

# Computer Vision Estimation of Stress and Anxiety Using a Gamified Mobile-based Ecological Momentary Assessment and Deep Learning: Research Protocol

---

---

Ali Kargarandehkordi\*, [kargaran@hawaii.edu](mailto:kargaran@hawaii.edu)  
Peter Washington\*, [pyw@hawaii.edu](mailto:pyw@hawaii.edu)

\*Department of Information and Computer Sciences  
University of Hawai'i at Mānoa

---

## Abstract

---

Stress and anxiety can contribute to the development of major health problems such as heart disease, depression, and obesity. Due to its subjective nature, it is challenging to precisely measure human affect by relying on automated approaches. We therefore propose a personalized prediction framework fine-tuned for each participant in lieu of the traditional “one-size-fits-all” machine learning approach. We aim to collect such individualized data via two distinct procedures: 1) a smartphone-based ecological momentary assessment of stress, and 2) Zoom calls. The data collected from these periodic self-reports will include selfie photographs and ecological momentary assessments of affect. To enhance user engagement during the data collection process, we propose the use of gamification, an emerging trend which involves influencing user behavior and lifestyle by incorporating fun and engaging game elements into non-game contexts (e.g., health-related tasks). In addition to developing a standardized platform to achieve more accurate recognition of stress and anxiety, we plan to conduct a concurrent study in which we will capture videos of our subjects undertaking the Stroop Color Word and Amygdala Test and analyze the footage to identify additional significant characteristics that relate to anxiety. This could include features such as head and mouth movements, lip and cheek deformations, eye gaze, and blinking rates. The final results will provide a comparative evaluation of both objective measures of stress. This research project was approved by the University of Hawaii Institutional Review Board.

**Keywords:** anxiety, stress, machine learning, gamification, computer vision, mHealth

## Introduction

---

Consistent exposure to stress and repetitive feelings of anxiety can lead to detrimental impacts on life quality and expectancy [1]. Several of the leading causes of death (e.g., cardiovascular disease) are linked to these adverse mental states [2, 3]. Despite this large burden on the health system, chronic stress and anxiety are largely left undertreated [4, 5]. The recognition of stress and anxiety is a highly subjective endeavor, leading to challenges with quantifying such affects in an objective manner. Generally, stress and anxiety are measured through clinical questionnaires designed by psychological experts [6].

In certain cases, stress and anxiety have been documented to produce a noticeable effect on facial characteristics, including modifications on pupil size, eyelid closure, blinking rate, mouth activity, lip deformation, cheek deformation, and head movement[7-9]. However, this physical manifestation varies significantly across individuals. Therefore, developing an accurate single-purpose machine learning (ML) model to recognize stress and anxiety is likely outside of the realm of feasibility.

To overcome this challenge, we propose the utilization of a personalized ML approach, where an individual ML model is fine-tuned to each individual participant's data [10, 11]. Such personalization can provide an opportunity to eventually develop digital therapeutics which intervene when stress or anxiety is recognized.

In this protocol documentation, we first present a survey of methods in the field of digital mental health. We then discuss a data collection protocol for curating a dataset of facial selfies and corresponding labels of stress and anxiety through an ecological momentary assessment (EMA). We next discuss a series of personalized ML experiments which we will conduct using the resulting dataset.

## Prior Works

---

### Conventional Methods for Stress and Anxiety Measurement

There are two commonly used methods to assess anxiety disorders: (1) analyzing data collected through clinician interviews and (2) self-reports using standardized questionnaires [12]. These methods involve asking patients to report on the severity and intensity of their anxiety symptoms. However, these measurements are subjective and may result in inconsistent or unreliable data [13] due to numerous factors including the patient's inability to recall information accurately and discomfort with the clinician [14]. Other factors that can compromise data reliability include question ambiguity, psychopathology, and biases [15, 16].

Another way to measure stress and anxiety is to evaluate cortisol levels produced by the HPA axis. This involves an intensive analyzation of cortisol from different sources such as blood, urine, hair, and saliva [17, 18]. Other physiological indicators include assessing brain activity using EEG, heart activity using ECG, muscle activity using EMG, and skin conductance using electrodermal response. Changes in breathing patterns and speech characteristics can also indicate anxiety [6, 19, 20]. Researchers are particularly interested in changes in fundamental frequencies of the voice as a potential indicator of anxiety [7, 21].

### Facial Signs and Expressions

Charles Darwin's book "The Expression of the Emotions in Man and Animals" in the late 1800s first highlighted the connection between emotional states and facial expressions [22]. Later studies in the 1900s and early 2000s provided more evidence to support this connection, particularly with advanced methods like HPA axis and cardiovascular responses to stress. Recent studies suggest that facial signs can provide insights into classifying

anxiety[7, 8]. Anxiety can affect various facial expressions including eyes (pupil size variation, eye aperture, blinking rate, and gaze distribution), mouth activity [23], specific deformations on the lips [24], cheeks, head movements and velocity [25, 26]. However, covarying factors such as lying, depression, Parkinson's [27], schizophrenia [28], and environmental conditions like lighting and humidity can also affect facial expressions, particularly blinking rates [29].

## **Machine Learning for Affective Computing**

Machine learning (ML) models are increasingly applied to healthcare for various purposes such as therapy planning, disease diagnosis, and illness prediction. These tools have been shown to improve the quality and efficiency of medical care and are considered promising for medical experts [30, 31]. Although machine learning models have been successfully used in mental health to predict psychological and perceived anxiety, there are challenges related to reliability due to limitations in data collection [32], limited research-only resources [33], confidentiality of the data and inability to disclose it [34], and lack of effective commercial licenses [35]. These challenges make it difficult to use machine learning models in real healthcare scenarios.

Due to the inherent heterogeneity and subjectivity of stress, recognizing stress and anxiety typically involves using multiple forms of data. Analyzing and merging these multimodal data can help create a reliable and stable stress and anxiety recognition models. There are various ways to measure stress, including using stressors like personal factors, environment, and individuals' physiological, psychological, and physical reactions to stressors. Eye data and vehicle dynamics data are easily accessible, and some researchers have explored their connection to stress [36]. Past research has mostly used conventional machine learning techniques to manually extract features from data and create stress recognition models. Although this method has yielded favorable results, it can be difficult to accurately identify important and representative features, and it requires specialized knowledge and is less resilient to noise and data variations [37]. Some researchers have developed stress level recognition models using multimodal data. Benoit et al. [38] developed a driver simulator that utilized video data (facial activity) and physiological data, such as blinking, yawning, and head turning, as well as ECG and electrical skin responses to measure the driver's stress levels. Rigas et al. [39] developed a model that used physiological signals, video recordings (eye data, head movement), and environmental data to classify driver stress levels and predict driving performance. The model achieved an 86% accuracy rate in distinguishing between two stress levels (no stress, stress) using an SVM classifier.

Generalized machine learning models have limitations in recognizing emotional states due to high variability in physical and personal characteristics among individuals, leading to poor performance [40]. Personalized machine learning, by contrast, has been suggested as a more effective technique for precision health machine learning with high-dimensional data on subjective and person-specific prediction classes [40, 41]. This approach involves further research on affective variations within the individual, where each model is trained and tested

solely on an individual's data to examine the effects of variability on recognizing emotional transitions and states.

### **Smart-sensing and Passive Monitoring**

Observing and evaluating patients' behavior as an indicator of changes in the severity levels of mental disorder symptoms can be an alternative to traditional methods. Using mobile phones as sensing devices enables researchers to passively record specific behaviors related to different mental disorders. Mobile phones are useful tools due to their various sensors and software for evaluating behaviors and contextual awareness as well as their constant presence with the owner [42-44]. In addition to subjective information from interviews or self-reports, mobile phones can also provide objective data through their camera, allowing researchers to make objective assessments of mental health [45].

### **Reminders and Alerts**

Adding alert-based components such as reminders to digital health interventions can improve engagement and encourage adherence to the intervention. These reminders can be sent through various channels, including email, SMS, or push notifications within mobile applications. Incorporating reminders has been shown to increase engagement levels in mental healthcare research, making it more effective than interventions without this feature [46-48]. Using reminders in digital health interventions can also lead to higher levels of adherence to the research and lower non-usage attrition, which are common issues with online-based interventions [49-51]. Non-usage attrition refers to participants who stop using the digital intervention while still participating in the research through other protocols, such as filling out questionnaires.

### **Gamification**

Gamification, which involves using game design elements in non-game contexts, is another popular trend to improve the user engagement and activities in digital interventions. In an academic context, gamification refers to using gamified elements and incentives, along with visually appealing designs, to enhance user engagement, motivation, and adherence to research activities [52-54]. Gamification has been implemented in various healthcare services to increase user activity and productivity of behaviors, using different elements such as points, leaderboards, challenges, feedback [55-60], rewards, story/theme [61, 62], achievements/badges [55], progress tracking, and levels [58, 60].

### **Study Procedure and Methodologies**

---

The Ecological Momentary Assessment (EMA), also known as the Experience Sample Method or Ambulatory Assessment, is a useful tool for classifying momentary moods in natural and real-life environments [63-65]. This method allows individuals to report their current mental state in real and familiar environments rather than in unfamiliar settings through standardized instruments [65-67]. EMA involves regular self-reporting of a psychiatric property while carrying out daily tasks over a specified period of time. Digital healthcare

interventions can utilize EMA, which has methodological strengths including increased ecological validity, reduced retrospective bias and errors, and the ability to collect data periodically across different real-life situations [65].

This study will use both a smartphone-based EMA application (which we call STAND) and Zoom calls to measure stress. The first procedure involves assessing stress levels by asking study participants to take daily selfies and self-report their stress and anxiety levels as well as a description of how they feel at the moment (including any symptoms related to stress/anxiety). In this procedure, the data collected will consist of both images and text, which we will use to train single modality and multimodal deep learning models for predicting stress and anxiety levels. We will evaluate the extent to which each modality improves the accuracy of the prediction. The deep learning CNN/LSTM network architecture we intend to employ for training the model in the first procedure is depicted in Figure 2.

During the second phase of the study, participants will be required to participate in periodic Zoom meetings at predetermined intervals throughout the one-month data collection period. These meetings will involve engaging in psychological tests such as the "Stroop Color Word Test" and "Amygdala Test," which are designed to be game-like in nature. Participants will be recorded during these tests, and the videos will be carefully analyzed to identify any additional facial expressions that may be associated with stress and anxiety.

Based on an EMA based intervention, this study intends to utilize all the techniques mentioned in the Prior Works section to efficiently recognize stress, a commonly experienced emotional state. The study will involve collecting data from participants for a month and then analyzing the data to identify the key features associated with accurately recognizing stress and anxiety. The experiment will be conducted in three phases, which are shown in Figure 1. Advantages of using an EMA-based mobile application for digital intervention in the healthcare field include greater availability, accessibility, anonymity, flexibility, cost-effectiveness, and interactivity through visually attractive designs [68-70]. We aim to explore the impact of incorporating personalized app modifications such as animations and informal language in increasing user engagement with the digital intervention. An example of this is using an interactive question related to a user's preference for dogs or cats as part of the app's registrations process (a dog person, or a cat person?). The app selects and arranges animations in various sections based on the user's response to this question.

Under an expedited review procedure, this research project was approved on December 20, 2022, by the University of Hawaii Institutional Review Board (UH IRB). The application qualified for expedited review under 45 CFR 46.110, Category 6, 7. Per 45 CFR 46.109. This study targets students as the main subject for analysis and focuses to screen stress among this population. Data associated with the students are prone to challenges resulting from barriers such as lack of time, stigma, scheduling issues, and significant workloads [71]. Stress is among the major pervasive mental health concerns within the school community, and many students around the globe undergo increased levels of stress in their everyday life [72, 73]. We aim to recruit at least 20 students (gender balanced) aged  $28 \pm 10$  years.

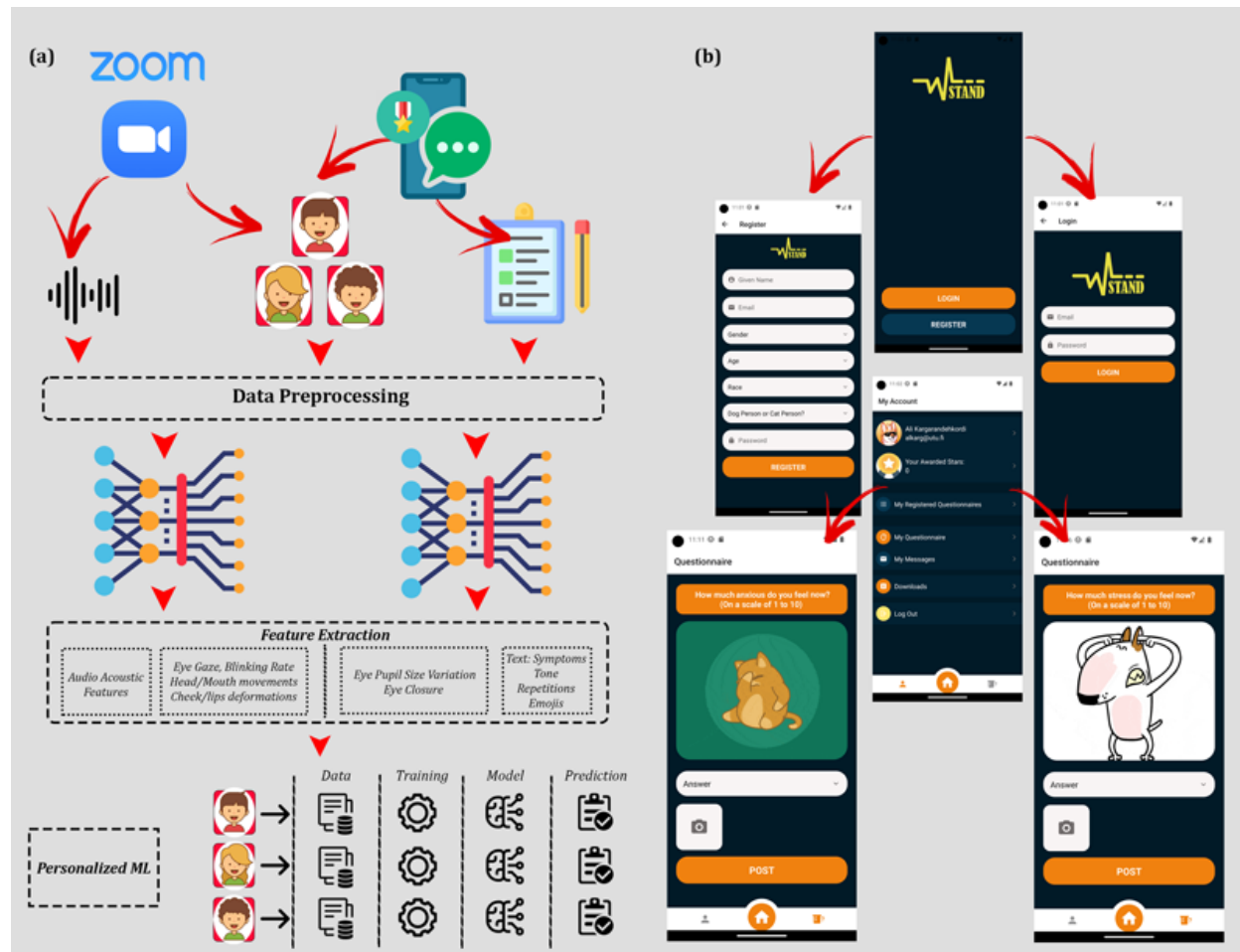


Figure 1: a) Workflow of the study. b) Flow of different screens in the STAND app.

## Stage 1: Mobile Application

A mobile application that can work on both Android and iOS platforms has been developed for the purpose of investigating the development of a subjective measure of stress. Study participants will receive questionnaires through daily notifications on the STAND app, in which there is an option for taking selfies. The photographs will be used to analyze potential stress related facial expressions (objective measurement). Initially, users will receive two separate notifications throughout the day that inquire about their levels of stress and anxiety. There will be a space provided where participants can input their current emotions and any related symptoms they may be experiencing such as tiredness, butterfly in the stomach, lack of focus etc.

To introduce game-like aspects to the study, a system of rewarding stars has been included in the app. The participants will be awarded stars according to the number of points they obtained for their enthusiastic and constructive involvement during the data gathering stage, such as completing assignments within 30 minutes of receiving a notification and providing complete responses to all questions each day during the initial phase. Participants who earn

more stars will have a higher chance of winning rewards, such as grocery shopping gift cards, through a fair lottery drawing at the end of the data collection phase.

## **Stage 2: Ecological Momentary Assessment Study and Periodic Zoom Calls**

In the second stage of the study, the process mentioned above will be carried out for one month to gather adequate data needed for creating personalized machine learning algorithms. To evaluate relevant psychometric properties in the third stage, participants will be requested to complete a self-report questionnaire every day. The questionnaire consists of two questions and will be distributed as multiple notifications throughout the day, with each notification containing one question. Every time participants receive a notification; they will need to respond to the same questions they have answered in the previous days on a scale ranging from 1 to 10:

- How much stress do you feel right now?
- How much anxiety do you feel right now?

In a second concurrent procedure, participants will also attend periodic Zoom meetings (at fixed intervals within the one-month data collection experiment) to take part in a game-like psychological examination known as “Stroop Color Word Test” and “Amygdala Test” while they are recorded. The recorded videos will be scrutinized to identify additional facial expressions associated with stress such as eye gaze distribution, blinking rates, lips and cheeks deformations, mouth activity, and head movements. Extracted audio signals from recorded videos will also be analyzed alongside facial expressions in a multi modal deep learning network (Figure 3).

During this process, individuals who have already been diagnosed with anxiety will be requested to take part in the two psychological tests mentioned earlier. The amygdala test requires the participants to search through arrays of faces with a specific emotional intensity level. Previous research has suggested that anxious individuals are quicker to detect angry facial expressions than positive ones. [74]. Therefore, individuals with anxiety tend to exhibit an increased attentional bias towards cues that indicate any form of threat in comparison to individuals without anxiety. [75]

## **Stage 3: Deep Phenotyping of Stress and Anxiety**

In the third and final phase, we will train personalized deep learning methods to recognize stress and anxiety. During the first procedure, two modalities will be utilized both individually and concurrently via separate network architectures to determine the marginal impact of each modality on prediction accuracy. These modalities consist of facial expressions taken from daily selfies and text descriptions provided by the participants concerning their momentary feeling. The overall CNN/LSTM network architecture of the model that will be employed to train each modality is displayed in Figure 2. Alongside a convolutional network, LSTMs [76] will be used to capture temporal relationships. LSTMs are useful for capturing long-range dependencies in input sequences such as videos [77]. The

LSTM's output will be transmitted into a fully-connected network layer, which will then be transmitted into another dense layer. This dense layer will then be forwarded to a final dense layer that utilizes softmax activation to estimate stress and anxiety levels.

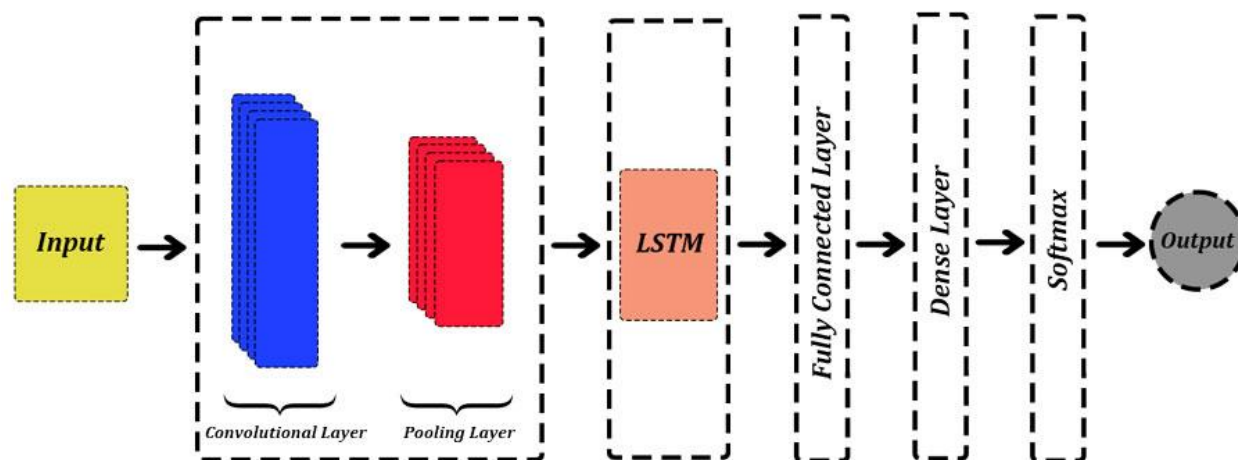


Figure 2: Neural network architecture of the proposed single modality CNN/LSTM network.

In addition to an enhanced subjective and objective measurement of stress (first procedure), this work presents a method for recognizing stress in students using a personalized deep learning framework that incorporates audio and facial expressions extracted from Zoom calls to predict self-reported stress and anxiety. The proposed multimodal framework involves four main steps: preprocessing, feature extraction, feature fusion, and classification (Figure 3).

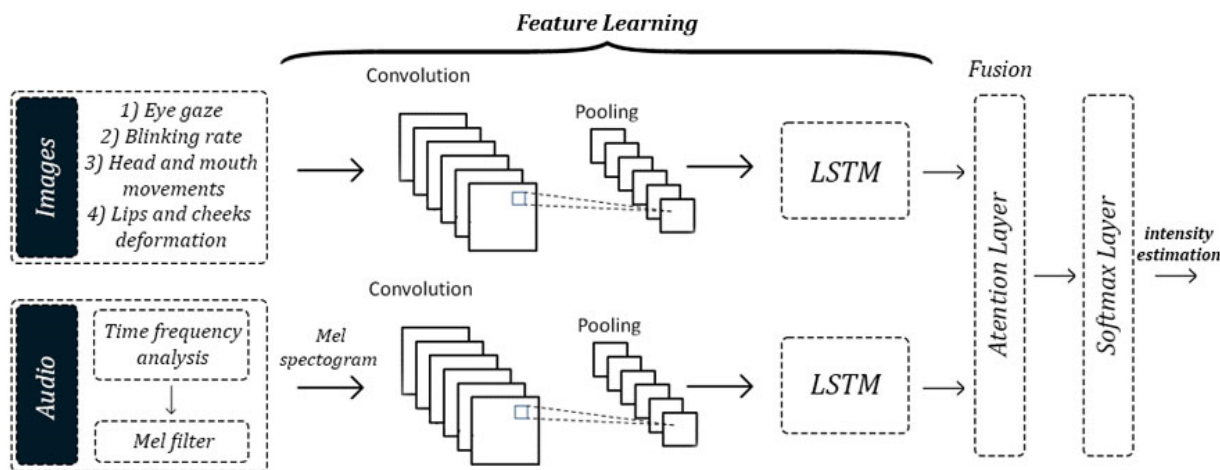


Figure 3: Neural network architecture of the proposed CNN/LSTM multimodal network.

The model will extract stress-related information by fusing data from multiple sources, including facial expressions (e.g., eye, head, cheek, mouth, and lips features), and audio signals. The voice data will be transformed into a time-frequency representation, and the facial expressions are extracted from photographs and video frames. The multidimensional features of each modality will then be extracted using deep learning techniques. The fully



connected layers in the framework obtain information about the stress state, which is fused into a global matrix for stress and anxiety recognition and intensity level. The ResNet50 and I3D models will be used for audio and facial expression feature extraction, respectively, with the I3D model incorporating a temporal attention module to identify important frames. The proposed multimodal fusion model aims to classify stress and anxiety levels in students in real time using multiple sources of data.

This study features a personalized analysis approach that utilizes deep learning to learn different features associated with a person's subjective symptoms, head and mouth movements, deformations of cheeks and lips, eye gaze, variations in pupil size, and blink rate. This approach enables the models to be overfitted to individual human subjects (i.e., “personalized machine learning”) by self-supervised learning. By analyzing the top predictors in each individual, we can identify the distinct features that contribute to their stress and anxiety over time. All the features used in the personalized models can be targeted for intervention. Therefore, the results of the personalized models can be used to create individualized interventions with a single or multiple feature-based approach, which can be used to recognize stress and anxiety levels using a personalized multimodal strategy. This method creates a unique model for each patient that is specifically designed to classify their stress and anxiety levels, and it will only be used for that particular patient when deployed into clinical contexts.

## **Conclusion**

Previous research studies have shown that mental health applications can be an effective tool for increasing user engagement even without incorporating gamification into their design. We anticipate that the STAND app will result in greater user participation. It will be beneficial to explore the addition of more personalized and visually appealing design elements. By incorporating a second concurrent procedure into this research protocol and analyzing video frames facial expression and audio signals, we will have the opportunity to examine and compare the performance of our smartphone-based EMA application with another robust procedure that both can objectively measure the intensity of stress and anxiety. The use of personalized machine learning techniques in digital health has unlocked promising opportunities for creating and optimizing models in the field of mental health diagnosis, management, and treatment. The results of this study will compare the accuracy of both subjective and objective measurements of stress and anxiety intensity, which can serve as a resource for future studies.

## **Discussion and Future Works**

The proposed study aims to provide evidence for the efficacy of using state-of-the-art technologies (i.e., ecological momentary assessment and personalized machine learning) to classify stress and anxiety levels among young adults. Combining these technologies to make digital mental health interventions might be a promising alternative to the current standard of care. Although we will apply novel techniques to quantify anxiety and identify the potential relationships between facial cues for anxiety and stress, there are still a few

technologies that need to be investigated and combined in future works to enhance the efficiency of such recognition mechanisms.

Crowdsourcing is one of these technologies that has emerged in the field of digital phenotyping. Judging and validating subjective details often requires human intervention for potential improvements. Innovations in this field include algorithms for data efficiency and complexity management [78, 79] and debugging and verification of AI models [80, 81].

Another solution to increase the efficiency of mental health interventions lies in modern smartphones and their sensors. The prevalent usage of smartphones in society, and the fact that they became an integral part of individuals' lives represents a new frontier technology in digital health. Physical activities approximated by acceleration sensors[82], duration of stays and their specific time of the day at known locations approximated by Wi-Fi and GPS information [83], and the frequencies of interactions and social activities [84] can be highly related to the individual's mental health state and can be screened and recorded passively benefiting from these portable devices.

Our plan is to develop the STAND app into a versatile tool that can be customized for researching and gathering computer vision data of the face to build personalized machine learning models which can support digital health interventions related to mental states and developmental disorders. This work can integrate with existing research in using automatic emotion recognition for a variety of contexts, including mental illness diagnosis, recognizing human social and physiological interactions, and developing sociable robotics and other human-computer interaction systems [57, 85-89]. For example, emotional expressions have a crucial role in recognizing certain types of developmental disorders. Autism spectrum disorder (ASD) affects almost 1 in 44 people in America [90] and it is the fastest-growing developmental disorder in the United States [91, 92]. Children with autism exhibit emotions differently than their neurotypical peers, and they often struggle to display appropriate facial expressions. [93-95]. In order to enhance social communication, it is anticipated that creating an optimal model for human-computer interactions would involve accurately classifying and responding to human emotions e.g., stress and anxiety [96-105]. Initial attempts have been made to utilize digital and wearable devices to enable families to conduct therapy sessions in the comfort of their own home. The intervention structure can be tailored to meet the specific needs of the child [106-112]. The personalized affective AI proposed here could be integrated into these interventions to provide enhanced digital and mobile therapies to children with ASD and other developmental delays.

## References

---

- [1] E. S. Epel *et al.*, "Accelerated telomere shortening in response to life stress," *Proc Natl Acad Sci U S A*, vol. 101, no. 49, pp. 17312-5, Dec 7 2004, doi: 10.1073/pnas.0407162101.
- [2] R. S. Lazarus, *Psychological stress and the coping process* (McGraw-Hill series in psychology). New York: McGraw-Hill (in English ; Undetermined), 1966.

- [3] J. Tomaka, J. Blascovich, R. M. Kelsey, and C. L. Leitten, "Subjective, physiological, and behavioral effects of threat and challenge appraisal," *Journal of personality and social psychology*, vol. 65, no. 2, p. 248, 1993.
- [4] L. Welty, A. Harrison, K. Abram, N. Olson, D. Aaby, and K. McCoy, "Substance abuse and mental health services administration.(2017). Key substance use and mental health indicators in the United States: results from the 2016 national survey on drug use and health (HHS Publication No. SMA 17-5044, NSDUH Series H-52). Rockville, MD: Center for Behavioral Health Statistics and Quality," *Substance Abuse and Mental Health Services Administration. Retrieved. College of Health Sciences*, vol. 106, no. 5, p. 128, 2019.
- [5] D. Chisholm *et al.*, "Scaling-up treatment of depression and anxiety: a global return on investment analysis," *The Lancet Psychiatry*, vol. 3, no. 5, pp. 415-424, 2016.
- [6] S. Greene, H. Thapliyal, and A. Caban-Holt, "A survey of affective computing for stress detection: Evaluating technologies in stress detection for better health," *IEEE Consumer Electronics Magazine*, vol. 5, no. 4, pp. 44-56, 2016.
- [7] N. Sharma and T. Gedeon, "Objective measures, sensors and computational techniques for stress recognition and classification: A survey," *Computer methods and programs in biomedicine*, vol. 108, no. 3, pp. 1287-1301, 2012.
- [8] N. Sharma and T. Gedeon, "Modeling observer stress for typical real environments," *Expert Systems with Applications*, vol. 41, no. 5, pp. 2231-2238, 2014.
- [9] D. F. Dinges *et al.*, "Optical computer recognition of facial expressions associated with stress induced by performance demands," *Aviation, space, and environmental medicine*, vol. 76, no. 6, pp. B172-B182, 2005.
- [10] A. Halaris, "A primary care focus on the diagnosis and treatment of major depressive disorder in adults," *Journal of Psychiatric Practice*<sup>®</sup>, vol. 17, no. 5, pp. 340-350, 2011.
- [11] R. V. Shah *et al.*, "Personalized machine learning of depressed mood using wearables," *Translational psychiatry*, vol. 11, no. 1, pp. 1-18, 2021.
- [12] E. Andreou *et al.*, "Perceived stress scale: reliability and validity study in Greece," *International journal of environmental research and public health*, vol. 8, no. 8, pp. 3287-3298, 2011.
- [13] G. J. Meyer *et al.*, "Psychological testing and psychological assessment: A review of evidence and issues," *American psychologist*, vol. 56, no. 2, p. 128, 2001.
- [14] L. H. Rogler, R. G. Malgady, and W. W. Tryon, "Evaluation of mental health: Issues of memory in the Diagnostic Interview Schedule," *The Journal of nervous and mental disease*, vol. 180, no. 4, pp. 215-222, 1992.
- [15] T. F. Van de Mortel, "Faking it: social desirability response bias in self-report research," *Australian Journal of Advanced Nursing, The*, vol. 25, no. 4, pp. 40-48, 2008.
- [16] O. P. John and R. W. Robins, "Accuracy and bias in self-perception: individual differences in self-enhancement and the role of narcissism," *Journal of personality and social psychology*, vol. 66, no. 1, p. 206, 1994.
- [17] E. Russell, G. Koren, M. Rieder, and S. Van Uum, "Hair cortisol as a biological marker of chronic stress: current status, future directions and unanswered questions," *Psychoneuroendocrinology*, vol. 37, no. 5, pp. 589-601, 2012.
- [18] D. H. Hellhammer, S. Wüst, and B. M. Kudielka, "Salivary cortisol as a biomarker in stress research," *Psychoneuroendocrinology*, vol. 34, no. 2, pp. 163-171, 2009.
- [19] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Tröster, and U. Ehlert, "Discriminating stress from cognitive load using a wearable EDA device," *IEEE Transactions on information technology in biomedicine*, vol. 14, no. 2, pp. 410-417, 2009.

- [20] A. de Santos Sierra, C. S. Ávila, J. G. Casanova, and G. B. Del Pozo, "A stress-detection system based on physiological signals and fuzzy logic," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 10, pp. 4857-4865, 2011.
- [21] P. Wittels, B. Johannes, R. Enne, K. Kirsch, and H.-C. Gunga, "Voice monitoring to measure emotional load during short-term stress," *European journal of applied physiology*, vol. 87, no. 3, pp. 278-282, 2002.
- [22] C. Darwin and P. Prodger, *The expression of the emotions in man and animals*. Oxford University Press, USA, 1998.
- [23] W. Liao, W. Zhang, Z. Zhu, and Q. Ji, "A decision theoretic model for stress recognition and user assistance," in *AAAI*, 2005, vol. 2005, pp. 529-534.
- [24] D. Metaxas, S. Venkataraman, and C. Vogler, "Image-based stress recognition using a model-based dynamic face tracking system," in *International Conference on Computational Science*, 2004: Springer, pp. 813-821.
- [25] U. Hadar, T. J. Steiner, E. C. Grant, and F. Clifford Rose, "Head movement correlates of juncture and stress at sentence level," *Language and speech*, vol. 26, no. 2, pp. 117-129, 1983.
- [26] A. Adams, M. Mahmoud, T. Baltrušaitis, and P. Robinson, "Decoupling facial expressions and head motions in complex emotions," in *2015 International conference on affective computing and intelligent interaction (ACII)*, 2015: IEEE, pp. 274-280.
- [27] J. Mackintosh, R. Kumar, and T. Kitamura, "Blink rate in psychiatric illness," *The British Journal of Psychiatry*, vol. 143, no. 1, pp. 55-57, 1983.
- [28] R. C. Chan and E. Y. Chen, "Blink rate does matter: a study of blink rate, sustained attention, and neurological signs in schizophrenia," *The Journal of nervous and mental disease*, vol. 192, no. 11, pp. 781-783, 2004.
- [29] G. W. Ousler III, K. W. Hagberg, M. Schindelar, D. Welch, and M. B. Abelson, "The ocular protection index," *Cornea*, vol. 27, no. 5, pp. 509-513, 2008.
- [30] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255-260, 2015.
- [31] D. D. Miller and E. W. Brown, "Artificial intelligence in medical practice: the question to the answer?," *The American journal of medicine*, vol. 131, no. 2, pp. 129-133, 2018.
- [32] H. Suominen, L. Zhou, L. Hanlen, and G. Ferraro, "Benchmarking clinical speech recognition and information extraction: new data, methods, and evaluations," *JMIR medical informatics*, vol. 3, no. 2, p. e4321, 2015.
- [33] D. I. R. Boyle and N. Rafael, "BioGrid Australia and GRHANITE™: Privacy-protecting subject matching," in *Health Informatics: The Transformative Power of Innovation*: IOS Press, 2011, pp. 24-34.
- [34] D. S. Char, N. H. Shah, and D. Magnus, "Implementing machine learning in health care—addressing ethical challenges," *The New England journal of medicine*, vol. 378, no. 11, p. 981, 2018.
- [35] D. Graff and C. Cieri, "English gigaword, linguistic data consortium," ed, 2003.
- [36] H. Boril, S. Omid Sadjadi, T. Kleinschmidt, and J. Hansen, "Analysis and detection of cognitive load and frustration in drivers' speech," in *Proceedings of the 11th annual conference of the international speech communication association*, 2010: International Speech Communication Association, pp. 502-505.
- [37] B. Nakisa, M. N. Rastgoo, D. Tjondronegoro, and V. Chandran, "Evolutionary computation algorithms for feature selection of EEG-based emotion recognition using mobile sensors," *Expert Systems with Applications*, vol. 93, pp. 143-155, 2018.
- [38] A. Benoit *et al.*, "Multimodal focus attention and stress detection and feedback in an augmented driver simulator," *Personal and Ubiquitous Computing*, vol. 13, pp. 33-41, 2009.

- [39] G. Rigas, Y. Goletsis, P. Bougia, and D. I. Fotiadis, "Towards driver's state recognition on real driving conditions," *International Journal of Vehicular Technology*, vol. 2011, 2011.
- [40] M. Sultana, M. Al-Jefri, and J. Lee, "Using machine learning and smartphone and smartwatch data to detect emotional states and transitions: exploratory study," *JMIR mHealth and uHealth*, vol. 8, no. 9, p. e17818, 2020.
- [41] A. Hart, D. Reis, E. Prestele, and N. C. Jacobson, "Using Smartphone Sensor Paradata and Personalized Machine Learning Models to Infer Participants' Well-being: Ecological Momentary Assessment," *Journal of Medical Internet Research*, vol. 24, no. 4, p. e34015, 2022.
- [42] G. M. Harari, S. R. Müller, M. S. Aung, and P. J. Rentfrow, "Smartphone sensing methods for studying behavior in everyday life," *Current opinion in behavioral sciences*, vol. 18, pp. 83-90, 2017.
- [43] R. Mafrur, I. Nugraha, and D. Choi, "Modeling and discovering human behavior from smartphone sensing life-log data for identification purpose," *Human-centric Computing and Information Sciences*, vol. 5, no. 1, pp. 1-18, 2015.
- [44] L. Pei *et al.*, "Human behavior cognition using smartphone sensors," *Sensors*, vol. 13, no. 2, pp. 1402-1424, 2013.
- [45] J. Asselbergs, J. Ruwaard, M. Ejdy, N. Schrader, M. Sijbrandij, and H. Riper, "Mobile phone-based unobtrusive ecological momentary assessment of day-to-day mood: an explorative study," *Journal of medical Internet research*, vol. 18, no. 3, p. e5505, 2016.
- [46] A. E. Whitton *et al.*, "Breaking open the black box: isolating the most potent features of a web and mobile phone-based intervention for depression, anxiety, and stress," *JMIR mental health*, vol. 2, no. 1, p. e3573, 2015.
- [47] C. E. Branson, P. Clemmey, and P. Mukherjee, "Text message reminders to improve outpatient therapy attendance among adolescents: a pilot study," *Psychological services*, vol. 10, no. 3, p. 298, 2013.
- [48] G. Clarke *et al.*, "Overcoming Depression on the Internet (ODIN)(2): a randomized trial of a self-help depression skills program with reminders," *Journal of medical Internet research*, vol. 7, no. 2, p. e412, 2005.
- [49] S. M. Kelders, R. N. Kok, H. C. Ossebaard, and J. E. Van Gemert-Pijnen, "Persuasive system design does matter: a systematic review of adherence to web-based interventions," *Journal of medical Internet research*, vol. 14, no. 6, p. e2104, 2012.
- [50] R. Neff and J. Fry, "Periodic prompts and reminders in health promotion and health behavior interventions: systematic review," *Journal of medical Internet research*, vol. 11, no. 2, p. e1138, 2009.
- [51] L. Maenhout, C. Peuters, G. Cardon, G. Crombez, A. DeSmet, and S. Compennolle, "Nonusage attrition of adolescents in an mHealth promotion intervention and the role of socioeconomic status: secondary analysis of a 2-arm cluster-controlled trial," *JMIR mHealth and uHealth*, vol. 10, no. 5, p. e36404, 2022.
- [52] D. King, F. Greaves, C. Exeter, and A. Darzi, "'Gamification': Influencing health behaviours with games," vol. 106, ed: SAGE Publications Sage UK: London, England, 2013, pp. 76-78.
- [53] C. Lister, J. H. West, B. Cannon, T. Sax, and D. Brodegard, "Just a fad? Gamification in health and fitness apps," *JMIR serious games*, vol. 2, no. 2, p. e3413, 2014.
- [54] M. Schmidt-Kraepelin, P. A. Toussaint, S. Thiebes, J. Hamari, and A. Sunyaev, "Archetypes of gamification: analysis of mHealth apps," *JMIR mHealth and uHealth*, vol. 8, no. 10, p. e19280, 2020.
- [55] C. Dichev and D. Dicheva, "Gamifying education: what is known, what is believed and what remains uncertain: a critical review," *International journal of educational technology in higher education*, vol. 14, no. 1, pp. 1-36, 2017.

- [56] H. Kalantarian, P. Washington, J. Schwartz, J. Daniels, N. Haber, and D. Wall, "A gamified mobile system for crowdsourcing video for autism research," in *2018 IEEE international conference on healthcare informatics (ICHI)*, 2018: IEEE, pp. 350-352.
- [57] H. Kalantarian, K. Jedoui, P. Washington, and D. P. Wall, "A mobile game for automatic emotion-labeling of images," *IEEE transactions on games*, vol. 12, no. 2, pp. 213-218, 2018.
- [58] R. Farzan, J. M. DiMicco, D. R. Millen, C. Dugan, W. Geyer, and E. A. Brownholtz, "Results from deploying a participation incentive mechanism within the enterprise," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2008, pp. 563-572.
- [59] C. Cheong, F. Cheong, and J. Filippou, "Quick quiz: A gamified approach for enhancing learning," 2013.
- [60] W. LI, T. GROSSMAN, and G. FITZMAURICE, "Gamified tutorial system for first time AutoCAD users," in *UIST*, 2012, vol. 12, pp. 7-10.
- [61] T. D.-L. Guin, R. Baker, J. Mechling, and E. Ruyle, "Myths and realities of respondent engagement in online surveys," *International Journal of Market Research*, vol. 54, no. 5, pp. 613-633, 2012.
- [62] A. L. Smith and L. Baker, "Getting a clue: creating student detectives and dragon slayers in your library," *Reference Services Review*, 2011.
- [63] L. Hoffman, *Longitudinal analysis: Modeling within-person fluctuation and change*. Routledge, 2015.
- [64] M. Csikszentmihalyi, *Handbook of research methods for studying daily life*. Guilford Press, 2011.
- [65] S. Shiffman, A. Stone, and M. Hufford, "Ecological momentary assessment. Annu. Rev. Clin. Psychol," *Ecological Momentary Assessment. Annu Rev Clin Psychol*, vol. 4, no. 1, 2008.
- [66] M. Balsamo, F. Cataldi, L. Carlucci, C. Padulo, and B. Fairfield, "Assessment of late-life depression via self-report measures: a review," *Clinical interventions in aging*, vol. 13, p. 2021, 2018.
- [67] W. H. Organization. Mental Health of Older Adults [Online] Available: <https://www.who.int/en/news-room/fact-sheets/detail/mental-health-of-older-adults>
- [68] J. Lin, D. Ebert, D. Lehr, M. Berking, and H. Baumeister, "Internet based cognitive behavioral interventions: state of the art and implementation possibilities in rehabilitation," *Die Rehabilitation*, vol. 52, no. 3, pp. 155-163, 2013.
- [69] M. Cunningham, V. Wuthrich, R. Rapee, H. Lyneham, C. Schniering, and J. Hudson, "The Cool Teens CD-ROM for anxiety disorders in adolescents," *European child & adolescent psychiatry*, vol. 18, no. 2, pp. 125-129, 2009.
- [70] A. L. Calear and H. Christensen, "Review of internet-based prevention and treatment programs for anxiety and depression in children and adolescents," *Medical Journal of Australia*, vol. 192, pp. S12-S14, 2010.
- [71] A. Gulliver, K. M. Griffiths, and H. Christensen, "Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review," *BMC psychiatry*, vol. 10, no. 1, pp. 1-9, 2010.
- [72] R. C. Kessler, P. Berglund, O. Demler, R. Jin, K. R. Merikangas, and E. E. Walters, "Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication," *Archives of general psychiatry*, vol. 62, no. 6, pp. 593-602, 2005.
- [73] A. J. Bravo, M. C. Villarosa-Hurlocker, and M. R. Pearson, "College student mental health: An evaluation of the DSM-5 self-rated Level 1 cross-cutting symptom measure," *Psychological Assessment*, vol. 30, no. 10, p. 1382, 2018.
- [74] A. Öhman, D. Lundqvist, and F. Esteves, "The face in the crowd revisited: a threat advantage with schematic stimuli," *Journal of personality and social psychology*, vol. 80, no. 3, p. 381, 2001.
- [75] K. Mogg and B. P. Bradley, "A cognitive-motivational analysis of anxiety," *Behaviour research and therapy*, vol. 36, no. 9, pp. 809-848, 1998.

- [76] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735-1780, 1997.
- [77] B. Sun, Q. Wei, L. Li, Q. Xu, J. He, and L. Yu, "LSTM for dynamic emotion and group emotion recognition in the wild," in *Proceedings of the 18th ACM international conference on multimodal interaction*, 2016, pp. 451-457.
- [78] Y. Zhang, X. Chen, D. Zhou, and M. I. Jordan, "Spectral methods meet EM: A provably optimal algorithm for crowdsourcing," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 3537-3580, 2016.
- [79] D. Zhou, S. Basu, Y. Mao, and J. Platt, "Learning from the wisdom of crowds by minimax entropy," *Advances in neural information processing systems*, vol. 25, 2012.
- [80] M. T. Ribeiro, S. Singh, and C. Guestrin, "Nothing else matters: Model-agnostic explanations by identifying prediction invariance," *arXiv preprint arXiv:1611.05817*, 2016.
- [81] J. Chang, S. Gerrish, C. Wang, J. Boyd-Graber, and D. Blei, "Reading tea leaves: How humans interpret topic models," *Advances in neural information processing systems*, vol. 22, 2009.
- [82] B. H. Hansen, E. Kolle, S. M. Dyrstad, I. Holme, and S. A. Anderssen, "Accelerometer-determined physical activity in adults and older people," *Medicine and science in sports and exercise*, vol. 44, no. 2, pp. 266-272, 2012.
- [83] S. Saeb *et al.*, "Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study," *Journal of medical Internet research*, vol. 17, no. 7, p. e4273, 2015.
- [84] L. K. George, D. G. Blazer, D. C. Hughes, and N. Fowler, "Social support and the outcome of major depression," *The British Journal of Psychiatry*, vol. 154, no. 4, pp. 478-485, 1989.
- [85] A. Banerjee, P. Washington, C. Mutlu, A. Kline, and D. P. Wall, "Training and profiling a pediatric emotion recognition classifier on mobile devices," *arXiv preprint arXiv:2108.11754*, 2021.
- [86] C. Hou, H. Kalantarian, P. Washington, K. Dunlap, and D. P. Wall, "Leveraging video data from a digital smartphone autism therapy to train an emotion detection classifier," *medRxiv*, p. 2021.07. 28.21260646, 2021.
- [87] H. Kalantarian *et al.*, "Labeling images with facial emotion and the potential for pediatric healthcare," *Artificial intelligence in medicine*, vol. 98, pp. 77-86, 2019.
- [88] P. Washington *et al.*, "Training an emotion detection classifier using frames from a mobile therapeutic game for children with developmental disorders," *arXiv preprint arXiv:2012.08678*, 2020.
- [89] P. Washington *et al.*, "Training affective computer vision models by crowdsourcing soft-target labels," *Cognitive computation*, vol. 13, pp. 1363-1373, 2021.
- [90] D. L. Christensen *et al.*, "Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2012," *MMWR Surveillance Summaries*, vol. 65, no. 13, p. 1, 2018.
- [91] K. Ardhanareeswaran and F. Volkmar, "Introduction. Focus: autism spectrum disorders," *The Yale Journal of Biology and Medicine*, vol. 88, no. 1, pp. 3-4, 2015.
- [92] E. Gordon-Lipkin, J. Foster, and G. Peacock, "Whittling down the wait time: exploring models to minimize the delay from initial concern to diagnosis and treatment of autism spectrum disorder," *Pediatric Clinics*, vol. 63, no. 5, pp. 851-859, 2016.
- [93] J. Manfredonia *et al.*, "Automatic recognition of posed facial expression of emotion in individuals with autism spectrum disorder," *Journal of autism and developmental disorders*, vol. 49, pp. 279-293, 2019.
- [94] A. Nag *et al.*, "Toward continuous social phenotyping: analyzing gaze patterns in an emotion recognition task for children with autism through wearable smart glasses," *Journal of medical Internet research*, vol. 22, no. 4, p. e13810, 2020.

- [95] A. Lakkapragada *et al.*, "The classification of abnormal hand movement to aid in autism detection: Machine learning study," *JMIR Biomedical Engineering*, vol. 7, no. 1, p. e33771, 2022.
- [96] P. Washington *et al.*, "Activity recognition with moving cameras and few training examples: applications for detection of autism-related headbanging," in *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1-7.
- [97] P. Washington *et al.*, "A wearable social interaction aid for children with autism," in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2016, pp. 2348-2354.
- [98] N. Haber *et al.*, "A wearable social interaction aid for children with autism," *arXiv preprint arXiv:2004.14281*, 2020.
- [99] P. Washington and D. P. Wall, "A Review of and Roadmap for Data Science and Machine Learning for the Neuropsychiatric Phenotype of Autism," *arXiv preprint arXiv:2303.03577*, 2023.
- [100] C. Voss, N. Haber, and D. P. Wall, "The potential for machine learning–based wearables to improve socialization in teenagers and adults with autism spectrum disorder—reply," *JAMA pediatrics*, vol. 173, no. 11, pp. 1106-1106, 2019.
- [101] J. Daniels *et al.*, "Exploratory study examining the at-home feasibility of a wearable tool for social-affective learning in children with autism," *NPJ digital medicine*, vol. 1, no. 1, p. 32, 2018.
- [102] C. Voss *et al.*, "Effect of wearable digital intervention for improving socialization in children with autism spectrum disorder: a randomized clinical trial," *JAMA pediatrics*, vol. 173, no. 5, pp. 446-454, 2019.
- [103] H. Kalantarian *et al.*, "The performance of emotion classifiers for children with parent-reported autism: quantitative feasibility study," *JMIR mental health*, vol. 7, no. 4, p. e13174, 2020.
- [104] H. Kalantarian, P. Washington, J. Schwartz, J. Daniels, N. Haber, and D. P. Wall, "Guess What? Towards Understanding Autism from Structured Video Using Facial Affect," *Journal of healthcare informatics research*, vol. 3, pp. 43-66, 2019.
- [105] Y. Qian *et al.*, "Computer Vision Estimation of Emotion Reaction Intensity in the Wild," *arXiv preprint arXiv:2303.10741*, 2023.
- [106] A. Kline *et al.*, "Superpower glass," *GetMobile: Mobile Computing and Communications*, vol. 23, no. 2, pp. 35-38, 2019.
- [107] C. Voss *et al.*, "Superpower glass: delivering unobtrusive real-time social cues in wearable systems," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, 2016, pp. 1218-1226.
- [108] P. Washington *et al.*, "Data-driven diagnostics and the potential of mobile artificial intelligence for digital therapeutic phenotyping in computational psychiatry," *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, vol. 5, no. 8, pp. 759-769, 2020.
- [109] N. Haber, C. Voss, and D. Wall, "Making emotions transparent: Google Glass helps autistic kids understand facial expressions through augmented-reality therapy," *IEEE Spectrum*, vol. 57, no. 4, pp. 46-52, 2020.
- [110] P. Washington *et al.*, "SuperpowerGlass: a wearable aid for the at-home therapy of children with autism," *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, vol. 1, no. 3, pp. 1-22, 2017.
- [111] J. Daniels *et al.*, "Feasibility testing of a wearable behavioral aid for social learning in children with autism," *Applied clinical informatics*, vol. 9, no. 01, pp. 129-140, 2018.
- [112] C. Voss *et al.*, "Designing a holistic at-home learning aid for autism," *arXiv preprint arXiv:2002.04263*, 2020.