensive psychiatric (clinical classification according to the DSM-V), learning, familial, environmental, and lifestyle assessments. The current goal is to classify the multimorbidity of children and adolescents based on resting state EEG data. In addition, demographic information as well as extensive cognitive and behavioral measures will be permitted to derive predictive models. This restriction is introduced due to the limited real-world practicability and economic viability of MRI and DTI measurements. Future AI challenges will potentially include other measures, such as task-related EEG and neuroimaging (T1-weighted MRI, DTI, and functional MRI) data to assess whether the previously achieved prediction accuracy can

be exceeded using these data. In the present challenge, the organizers will provide raw and preprocessed EEG data as well as specifically extracted EEG features, which has been shown relevant to different psychiatric developmental disorders (e.g. theta-beta ratio, frontal alpha asymmetry). Using this approach, we expect to attract both neuroimaging experts who want to employ their own EEG processing pipeline as well as participants without a neuroscience background, who are more interested in the machine learning aspect of the problem and may be happy to work on pre-extracted features. We believe that a community driven effort to derive predictive markers from these data using advanced AI algorithms can help to improve the diagnosis of psychiatric developmental disorders. More details about the activities of the topic group “AI for Psychiatry” can be found in the documents [FGAI4H-C-013](#) and [FGAI4H-C-013-A1](#). These can be accessed with a free ITU account (cf. “Get involved”).

Current members of the topic group “AI for Psychiatry” include Prof. Nicolas Langer and Dr. Stefan Haufe. Prof. Langer and Dr. Haufe share interest in using machine learning methods to analyze neurophysiological data in combination with behavioral and cognitive measures to advance the endeavor of biomarkers for psychiatric disorders. Such an integration of different types of brain and behavioral measures requires knowledge about the characteristics of the measurement modalities involved but also about the methodological approaches (e.g. multivariate analysis) to examine information that is encoded in the combination of the measures. During Nicolas Langer’s work with Prof. Nadine Gaab at the Harvard Medical School, Boston, USA, he has built an expertise in studying neurophysiological data and integrating this information behavioral and cognitive data. Prof. Langer has later joined Dr. Michael Milham at the Child Mind Institute in New York, USA, as a co-investigator of the HBN. With the HBN, they initiated a project that closely follows the goal to identify potential biomarkers for psychiatric developmental disorders. Dr. Stefan Haufe is a computer scientist developing machine learning and signal processing techniques for analyzing neuroimaging (in particular, EEG) data. He is currently an ERC junior group leader at Charité - Universitätsmedizin Berlin, Germany, where he is leading a five-year effort to characterize and predict psychiatric and neurological diseases using non-invasive brain electrophysiology. As such, he is very interested in benchmarking EEG-based biomarkers on large public clinical datasets. Prof. Langer and Dr. Haufe are interested to promote and support standardization and benchmarking efforts, which are crucial to the implementation of machine learning for diagnostics of psychiatric disorders. The topic group would benefit from further expertise of the medical and AI communities and from additional data. Thus, we like to invite any potential interested stakeholder. For example, clinicians, who could add their clinical expertise, ideally in multimorbidity of psychiatric disorders). In addition, clinical institutions could potentially provide additional (undisclosed) data, such as cognitive, behaviourally, neurophysiological data and clinical diagnosis. Furthermore, we aim to collaborate with researchers from cognitive neuroscience, machine learning experts, engineers and statisticians familiar with neurophysiological data as well as software developers.

3 Get involved

To join this topic group, please send an e-mail to the focus group secretariat (tsbfgai4h@itu.int) and the topic driver (n.langer@psychologie.uzh.ch). Please use a descriptive e-mail subject (e.g. "Participation topic group AI for Psychiatry"), 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.

References:

1. Cuthbert & Insel (2013). Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med*, 11(1), 126.
 2. Insel et al. (2010). Research domain criteria (RDoC): toward a new classification framework for research on mental disorders.
 3. Insel, T. R., & Cuthbert, B. N. (2015). Brain disorders? Precisely. *Science*, 348(6234), 499-500.
 4. Fair et al. (2012). Distinct neuropsychological subgroups in typically developing youth inform heterogeneity in children with ADHD. *PNAS*, 109(17), 6769-6774.
 5. Miranda-Dominguez et al. (2014). Connectotyping: model based fingerprinting of the functional connectome. *PloS one*, 9(11), e111048.
 6. Van Dam, N. T. et al. Data-Driven Phenotypic Categorization for Neurobiological Analyses: Beyond DSM-5 Labels. *Biol. Psychiatry*, 81, 484–494 (2016).
-