# The Aggression Pattern: Evaluating Signals from Wearable Devices for Agitation Detection

Govind Molakalapalli[1], Atharv Gupta[1], Aymen Omara[1], Bill Kitchel[1], Antonia Winkler[1], Matthew Clark[2], and Afsaneh Doryab[1,2,*]
**Corresponding author: ad4ks@virginia.edu*

[1] *Department of Systems and Information Engineering*
*University of Virginia*
*Charlottesville, VA, USA*

[2] *Department of Computer Science*
*University of Virginia*
*Charlottesville, VA, USA*

***Abstract*—Reactive aggression occurs when individuals take impulsive, damaging actions in response to perceived threats or provocations. Previous work has identified physiological indicators of aggressive behavior, though none have thoroughly analyzed multimodal physiological indicators in a general population. As such, we recruited 22 college students to complete an intentionally agitating task while recording their electrodermal activity, blood volume pulse, heart rate, and skin temperature using Empatica Embrace Plus. We then analyzed the participants' physiological changes in relation to their interferences and imposed provocations during the game. Our analysis uses both statistical methods and machine learning to identify relationships between participants' gameplay and their physiological responses. All four physiological modalities contribute to predicting reactive agitation. However, the predictions regarding provocation are primarily based on skin conductance and temperature. Across all our analyses, we observe a clear lack of generalization in physiological changes across participants, indicating that agitation prediction systems must account for individualized physiological nuances.**

***Keywords*—Physiological Signals, Manifestations of Agitation, Machine Learning**

## I. Introduction

Reactionary agitation, an impulsive response to perceived provocation [1], is a key contributor to interpersonal violence, including those that result in gun violence. Identifying early manifestations of these reactive states presents an opportunity to develop and deploy interventions to reduce harmful outcomes. Therefore, analysis of measurable physiological changes associated with these states may enable earlier recognition of risk and support the development of preventative intervention strategies. Advancements in wearable devices enable accurate tracking of signals, creating opportunities to innovate towards early-detection systems for reactive agitation. The ability to track specific physiological signals during acute frustration may enable the detection of the escalation phase preceding a reactive outburst. If these physiological patterns can be reliably modeled, commercial wearables could evolve from passive health trackers into active de-escalation tools, alerting users to heightened agitation states before cognitive control is bypassed.

While some existing research has demonstrated that changes in physiological signals, including heart rate and skin conductance, are associated with aggressive incidents [1]–[5], the evidence is limited and inconsistent, particularly for reactive agitation. This study investigates whether reactive aggression produces measurable changes in wearable-derived physiological signals, including electrodermal activity, blood volume pulse, heart rate, and skin temperature, in a general population. We developed a game that systematically varied levels of provocation to induce agitation and capture corresponding behavioral and physiological responses. We find that although no single physiological modality shows significant relationships with reactive agitation or provocation levels, they can be predicted using a multi-modal feature space. We also observe that while physiological relationships to agitation and provocation levels are present, they are highly individualized. These observations suggest that wearable sensing can capture physiological changes associated with agitation, but effective detection may require multimodal, person-specific modeling.

## II. Related Work

Advances in multi-modal sensing technologies have demonstrated the ability to detect aggressive behavior [1]–[5]. However, studies investigating the relationship between reactive agitation and physiological signals report conflicting results. While Boccadoro et al. [1] found no significant relationship between skin conductance responses and reactive agitation measures, Choi et al. [5] and Thomson et al. [6] demonstrated increases in electrodermal activity and sympathetic nervous system arousal in relation to evoked aggression. More pronounced changes in physiological signals have been observed in patient populations. For example, one study found that both heart rate and electrodermal activity increase when forensic psychiatric inpatients are in a higher state of aggression [2]. A study on people with dementia found that skin conductance and temperature features were significant in predicting aggressive events although the importance of these features varied across participants [7]. Another study on youth with autism using the same physiological signals found that those signals contributed explanatory information to model predictions but the authors did not examine the characteristics of the features that indicated aggression [3]. These works have not investigated how reactive aggression relates to multimodal physiological signals in a general population, which this paper explores.

## III. Game-Based Study for Inducing Agitation

We conducted a human-subjects study to evaluate the degree to which subjective agitation corresponds to changes in physiological responses. In this study, participants played against a computer opponent to complete competitive tasks while wearing an Empatica Embrace Plus [8] to track physiology. Players and opponents could interfere with each other by increasing the task's difficulty. Following Institutional Review Board approval, 22 undergraduates and graduate students (male = 9, female = 13, 18-30 years old) at our university were recruited.

### A. Game Design

We designed a simple game inspired by the Taylor Aggression Paradigm [1], [9], in which the player must determine whether targets are present in a 5x5 grid displayed to them. Participants were informed they would be playing against an opponent to identify the number of squares that were a specific shade of dark green as quickly and accurately as possible. To induce frustration, both the participant and the opponent could make the game harder for each other by applying slowed buttons, distracting grids, or a combination of the two (see Figure 1 for grid previews). The provocation level, i.e., the difficulty level applied by the virtual opponent, was intended to induce agitation in participants towards the opponent. Interference level, on the other hand, was the difficulty level participants selected to make the game harder/easier for the virtual opponent. We used interference as a metric to record participants' agitation level. Both provocation and interference have levels 1 through 4, where higher values correspond to more difficult distractions, as shown in Figure 1. Time constraints were set to create a sense of urgency. Participants had a maximum of 5 seconds to select their interference level, 30 seconds to identify targets and respond, and 15 seconds to review the leaderboard. These three screens together made one trial, and 15 trials made one round. All participants completed three rounds. To induce varied interference responses from the player, the opponent follows different provocation patterns during each round.

### B. Study Format

The study was designed as follows: participants completed a pre-survey, then played a game round and completed

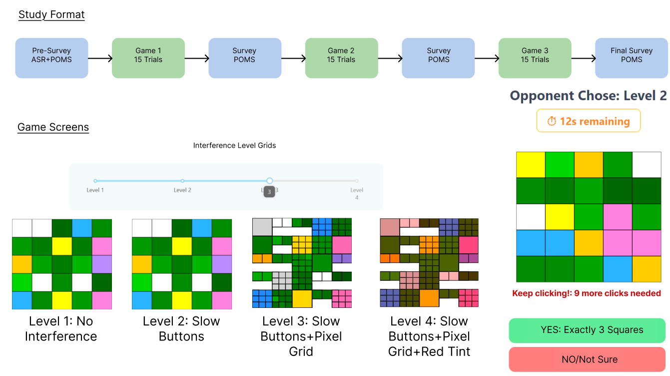



Fig. 1: Study process and game screens. Participants played three rounds of the game against virtual opponents and completed questionnaires before and after each round. The game asked participants to indicate the number of dark green squares in a 5x5 grid. Participants could interfere, changing the grid to make the number of green squares harder to count. Simultaneously, the virtual opponent could interfere with the participant.

the Profile of Mood States (POMS) [10]. During the pre-survey segment, participants completed the 30-item Anger Self-Report (ASR) [11] and the POMS. ASR responses were measured on a 6-point ordinal scale and numerically encoded to compute an average agitation tendency score for each participant. Following this, participants were briefed on game rules. The study format can be reviewed in Figure 1.

Participants played three game rounds (15 trials each), each followed by the POMS survey and a short break to allow physiological signals to return to baseline. In round 1, the opponent increased provocation over time (starting at 1 or 2 and increasing); in round 2, it was aggressive throughout (3-4 only); and in round 3, it was normal (very low provocation). The spread of the provocation levels and the associated interference response are shown in Table I. At the end of all rounds, participants completed a final POMS survey to identify lasting shifts in agitation.

## IV. Analysis and Results

The analysis is organized around two main research questions (RQs):

- **RQ1:** What physiological signals are indicative of reactive agitation?
- **RQ2:** How do physiological signals indicative of reactive agitation vary between individuals, along with their qualitative self-reported agitation?

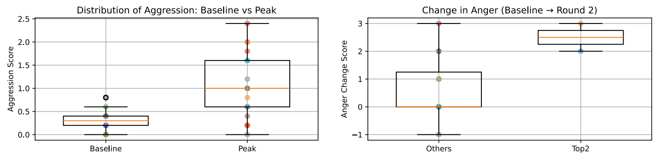
In the following sections, we describe the processing and modeling steps for analyzing the collected data. Our method

TABLE I: Provocation & interference summary stats (both use scale from 1-4)

| Game # | Provocation (avg ± std) | Interference (avg ± std) |
|---|---|---|
| 1 | 2.23 ± 0.85 | 2.25 ± 1.27 |
| 2 | 3.56 ± 0.50 | 2.84 ± 1.22 |
| 3 | 1.09 ± 0.41 | 2.56 ± 1.24 |

TABLE II: Distribution of times each provocation and interference level was selected.

| Level | Provocation Count | Decision Count |
|---|---|---|
| 1 | 374 | 316 |
| 2 | 160 | 126 |
| 3 | 207 | 163 |
| 4 | 204 | 340 |




a) Baseline vs Peak Agitation b) Peak Anger Change (Top 2 vs Others)

Fig. 2: (a) Baseline vs Peak Agitation POMS (b) Peak anger change by group ("generally aggressive" vs others)

consists of machine learning, mixed-effects models, time series, and statistical analysis to explore common and individual patterns that may answer our research questions.

### *A. Data Processing*

Data from the game, including associated timestamps for each window, were extracted and merged with high-resolution Empatica data for comprehensive analysis. Empatica data was aggregated to the second, logging second-by-second electrodermal activity (EDA), blood volume pulse (BVP), heart rate, and skin temperature. This was then joined with gameplay logs, which tracked the response time, accuracy, interference level, opponent provocation level, and time spent in each window (interference, game, leaderboard). While processing the game data, one participant's data was lost. The remaining 21 participants were used for the joint analysis of the game and physiological signals. To prepare data for machine learning, trial-level features were extracted from the aligned dataset by aggregating physiological signals within each trial window. For each signal, summary statistics such as mean, minimum, maximum, and standard deviation were computed, resulting in a structured trial-level dataset used for modeling. The aligned and aggregated data were then used for time-series analysis, statistical evaluation, and machine learning.

### *B. Measuring Game-Induced Agitation*

To verify that the designed game successfully induced agitation-related responses, we conducted a statistical analysis on self-reported mood data collected through the POMS survey. Numerical agitation scores were derived by averaging the numerical scores of the 5 agitation-related mood states (angry, annoyed, resentful, on-edge, and miserable). The numerical value from 0 to 5 was derived from the ordinal nature of the POMS survey responses. We compared baseline pre-survey responses to peak agitation values (which varied across participants) observed in Figure 2. Given the paired and ordinal nature of the data, a Wilcoxon signed-rank test was performed. Results indicated a statistically significant increase in agitation scores, with median values rising from 0.300 at baseline to 1.000 at peak ($W = 13.50, p = 0.0002$), and 20 out of 22 participants exhibited an increase. These findings confirm that the game effectively elicited heightened agitation-related states, supporting its use for subsequent physiological and behavioral analysis.

To understand how baseline agitation tendencies relate to induced emotional responses, participants were stratified using their Anger Self Report (ASR) responses collected prior to the task. Two participants had a positive average score of 0.286 (higher scores indicate more agitation) and were classified in the "more aggressive" subgroup, while the other group had an average score of -1.5. To compare peak changes in POMS-derived anger scores (defined as the change from baseline to the maximum agitation value observed across rounds), a Mann-Whitney U test was conducted between this subgroup and the rest of the cohort. As shown in Figure 2, participants with higher baseline agitation reported consistently higher increases in agitation scores compared to those of other participants. However, the difference between the two groups did not reach statistical significance ($U = 36.5, p = 0.053$), which is possibly due to the small subgroup size (n = 2). Despite this, the observed trend suggests that individuals with higher baseline agitation exhibited stronger emotional responses to provocation imposed by the virtual opponent in the game. We found that the interference response in the game data is skewed towards the two extremes of the level categories, with levels 1 and 4 being chosen 316 and 340 times, respectively, as shown in Table II. It is worth noting, however, that if the five-second window to apply an interference level expired, the player had level 1 automatically selected as their response. This, in turn, made the game harder for them to win.

### *C. Predicting Interference and Provocation from Physiological Responses*

*Machine Learning Modeling:* First, we explored whether machine learning could predict participants' interference levels using physiological data. This analysis used a Random Forest classification algorithm and utilized physiological features to predict interference and provocation levels. To minimize overfitting and perform cross-validation, we used Leave-One-Participant-Out (LOPO) and per-participant Leave-One-Trial-Out (LOTO) cross-validation. A per-participant overview of model accuracy for both of these cross-validation strategies is shown in Figure 3. Utilizing LOTO to predict interference level, the model achieved a mean accuracy of approximately 0.63 ($\sigma = 0.28; 95\%$Confidence Interval (CI) $: 0.50 - 0.76$), slightly above the majority-class baseline of 0.61 (paired t-test: $p = 0.53$). This suggests the model does not consistently outperform the baseline. In particular, there was substantial variation in predicting the selected interference level across participants, suggesting that the physiology of some participants may be more predictive of their agitation than others. Meanwhile, when using a LOPO cross-validation for interference level prediction, the accuracy is 0.315 ($\sigma = 0.28; 95\%$ CI $: 0.19 - 0.44$), which is negligibly different than the majority-class baseline of 0.323 ($p = 0.93$), suggesting this model did not perform any better than always guessing the most common label. Therefore, these results indicate individualized models may be more effective at predicting agitation, and that model performance may differ greatly between individuals.

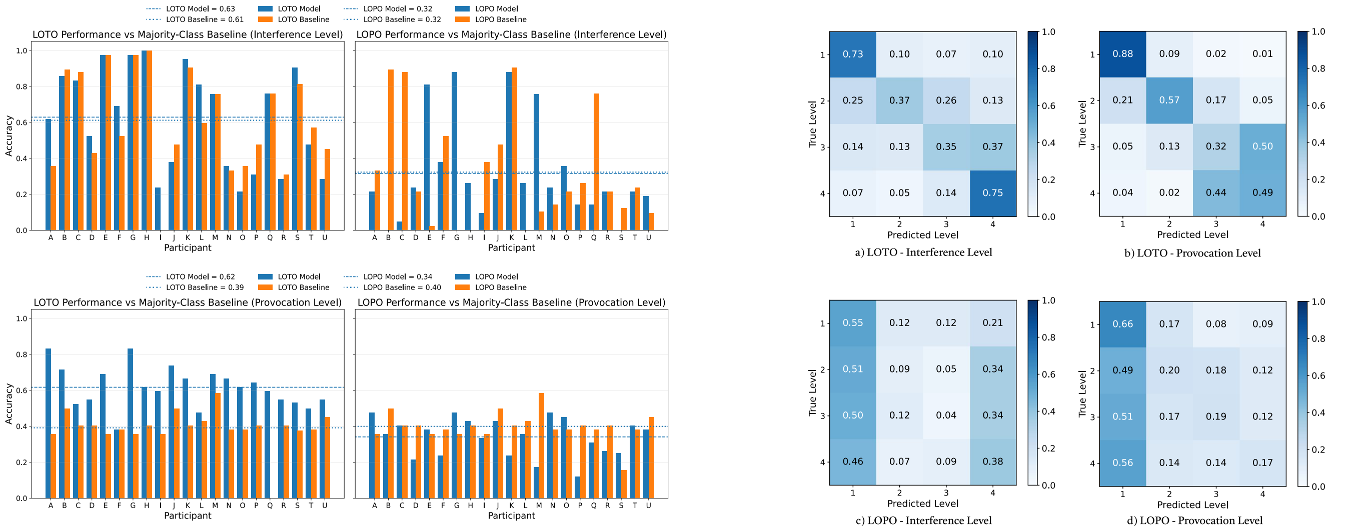



Fig. 3: Model performance and classification results for interference and provocation level prediction with LOTO and (LOPO) validation. Panels on the left show per-participant accuracy for LOTO and LOPO, with bars representing model and baseline performance and horizontal lines indicating mean accuracies. Panels on the right present normalized confusion matrices, where rows correspond to true levels and columns to predicted levels, for both interference and provocation across LOTO (top row) and LOPO (bottom row).

When predicting opponent provocation level, LOTO achieved a mean accuracy of approximately 0.62 ($\sigma = 0.11; 95\%\ \text{CI} : 0.57 - 0.67$), significantly above the majority-class baseline of 0.39 ($p < 0.001$). Within a LOPO cross-validation, the model achieved an accuracy of approximately 0.34 ($\sigma = 0.11; 95\%\ \text{CI} : 0.29 - 0.39$), compared to a majority-class baseline of 0.40 ($p = 0.067$). The performance of these models indicates that provocation level is also more accurately predicted using personalized models, suggesting the presence of individualized physiological patterns.

The confusion matrices for both validations are shown in Figure 3. LOTO cross-validation shows a clear diagonal of correctly predicted labels. Interestingly, we also observe that most model errors were off by only a single interference level (e.g., level 3 was frequently predicted as level 4), suggesting the potential presence of physiological differences between levels, even if the exact granular level cannot be predicted. However, the LOPO models show a different pattern. For both interference and provocation level, samples were often predicted to be level 1, regardless of what participants actually selected. In contrast, the second-most-commonly predicted label across all ground-truth interference levels was 4. This may be indicative of common strategies for the game, where some participants routinely inflict minimal interference on their opponents, while others inflict the most interference. This is supported by Table I, which shows that levels 1 and 4 were the most commonly selected interference levels. The model's performance suggests it could identify these common strategies but could not use physiology to predict which strategy participants would take. Taken together, these confusion matrices provide clear insight into the superior performance of the LOTO models, further emphasizing the benefit of personalized predicted modeling.

*Important Physiological Features in Predicting Interference and Provocation Levels:* To identify the physiological features most indicative of participants' selected interference level, we calculated the permutation-based importance for each feature across folds of the LOTO cross-validation. These values were used to identify the most frequently important features. This set is defined by counting the number of folds where each feature was more important than average. We then calculate the median number of occurrences and save the features that occur more frequently than average. The results of this analysis are shown in Figure 4.

The frequently important physiological features varied between models that predicted interference and provocation levels. When predicting participants' selected interference levels, the model frequently used features from all four modalities. Of the 5 features from each modality, 3 skin temperature, 3 EDA, 2 BVP, and 2 heart rate features were included in the frequently important feature set. However, when predicting provocation level, skin temperature (4/5 features) and EDA (4/5 features) accounted for almost all the most frequently important features. These differences suggest that predicting interference level relies on multimodal modeling of participants' physiology, whereas predicting provocation level relies on skin temperature and EDA.

*Mixed-effects Model:* To evaluate how physiological signals, provocation level, and interference level were interrelated, a linear mixed-effects model was applied to test for relationships within the game logs and physiological data. In particular, these models were used to identify significant multivariate relationships between participants' gameplay and the mean and standard deviation of each physiological signal while accounting for individualized gameplay styles. This

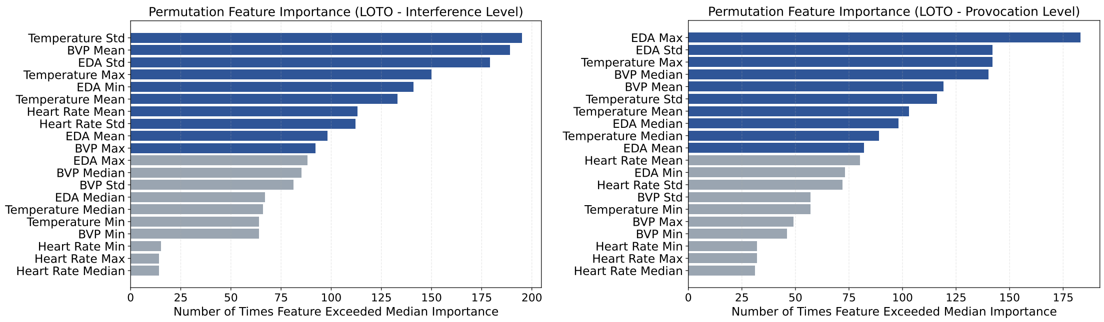



Fig. 4: Permutation feature importance results across participants for interference (left) and provocation (right), highlighting the most consistently influential physiological features.

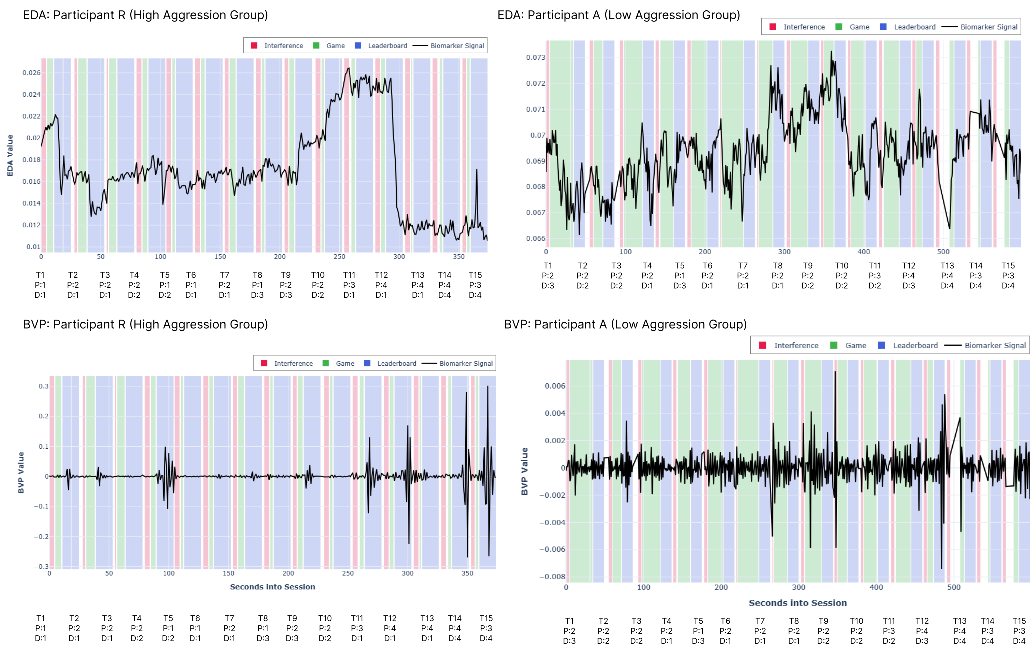



Fig. 5: High baseline participants by ASR, participant R and low Baseline by ASR, participant A

analysis used 845 samples from 21 participants.

There were no significant relationships between EDA, BVP, skin temperature, or heart rate and the selected interference levels ($p > 0.050$) or provocation levels ($p > 0.090$). Notably, the standard deviation of skin temperature showed a barely insignificant relationship with the interference level ($\beta = -0.4927, p = 0.053$). However, several significant relationships were identified within the gameplay variables. In particular, provocation and interference were interrelated ($\beta = 0.133, p = 0.028$), and both had significant relationships with the trial number ($\beta_{\text{Interference}} = 0.031, p_{\text{Interference}} < 0.001, \beta_{\text{Provocation}} = 0.042, p_{\text{ Provocation}} < 0.001$) and game number ($\beta_{\text{Interference}} = 0.0.230, p_{\text{Interference}} < 0.001, \beta_{\text{Provocation}} = -0.635, p_{\text{ Provocation}} < 0.001$). This suggests that, while participants generally copied their opponent's aggressiveness, their play styles became more aggressive over time, even when the virtual opponent played less aggressively.

*Time Series Comparison of Participant Physiology in Relation to Self-Reported Mood and Aggressive Tendencies:* Finally, we used time-series analysis to compare changes in physiological signals among participants who self-reported experiencing different moods and aggressive tendencies during our study. As mentioned before, two participants had the highest baseline agitation scores in the group according to the pre-survey analysis. These two individuals were placed in the "generally aggressive" subgroup. A comparison of physiological signals for two participants while playing the game is shown in Figure 5, where participant R is in the generally aggressive subgroup and participant A is in the low aggression subgroup. Interestingly, comparing the two participants shows that participant A closely follows the opponent's provocation level from the previous trial. Meanwhile, participant R shows higher interference only towards the end of the first round. As shown in Figure 5, participants exhibit similar spikes when the opponent applies a higher level of provocation, likely due to increased agitation from the more challenging game.

## V. Discussion

Our analysis focused on investigating physiological signals indicative of reactive agitation (RQ1) and differences in response between individuals (RQ2). The findings reveal the complex relationship between physiological spikes and reactive agitation. We discovered that agitation as an emotion is difficult to categorize, but the underlying trends in physiological signals may provide insight into how an individual might react.

The results of our feature importance analysis indicate that multiple physiological signals contribute to predicting reactive agitation. The most frequently important feature for predicting the provocation level was maximum EDA. This roughly aligns with findings from de Looff's study, which found that both EDA and heart rate spiked in the 30 minutes preceding an aggressive event, during which patients lashed out verbally and/or physically [2]. However, the models for predicting interference levels considered features from all four modalities. This suggests that predicting agitation levels in the general population requires a more comprehensive understanding of participants' physiology compared to those in mental health inpatient settings.

Our results also indicate a significant difference in accuracy between population and personalized models. This trend aligns with previous literature [3], which predicted aggressive behavior for individuals with Autism Spectrum Disorder (ASD). Our work suggests that these individual differences in physiology and agitation are not specific to inpatients and must also be considered when predicting aggression in the general population.

### A. Limitations and Future Directions

The primary limitations of this study are the subjective nature of mapping physiological signals to specific emotional states and the constrained sample size. Whereas statistical significance was achieved in several metrics, isolating pure agitation from general spikes or frustration remains challenging. To calculate a proxy for the agitation score, we assessed a range of related mood states, with agitated responses linked to feelings of frustration, annoyance, and anxiety, as captured in survey data. Furthermore, with a final recruitment of 22 participants, one of whom was excluded due to incomplete game data, the sample size is relatively small. Consequently, it is challenging to confidently state that the findings can be generalized to a larger population. Future research should focus on larger participant cohorts to validate person-dependent modeling approaches and to better differentiate between overlapping states of agitation.

## VI. Conclusion

This study investigated the feasibility of using commercial wearable devices to detect physiological manifestations of reactive agitation during a controlled, competitive game. Analyzing physiological data revealed the immense complexity of agitation prediction. While statistical model evaluation revealed that opponent provocation and study duration are strong indicators of aggressive behavior, physiological signals and game data from participants did not reach broad statistical significance. Furthermore, machine learning evaluations confirmed initial hypotheses that cardiovascular features, such as BVP and heart rate, are the most critical indicators of aggressive patterns, but predictive performance varied across participants and did not generalize. Ultimately, these findings highlight the individualized nature of physiological reaction. Future efforts in wearable-based early warning systems must pivot toward person-dependent, personalized baseline models.


## Acknowledgment

The research team would like to thank Bobby Doyle for his ongoing knowledge and support of this work. Furthermore, we would like to thank Dr. Peter de Looff for meeting with us to discuss insights into his work.


## References


[1] S. Boccadoro, L. Wagels, A. T. Henn, P. Hüpen, L. Graben, A. Raine, and I. Neuner, "Proactive vs. reactive aggression within two modified versions of the taylor aggression paradigm," *Frontiers in Behavioral Neuroscience*, vol. 15, p. 749041, 2021.

[2] P. de Looff, M. L. Noordzij, M. Moerbeek, H. Nijman, R. Didden, and P. Embregts, "Changes in heart rate and skin conductance in the 30 min preceding aggressive behavior," *Psychophysiology*, vol. 56, p. e13420, 2019.

[3] M. S. Goodwin, C. A. Mazefsky, S. Ioannidis, D. Erdogmus, and M. Siegel, "Predicting aggression to others in youth with autism using a wearable biosensor," *Autism Research*, vol. 11, no. 11, pp. 1534–1546, 2019.

[4] P. C. De Looff, L. J. M. Cornet, C. H. De Kogel, B. Fernández-Castilla, P. J. C. M. Embregts, R. Didden, and H. L. I. Nijman, "Heart rate and skin conductance associations with physical aggression, psychopathy, antisocial personality disorder and conduct disorder: An updated meta-analysis," *Neuroscience & Biobehavioral Reviews*, vol. 132, pp. 553–582, 2022.

[5] M.-H. Choi, S.-J. Lee, J.-W. Yang, J.-H. Kim, J.-S. Choi, H.-S. Kim, J.-Y. Park, J.-H. Jun, G.-R. Tack, H.-J. Kim, and S.-C. Chung, "An analysis of the correlation between young males' personal aggression and their skin conductance levels during exposure to aggression images," *Psychiatry Research*, vol. 186, no. 2-3, pp. 441–442, 2011.

[6] N. D. Thomson, S. Kevorkian, J. Blair, J. Farrell, S. J. West, and J. M. Bjork, "Psychophysiological underpinnings of proactive and reactive aggression in young men and women," *Physiology Behavior*, vol. 242, 2021.

[7] A. Badawi, S. Elmoghazy, S. Choudhury, S. Elgazzar, K. Elgazzar, and A. Burhan, "ultimodal detection of agitation in people with dementia in clinical settings: Observational pilot study," *JMIR Aging*, vol. 8, 2025.

[8] Empatica Inc., "Embrace Plus." [Online]. Available: https://www.empatica.com/embraceplus/. Accessed: 12-Apr-2026.

[9] R. J. McCarthy and M. Elson, "A conceptual review of lab-based aggression paradigms," *Collabra: Psychology*, vol. 4, no. 1, pp. 1–12, 2018.

[10] S. L. Curran, M. A. Andrykowski, and J. L. Studts, "Short form of the profile of mood states (poms-sf): psychometric information.," *Psychological assessment*, vol. 7, no. 1, p. 80, 1995.

[11] N. S. Reynolds, F. H. Walkey, and D. E. Green, "The anger self report: A psychometrically sound (30 item) version," *New Zealand Journal of Psychology*, vol. 23, pp. 64–64, 1994.