

Objective measures of psychiatric symptoms through speech and visual features during clinical interaction

Submission ID 3001003

SUBMISSION DETAILS

I agree to provide poster pdf for attendee download. Yes

Methodological Issue Being Addressed Compared to the progress seen in other branches of medicine, psychiatry lacks reliable and objective biomarkers to individually tailor treatments. Today, clinical states are measured by using question-based scales related to specific symptomatic domains which may be subject to biases. Thus, identifying objective markers of psychiatric disease states, including behavioral-based phenotypes, to support transdiagnostic dimensional approaches, is necessary for disease classification.

Since psychiatric disorders are disorders of social interaction emphasizing the importance of studying behavioral dynamics in real-life social interaction at the dyadic level rather than solely individual behavior. Interaction-based phenotyping to quantify the level of social behavior exchange or interpersonal synchrony of behavior could deliver additional data to generate an observer-independent picture of a patient's mental state ultimately leading to earlier and more sensitive identification of abnormalities. An Interaction-based sociometric approach combined with new automated monitoring technologies delivers novel state-associated biomarkers for diagnostic and therapeutic guidance. It leverages advancements in machine learning, computer vision, and natural language processing to extract meaningful information from patients' speech and visual cues.

Recent progress in computer vision, speech processing and machine learning has enabled detailed and objective characterization of human interaction behavior. Applying these advanced methods of AI provides new opportunities to identify digital markers of patient behaviour. Thus, merging computer vision-based measurement and speech feature in a multi-modal approach would enhance the quality of analysis by allowing to detect changes in the quality of communication as alterations in the dyadic interaction patterns.

Introduction We aim to identify and formalize a set of novel multimodal digital biomarkers derived from audio and visual interaction data and to develop predictive models within the scope of depression and schizophrenia. On this basis, we aim to develop models aiding in differential diagnosis, forecasting the patient's status (e.g., relapse prediction), and predicting therapeutic alliance. Important outcomes include technical and organizational methods for the management of medical data, demonstration scenarios covering patients' journeys including early detection, diagnosis support, relapse prediction, therapy support, an annotated corpus.

Methods A longitudinal multicenter observational study is performed across three clinical sites in Germany and France for which clinical interviews with psychiatric patients are audio and video recorded after they give consent. The following types of data is being analyzed: clinical scores

(questionnaires and scales), speech, video and physiological measures.

The audio analysis from dyadic interactions, the data will be pre-processed and speaker diarization and speaker labelling will be performed. In addition, the speech will be either manually annotated or automatically transcribed. Both paralinguistic and linguistic features will be extracted on the utterance level. Examples of speech features include pitch, speech rate, spectral and prosodic characteristics as well as word usage and content.

From the video streams, we will extract a comprehensive set of visual behaviour descriptors from both patient and clinician. These include gaze, head pose, facial expressions, posture, and body movements. We will make use of state-of-the-art approaches in computer vision that enable automatic extraction of such behaviours.

To achieve the primary objective of the study a comparison analysis will be performed between the new digital markers and the standard clinical measures. We will perform a multi-modal analysis using combined data modalities as well.

Results The study is currently still ongoing with so far over 200 hours of collected recordings of more than 100 psychiatric patients. Preliminary analysis of a sample size of 50 patients will be presented at the conference.

We will calculate the relationship between the extracted audio and video measures and several clinical scales, thus symptom severity in an exploratory analysis using Spearman's rank correlation coefficient. Moreover, to make sense of the extracted speech and visual features, machine learning models are employed containing examples of various psychiatric states. The goal is to develop algorithms that can accurately classify and predict these states based.

We hypothesize that changes in symptoms' severity are highly reflected in changes in speech and visual measures and that we will find significant correlations and precise prediction on clinical outcome. Integrated with traditional diagnostic methods, these measures could help improve the accuracy and objectivity of psychiatric assessments.

Conclusion By combining speech and visual features with machine learning algorithms, we aim to develop reliable and objective measures of psychiatric symptoms. These measures can assist clinicians in diagnosing and monitoring the severity of symptoms, track changes over time, and evaluate the effectiveness of interventions and treatment strategies.

Co-Authors

* Presenting Author

First Name	Last Name	Affiliation
Alexandra *	König *	Ki:elements

Hali	Lindsay	German Research Center for Artificial Intelligence (DFKI)
Philippe	Mueller	German Research Center for Artificial Intelligence (DFKI)
Danilo	Postin	Department of Psychiatry, School of Medicine and Health Sciences, Carl von Ossietzky University of Oldenburg, Bad Zwischenahn 26160, Germany
Eric	Ettore	Department of Psychiatry, Hopital Pasteur, CHU de Nice, 06000 Nice, France
Michal	Balazia	Institut national de recherche en informatique et en automatique (INRIA), Stars Team, France
Amandine	Lecomte	Institut national de recherche en informatique et en automatique (INRIA). Semagramme Team. Nancy. France
Benedikt	Wirth	German Research Center for Artificial Intelligence (DFKI)

Keywords

Keywords
digital phenotypes
psychiatric symptoms
speech language processing
computer vision
clinical interaction

Guidelines I have read and understand the Poster Guidelines

Disclosures if applicable Nothing to disclose