Depression ClinRO inference using momentary multimodal behavioral and physiological digital biomarkers from decentralized data collection

Submitter Jeffrey Cohn

Affiliation Deliberate.AI, University of Pittsburgh

SUBMISSION DETAILS

I agree to provide poster pdf for attendee download. Yes

Poster PDF for download <blank>

What is the Methodological Question Being Addressed? Can multimodal machine learning (MML) using objective, continuous measures of facial expression, body motion, voice, and physiology reliably measure depression severity?

Introduction Current methods of assessing depression severity are limited to subjective measures of patient self-report and clinical interview. Self-report is limited by patients' reading ability, idiosyncratic use, inconsistent metric properties across scale dimensions, reactivity, and differences between clinicians' and patients' conceptualization of symptoms. Clinician interviews enable more consistent use, but are time-intensive, difficult to standardize across settings, inherently subjective, and susceptible to reactivity effects and bias. Suboptimal interrater reliability is a key contributor to failed phase 3 clinical trials, leading to most studies being underpowered (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6078483/).

Multimodal machine learning (MML) using face, head movement, and vocal prosody presents a powerful alternative to conventional measures. In a recent clinical trial, MML achieved kappa = 0.73 for three levels of depression severity [1]. In a multi-site study, MML using vocal prosody, gaze, and head dynamics generalized well across datasets from three different Western countries [2]. Sample sizes were relatively small, however, and severity was measured with only two or three ordinal groupings. For use in clinical trials, reliability must be demonstrated in large sample sizes with severity measured with point-level accuracy across the full range of possible scores. We will test the hypothesis: MML can achieve high concurrent validity for point-wise measurement with clinician reported (ClinRo) HAM-D and MADRS interviews.

Methods Five hundred participants (2:1 female and 40% non-Caucasian) with a full range of depression severity from multiple settings will be interviewed using the current gold-standard Hamilton Depression Rating Scale (HAM-D/HDRS) and Montgomery-Asberg Depression Rating Scale (MADRS). To maximize effective inter-observer reliability for training the MML model, interviews will be independently rated by three highly-trained interviewers and combined. Because depression is in part interpersonal, both participant and interviewer behavior will be recorded using synchronized high-resolution cameras, directional microphones, and passive sensors for heart rate, respiration, and electrodermal. To ensure interpretability, we will use a combination of shallow and deep

learning approaches. The interpretable feature set identified by [3] along with physiology will be of particular focus.

Results Data collection will be completed at time of poster presentation for the first 96 of 500 participants in the initial two clinical settings. At minimum, the following results will be included

- Descriptive statistics:
  - o HAM-D and MADRS scores frequency histogram
  - o Participants characteristics: age, gender, ethnicity.
  - o Rater characteristics: TBC
- Interrater reliability calculated as two-way mixed intraclass correlation coefficient (ICC)
- Regression model performance using a k-fold internal cross-validation approach:
  - o Mean absolute error (MAE) using averaged HAM-D and MADRS scores as ground-truth
  - o 'Differential' ICCs: recalculated ICC using two approaches: a) addition: model added to all existing raters, and b) swap: each individual rater replaced by model

References

1. Alghowinem, S., et al., Depression detection model interpretation via feature selection methods. IEEE Transactions on Affective Computing, 2020.
2. Dibeklioglu, H., Z. Hammal, and J.F. Cohn, Dynamic multimodal measurement of depression severity using deep autoencoding. IEEE Journal of Biomedical Health Informatics, 2018. 22(2): p. 525-536.

Conclusion We expect our preliminary analysis to provide insight into the hypothesis that MML can achieve high concurrent validity (ICC > 0.80) for point-wise measurement with clinician reported (ClinRo) HAM-D and MADRS interviews.

Co-Authors

* Presenting Author

First Name	Last Name	Affiliation
Jeffrey *	Cohn *	Deliberate.AI, University of Pittsburgh
Marc	Aafjes	Deliberate.Ai
Marcelo	Ciconet	Deliberate.Ai
Katie	Aafjes van Dorn	Deliberate.Ai

Keywords

Keywords
Depression
Multimodal machine learning
HAM-D/HDRS

MADRS
ClinRO

Guidelines I have read and understand the Poster Guidelines

Disclosures All authors are shareholders of and work with Deliberate Solutions, Inc. and report no conflicts of interest in this work.

Related Tables and Supporting Materials <blank>