Commentary on Five Most Significant Publications

[P1] Harshit Sharma, Yi Xiao, Victoria Tumanova, **Asif Salekin**, "Psychophysiological Arousal in Young Children Who Stutter: An Interpretable AI Approach", Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT), 2022. https://doi.org/10.1145/3550326

Problem and Motivation: Speech production is a complex process requiring precise vocal tract coordination while simultaneously processing cognitive-linguistic information. Naturally, speech production can be affected by the speaker's physiological arousal. Literature shows that young children who stutter (CWS) are especially vulnerable to such influences. Given that preschool age is the time when essential communication skills are undergoing the most significant development, and also when some children develop stuttering, it is essential for our understanding of stuttering to examine young children's physiological responses during speech production. *This study presents an interpretable AI approach to identify the second-by-second fluctuations and pattern differences in physiological arousal of preschool-age children who stutter (CWS) compared to others who don't during various speaking tasks.*

Paper's Novelty and Contribution: The challenges of the paper include a lack of data annotations since the subtle physiological markers of CWS are yet unknown. However, we know that the data belong to CWS or others, making the data labels weakly labeled. Our preliminary analysis demonstrated that the latent patterns indicative of CWS's unique situational response from different physiological sensors, such as heart rate (HR), EDA, and respiratory rate (RSP), are sparse and do not emerge simultaneously; hence, there is a need for modality-wise distinctive pattern identification. However, literature has shown that cross-modality integration of these sensors is also highly effective in affective state and physiological arousal assessment.

Hence, we developed an approach, MI-MIL, that identifies the CWS's distinctive sparse patterns independently from each physiological modality without any available annotations of such patterns and captures and leverages the cross-relationships of the identified modality-specific sparse patterns for effective CWS vs. CWNS classification. To address the weakly labeled data challenge and identify modality-wise distinct patterns, MI-MIL applies the modality-wise multiple-instance-learning (MIL) paradigm in each physiological modality independently. MIL paradigm is designed to extract sparse and subtle patterns from weakly labeled data (i.e., without any fine-grain annotations of the region, timestamps, or duration of the patterns in the data). To capture the cross-modality-relations, MI-MIL presents a novel modality-fusion network that identifies the cross-relations of each modality's CWS indicative sparse patterns.

Moreover, leveraging SHAP, we visualize and discuss the fine-grain, second-by-second, temporal, and distinctive physiological response (represented through physiological parameters: HR, EDA, RSP) patterns of CWS from others during speech production in different challenging conditions. This paper is the first to leverage and discuss machine learning model interpretation and its findings on stuttering individuals. Such visualization and identification of patterns have both group-wise and personalized impacts. Identifying and visualizing group-wise patterns would enhance our understanding of stuttering etiology and development that eventually can improve clinical services for people who stutter. Personalized temporal pattern identification would enable remote, continuous, and real-time monitoring of stuttering children's physiological arousal, which may lead to personalized, just-in-time interventions to mitigate their arousal responses during speaking, resulting in an improvement in speech fluency.

Professional Significance: The paper published in csranking.org listed top conference Ubicomp/IMWUT 2022. The work resulted from my lab's ongoing collaboration with the Syracuse University Speech clinic, funded by two NIH grants ($\#R21DC018103$ and $\# R01DC017476-S2$). My Ph.D. student Harshit Sharma is the first author, my Ph.D. student Yi Xiao is the second author, and I am the corresponding author.

[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 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.

[P3] Jingyu Xin, Vir V. Phoha, **Asif Salekin**, "Combating False Data Injection Attacks on Human-Centric Sensing Applications", Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT), 2022. https://doi.org/10.1145/3534577

Problem and Motivation: Human-centric sensing applications on smart wearables are vulnerable to false data injection attacks (FDIA) that alter a portion of the victim's sensory signal with forged data comprising a targeted trait. Such a mixture of forged and valid signals successfully deceives the continuous authentication system to accept it as an authentic signal. Simultaneously, introducing a targeted trait in the signal misleads human-centric applications to generate specific targeted inferences; that may cause adverse outcomes. Take a health monitoring system as an example that utilizes smartwatch physiological sensory data to assess a patient's health. Suppose the FDIA attacker injects other unhealthy individuals' physiological signals into the genuine healthy user's physiological sensory signals. In that case, it can cause wrong health assessment that may lead to unnecessary interventions and harm the patient's health. This paper evaluates false data injection attack (FDIA), its deception efficacy on sensor-based authentication and human-centric sensing applications simultaneously, and presents a novel attack detection approach.

Paper's Novelty and Contribution: FDIA modifies the sensory data such that the sensory data stream of another user is injected into the target user's sensory stream (victim), so no knowledge of the victim is assumed. If the attacker wants to generate a specific output, they can use signals with targeting information to influence the application's outcome. This paper particularly focuses on such targeting FDIAs aiming to misinform human-centric sensing applications with certain forged information.
However, several continuous authentication approaches have been proposed in the literature for human sensing applications. Hence, this paper's threat model considers a harder situation where a continuous authentication system is working in the background to protect the smart device. Other sensing applications can further accept only the signal verified by the authentication system. Hence, a successful FDIA sample must deceive the authentication system into thinking it is an authentic signal and the human-centric application to generate a misled inference. We evaluated a wide variety of FDIA samples and showed that an equal proportion of forged and legit signals mixed could deceive both the authentication and human event detection approaches.

Notably, we developed a novel FDIA detection approach from a single sensor stream. Literature shows that, in human-centric sensing, the sensory streams convey the user's unique traits. This paper's novelty comes from formulating the FDIA detection problem as a multiple instance learning (MIL) problem. The FDIA detection identifies if a signal sample (subject to inspection) comprises at least a pair of segments belonging to different individuals. This paper performs this task through a novel framework named Siamese-MIL that leverages the MIL paradigms, Siamese network structure, and a unique sensor data representation. The Siamese-MIL segments a signal sample into a set of all possible segment pairs. A Siamese neural network (SNN) is trained to identify any segment pair that contains signals from different individuals. If the trained SNN identifies at least one mismatched segment pair in a set, the respective signal sample is considered an FDIA sample. The approach differs from supervised learning or voting mechanisms on how the SNN is trained. Following the MIL training paradigm, SNN is tailored to be highly effective in detecting matched (legit) segment pairs (i.e., high recall) where mismatched segment pair detection accuracy (true negative rate) can be lower. Such characteristics ensure high FDIA detection accuracy. The Siamese-MIL FDIA detection approach is designed to extend the conventional authentication systems, prohibiting any attack signal from reaching human-centric applications. Importantly, the approach is generalizable to detect FDIA in any single human-centric sensing data stream.

Professional Significance: This paper was published in csranking.org listed top conference Ubicomp/IMWUT 2022. The work is funded by my NSF SCH Medium grant, and my Ph.D. student Jingyu Xin is the first author, and I am the corresponding author. I am preparing an NSF SaTC proposal based on this work.

[P4] Fatih Altay, Guillermo Ramón Sánchez, Yanli James, Stephen V. Faraone, Senem Velipasalar, **Asif Salekin**. Preclinical Stage Alzheimer's Disease Detection Using Magnetic Resonance Image Scans, The Thirty-Third Annual Conference on Innovative Applications of Artificial Intelligence (IAAI), 2021. https://doi.org/10.1609/aaai.v35i17.17772

Problem and Motivation: Alzheimer's is a type of brain disease mostly seen in older age and starts without showing any symptoms. It is still not clear when or how AD begins. To fully understand the effects of AD, it is important to investigate the disease since its beginning. There are three stages of AD known today. The first stage is referred to as the preclinical stage, which is the main focus of this work. The preclinical stage is when neurons start degenerating, even if there are no visible symptoms. Notably, detecting Alzheimer's disease in its early stages (i.e., Preclinical), even if there are no visible symptoms, is an important task, especially when the potential benefits on human life and the economy is considered. In this paper, we propose two attention models for detecting Alzheimer's disease, in the earliest (preclinical) stage, from 3D MRI images.

Paper's Novelty and Contribution: Magnetic resonance imaging (MRI) of the brain produces detailed 3-dimensional (3D) images of the brain. MRI imaging has allowed doctors and researchers to investigate brain structure and brain-related diseases and has been leveraged by recent AI and deep-learning-based studies. One of the main differences between our proposed approach and the existing studies is that prior works address the detection of Alzheimer's disease in the second and third stages when the effects and symptoms of the disease and the changes in brain neurons are identifiable. However, the detection of preclinical AD (first stage) is one of the important goals for the medical community. Thus, we have focused on and developed a novel method for detecting AD in the first (preclinical) stage, even when there are no visible symptoms. Notably, we developed attention-based mechanisms by leveraging transformer and 3D recurrent visual attention models. Hence, the resulting approach can visualize the brain regions indicative of preclinical AD, which is significant for practical use by the medical community. According to our extensive evaluations, the developed approach can detect Preclinical AD about 8-12 years before any perceivable symptoms by the patient.

Professional Significance: The paper was published in the csranking.org listed top conference IAAI 2021. I was the corresponding author, and the paper was co-first authored by a master's student Guillermo Ramón Sánchez and a Ph.D. student Fatih Altay, under my mentorship. Notably, the paper received the prestigious '**IAAI Deployed Application Award**' at the IAAI conference.

[P5] Tianjia He, Lin Zhang, Fanxin Kong, and **Asif Salekin**. Exploring Inherent Sensor Redundancy for Automotive Anomaly Detection, The 57th Design Automation Conference (DAC), 2020. https://doi.org/10.1109/DAC18072.2020.9218557

Problem and Motivation: Modern automobiles follow open architectures enabling various promising services and applications such as vehicle-to-vehicle communication and self-driving. However, it also introduces potential security vulnerabilities to malicious attacks that are beyond conventional cyber-attacks, such as spoofing attacks on GPS sensors, noninvasive attacks on Antilock Braking Systems, cameras, and LiDAR systems, etc. These attacks can spoof the automobile controllers to perform actions that may result in adverse outcomes. Hence, there is a need to validate sensor data before the automobile controller act on them. This paper addresses the task of detecting anomalous sensory streams by exploiting inherent redundancy among heterogeneous sensors.

Paper's Novelty and Contribution: The paper's threat model considers that the attacker can maliciously alter a subset of all sensors and control their measurements given to the controller, and the training data is trustful.

This paper investigates and leverages the inherent redundancy among heterogeneous sensors to detect anomalies. Inherent sensor redundancy is defined as multiple sensors simultaneously responding to the same physical aspect in a related manner. For example, pressing the accelerator will increase engine RPM and vehicle speed as well as affect GPS readings.

The developed anomaly detection approach's primary idea is first to identify the consistency among sensor data and then utilize it to detect anomalous behaviors of sensor measurements. To realize this idea, we developed a deep autoencoder-based anomaly detection method.

The developed deep-autoencoder learns a consistent pattern from vehicle sensor data in normal states and utilizes it as the nominal behavior for the detection. We define a threshold based on the reconstruction error (of the autoencoder), where corrupted sensor measurements will result in higher reconstruction errors than the threshold, while normal data will not. The approach only relies on normal sensor data and does not restrict the existence of multiple clusters in the training data set. The proposed method is independent of anomalous data for training and the calculation of pairwise correlation among sensors. We use a real-world dataset to demonstrate the feasibility of our approach.

The developed approach is generalizable and extendible on multi-sensor systems, where multiple sensors measure the same event, conveying inherent redundancy.

Professional Significance: The paper was published in csranking.org listed top-tier conference DAC 2020. This paper was first authored by Tianjia He, a master's student at Syracuse University. After joining Syracuse University, I co-mentored the student with Dr. Fanxin Kong, and this is the first paper resulting from my mentorship of a Syracuse University student. The paper received 20 citations.