Voice markers of neuropsychiatric disorders: assessing the generalizability performance of machine learning models

Alberto Parola
Copenhagen University, Copenhagen, Denmark

Astrid Rybner
Aarhus University, Aarhus, Denmark

Emil Trenckner Jessen
Aarhus University, Aarhus, Denmark

Stine Nyhus Larsen
Aarhus University, Aarhus, Denmark

Marie Damsgaard Mortensen
Aarhus University, Aarhus, Denmark

Arndis Simonsen
Aarhus University, Aarhus, Denmark

Yuan Zhou
Chinese Academy of Sciences, Beijing, China

Katja Koelkebeck
Hospital and Institute of the University of Duisburg-Essen, Essen, Germany

Vibeke Bliksted
Aarhus University, Aarhus, Denmark

Riccardo Fusaroli
Aarhus University, Aarhus, Denmark

Abstract

This research explores the potential of machine learning (ML) in identifying vocal markers for schizophrenia. While previous research showed that voice-based ML models can accurately predict schizophrenia diagnosis and symptoms, it is unclear to what extent such ML markers generalize to different clinical subpopulations and languages: the assessment of generalization performance is however crucial for testing their clinical applicability. We systematically examined voice-based ML model performance on a large cross-linguistic dataset (3 languages: Danish, German, Chinese). Employing a rigorous pipeline to minimize overfitting, including cross-validated training sets and multilingual models, we assessed generalization on participants with schizophrenia and controls speaking the same or different languages. Model performance was comparable to state-of-the art findings (F1-score 0.75) within the same language; however, models did not generalize well - showing a substantial decrease - when tested on new languages, and the performance of multilingual models was also generally low (F1-score 0.50).

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*