Commentary on paper [P2] Asif Salekin

[P2] **Asif Salekin**, Jeremy W. Eberle, Jeffrey J. Glenn, Bethany A. Teachman, and John A. Stankovic. 2018. A Weakly Supervised Learning Framework for Detecting Social Anxiety and Depression, ACM Interactive, Mobile, Wearable, and Ubiquitous Technologies (IMWUT), Vol. 2, No. 2, Article 81 (June 2018), 26 pages. https://doi.org/10.1145/3214284

Problem and Motivation: Although social anxiety and depression are common, they are often underdiagnosed and undertreated, in part due to difficulties identifying and accessing individuals in need of services. Current assessments rely on client self-report and clinician judgment, which are vulnerable to social desirability and other subjective biases. Identifying objective, non-burdensome markers of these mental health problems, such as features of speech, could help advance assessment, prevention, and treatment approaches. Prior studies showed that speech's prosodic, articulatory, and acoustic features could indicate disorders such as depression and social anxiety. Previous research examining speech detection methods has focused on fully supervised learning approaches employing strongly labeled data. However, strong labeling of persons high in disorder symptoms in speech audio data is impractical, partly because it is impossible to identify with high confidence which regions of a long speech indicate the person's disorder symptoms. We developed a weakly supervised learning framework for detecting social anxiety and depression from long audio clips.

Paper's Novelty and Contribution: We collected long speech audio samples from individuals already diagnosed with or high in symptoms of specific mental disorders from situations that may heighten the expression of the symptoms of respective disorders. This type of data is considered "weakly labeled," meaning that although they provide information about the presence or absence of disorder symptoms, they do not provide additional details, such as the precise times in the recording that indicate the disorder or the duration of those identifying regions. We developed a weakly supervised deep-learning framework for detecting social anxiety and depression from long audio clips.

Specifically, we presented a novel feature modeling (knowledge engineering and representation) technique named NN2Vec, which identifies and exploits the inherent relationship between speakers' vocal audio states and disorder symptoms/states. An interesting property of the generated NN2Vec feature representation is it captures the syntactic relations among the vocal acoustic signals. NN2Vec representations are similar for audio signal patterns with a similar probability of occurring in the same class/category. Neural networks typically respond in a similar manner to similar inputs. Generated distributed NN2Vec representations are designed to take advantage of this; audio signal patterns that should result in similar responses are represented by similar NN2Vec representations, and audio signal patterns that should result in different responses are represented by quite different NN2Vec representations. Hence, identifying sequences of vocal acoustic signal patterns indicative of a mental disorder should be easier for a weakly supervised classifier. In addition, we developed a new multiple-instance learning adaptation of a BLSTM classifier named BLSTM-MIL. The presented novel framework of using NN2Vec features with the BLSTM-MIL classifier achieves significantly higher F-1 scores in detecting speakers who are high in social anxiety and depression symptoms.

Notably, before this study, no existing dataset contained spontaneous speech labeled with speakers high in social anxiety. Hence, we built a dataset consisting of 3-minute samples of weakly labeled spontaneous speech from 105 participants. Readily accessible, not intrusive or burdensome, and free of extensive equipment, the NN2Vec and BLSTM-MIL framework is a scalable complement to healthcare providers' self-report, interview, and other assessment modalities.

Professional Significance: I was the first author of this paper. The paper received 55 citations. The paper was published in the csranking.org ranked top conference: Ubicomp/IMWUT 2018. Currently, the work is being utilized for caregivers' mental health assessment in collaboration with the Nursing and Behavioral Science department of the University of Tennessee.