

Stress Mindset Matters: Rethinking Mental Stress Detection with Multimodal Wearable Sensors

Lakmal Meegahapola
Nokia Bell Labs
Cambridge, United Kingdom
lakmal.meegahapola@nokia.com

Marios Constantinides
CYENS Centre of Excellence
Nicosia, Cyprus
marios.constantinides@cyens.org.cy

Zoran Radivojevic
Nokia Bell Labs
Cambridge, United Kingdom
zoran.radivojevic@nokia.com

Hongwei Li
Nokia Bell Labs
Cambridge, United Kingdom
hongwei.3.li@nokia.com

Michael S. Eggleston
Nokia Bell Labs
Murray Hill, New Jersey, USA
michael.eggleston@nokia.com

Daniele Quercia
Nokia Bell Labs
Cambridge, United Kingdom
Politecnico di Torino
Turin, Italy
quercia@cantab.net

Abstract

The mindset people have about stress is important to be studied because this core belief, that stress is either enhancing or debilitating, fundamentally alters a person's physiological and psychological responses to stressors. However, this crucial construct is rarely considered in prior research on momentary stress detection with wearables, leaving two fundamental questions unanswered: can wearable data identify an individual's stress mindset, and can mindset be leveraged to build better performing stress detection models? To investigate that, we conducted an in-lab study (N=23) with wearable devices by inducing mental stress in participants. First, we found that heart rate variability and electrodermal activity features carry signatures of stress mindset. Second, machine learning models can discriminate stress mindset with sensors, achieving AUCs upto 0.88. Finally, a random forest model trained for stress-is-enhancing participants outperformed a one-size-fits-all model (AUC=0.91 vs. 0.78, $p < 0.05$), for the task of stress detection. Our findings show that stress mindset leaves a measurable physiological footprint and that mindset-aware models open the potential for more personalized stress detection and interventions. To support future research, we publicly release the anonymized dataset at <https://social-dynamics.net/stress/mindset>

CCS Concepts

• **Human-centered computing** → Ubiquitous and mobile computing; *Empirical studies in ubiquitous and mobile computing*; *Ubiquitous and mobile devices*; • **Applied computing** → *Health informatics*; • **Computing methodologies** → Supervised learning.

Keywords

mental stress, stress detection, stress mindset, datasets, wearables, wearable sensors, machine learning, momentary stress

ACM Reference Format:

Lakmal Meegahapola, Marios Constantinides, Zoran Radivojevic, Hongwei Li, Michael S. Eggleston, and Daniele Quercia. 2026. Stress Mindset Matters: Rethinking Mental Stress Detection with Multimodal Wearable Sensors. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*, April 13–17, 2026, Barcelona, Spain. ACM, New York, NY, USA, 26 pages. <https://doi.org/10.1145/3772318.3791340>

1 Introduction

Stress is a natural physiological and psychological response to perceived challenges or threats that vary in intensity and duration, influencing both mental and physical health [120]. The ability to automatically detect momentary mental stress is a cornerstone of modern Affective Computing, promising to enable just-in-time interventions that support well-being [6, 15, 90, 94, 117]. In both commercial wearables and a vast body of research, stress is typically inferred from physiological signals captured by wrist sensors (e.g., electrodermal activity, heart rate, and accelerometry) [78, 83, 106, 126]. The standard pipeline involves processing these signals in sliding windows, engineering predictive features, and using machine-learning models to classify a user's state as "stressed" vs. "not stressed". Modern pipelines also include deep learning models such as convolutional neural networks and long short term memory networks [69, 122]. Despite significant progress, this entire paradigm is built on a critical oversimplification. Current pipelines largely ignore stable individual psychological differences that shape how stressors are appraised and responded to. Among these is the stress mindset: a person's core belief about whether stress is enhancing or debilitating [27]. Prior work in Human-Computer Interaction (HCI) and Ubiquitous Computing (UbiComp) rarely discuss the role of stress mindset in understanding mental stress better.

Pioneering work by psychologist Alia Crum [27] established that an individual's stress mindset is a tractable, malleable belief with measurable consequences for their physiology, affect, and performance. A stress-is-enhancing mindset is associated with appraising demands as challenges that can be leveraged for growth, whereas a stress-is-debilitating mindset is associated with threat-oriented appraisals. Consequently, one's mindset can fundamentally tilt the physiological response toward what psychologists term productive



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

CHI '26, Barcelona, Spain

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2278-3/2026/04

<https://doi.org/10.1145/3772318.3791340>

eustress or harmful distress [39, 103]. While these ideas are well-developed in psychology, they remain largely unexplored in broader computing literature, spanning areas such as HCI, UbiComp, and Affective Computing. Technical work in our community has rigorously focused on using novel modalities and examining choices such as time windows, feature engineering, and leakage-safe evaluation [17, 76, 83, 122, 126], but has rarely bridged this foundational psychological theory with sensing and modeling choices. This omission matters: without this crucial psychological context, our models and interventions risk conflating beneficial and harmful arousal, ultimately undermining user trust and efficacy of wearable systems [48].

To close this gap, we conducted an exploratory in-lab study with 23 participants using wearables that captured heart rate (HR), electrodermal activity (EDA), and accelerometer motion (ACC) data, and investigated whether stress mindset is reflected in sensor features. We collected data during alternating periods of induced stress and relaxation, alongside validated psychological questionnaires, including the Stress Mindset Measure [27]. We also explored whether the knowledge on stress mindset can be used to build better performing stress detectors. In line with the overview provided in Figure 1, by analyzing this dataset, we formulated three Research Questions (RQs):

RQ1: Do self-report and wearable features discriminate stress is enhancing versus debilitating mindsets?

RQ2: How accurately can we infer an individual’s stress mindset from user-level wearable features?

RQ3: Does including an individual’s stress mindset improve prediction of momentary stress from wearables?

In answering our three RQs, we made three main contributions:

Contribution 1: We conducted a controlled in-lab study with 23 participants to investigate the relationship between stress mindset and momentary stress (Section 4). Participants wore an Empatica E4 wristband and a Polar H10 chest strap while engaging in two guided meditation sessions and two stress-inducing tasks: the Sing-a-Song stress test (socio-evaluative stress) and the Stroop color-word test (cognitive stress). We pre-processed the dataset at a moving time window-level (2, 5, 10, 20, 30 seconds) and user-level. Using marginal logistic models with generalized estimating equations on the collected sensor data, we show that heart rate and electrodermal activity features carry information about mindset (Section 5.2). Importantly, heart rate variability (HRV) low-frequency power exhibits a robust and statistically significant between-person association with mindset. There were other features from HR, EDA, and ACC that showed high significance before multiple-test correction. These findings establish for the first time that stress mindset is not merely psychological noise, but leaves a measurable signal in sensor features of wearables.

Contribution 2: We showed that a user’s stress mindset can be reliably inferred from physiological data aggregated at the user level (Section 6.2). Using a robust evaluation setup with cross validation, a regularized logistic-regression model achieved a high performance with a mean AUC of 0.88 for a binary classification of enhancing vs. debilitating mindsets. For a more challenging three-class formulation (low, medium, and high mindset scores), a support

vector classifier reached a mean AUC of 0.81. This performance was achieved using user-level features derived from short 2s and 5s windows, demonstrating the feasibility of inferring a user’s stress mindset directly from data captured by consumer-grade wearables.

Contribution 3: We demonstrated two pathways for model personalization based on stress mindset (Section 7.2). First, adding binary stress mindset as a feature to a Random Forest classifier is neutral-to-beneficial, with the clearest gains of around 2% increase of mean AUC, at 20s time window, compared to a model that does not use stress mindset as a feature (Figure 6). However, the results do not show a statistically significant improvement across any time window. Second, training separate stress detection models for each mindset group yields a substantial performance improvement (Figure 7). The model for the stress-is-enhancing group peaked at an mean AUC of 0.91 (10s window), outperforming the one-size-fits-all baseline (mean AUC of 0.83, 30s window). The performance difference was statistically significant. This shows that personalization based on stress mindset is not just theoretically sound, but a practically effective strategy for enhancing the performance of momentary stress detection. We also found that the enhancing group’s model reaches its peak performance at a 10s window, which is shorter than the optimal windows for both the one-size-fits-all model and the debilitating group’s model. This departs from the common recommendation to favor longer windows for wearable stress detection. Practically, this suggests window length could be personalized by mindset rather than fixed a priori.

Building on these findings, we outline a research agenda for mindset-aware stress sensing. Future systems should jointly model stable psychological traits and wearable biosignals by (1) inferring stress mindset from wearables and using it to personalize model and design choices, (2) validating robustness in the wild across populations, stressors, and devices while studying longitudinal change, and (3) translating mindset awareness into just-in-time interventions that distinguish eustress from distress and support reframing for users with a debilitating mindset. Done well, this shift can move the field from generic arousal detection to psychologically grounded personalization, with safeguards that ensure mindset inference empowers users rather than profiles them.

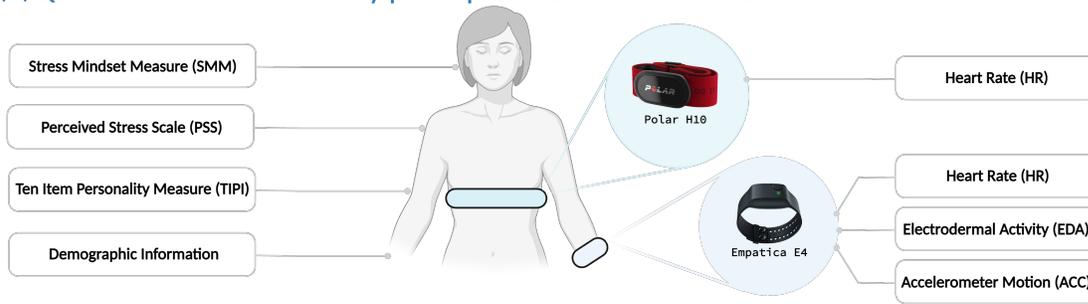
2 Related Work

This section situates our research at the intersection of two distinct but complementary fields: the technical domain of wearable-based stress detection (Section 2.1) and the theoretical domain of stress psychology (Section 2.2). Finally, we synthesize these areas to identify a critical research gap that our work aims to address (Section 2.3).

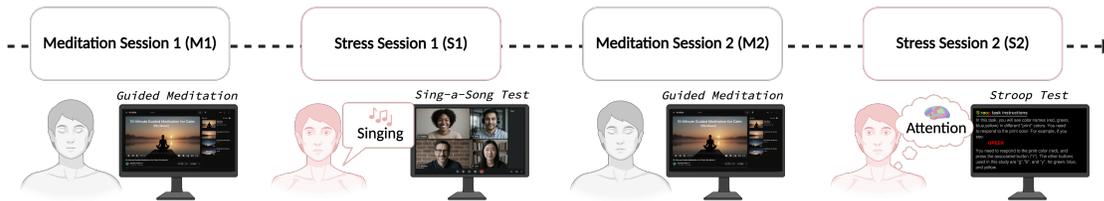
2.1 Wearable-Based Momentary Stress Detection

Research in ubiquitous computing and human-computer interaction has long pursued the goal of detecting human stress from multimodal data captured by wearable devices [31, 51, 88, 101]. This body of work has established a robust pipeline for stress sensing but has primarily concentrated on technical and methodological challenges, often treating stress as a binary construct: high versus

(A) Questionnaires answered by participants and the wearables used



(B) Study setup including the sessions for relaxing and inducing stress



(C) Overview of the experimental setup including dataset pre-processing, RQ1, RQ2, and RQ3

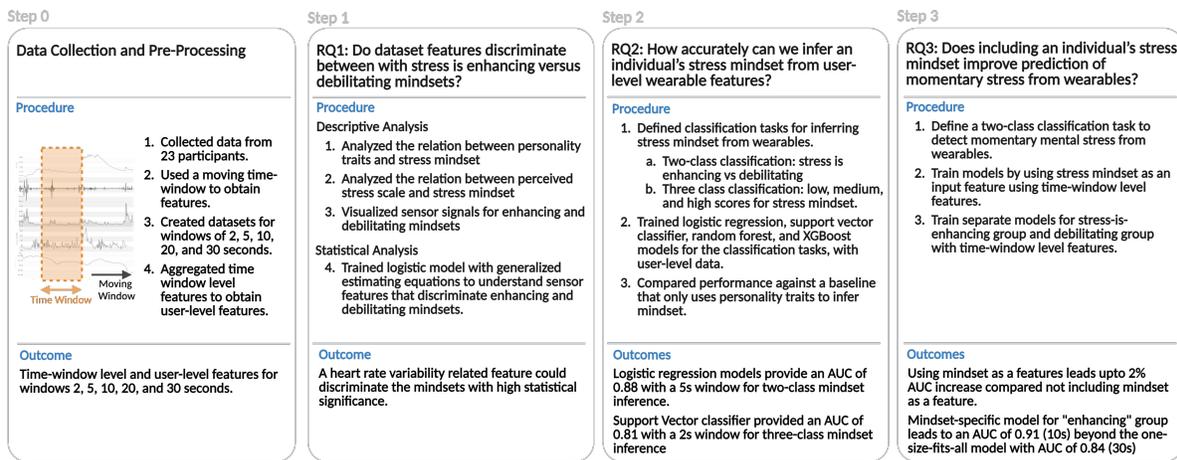


Figure 1: A conceptual overview of the study design, linking data collection, experimental protocol, and the analytical pipeline. (A) Data Collection: The study captured psychological data via self-report questionnaires (Stress Mindset Measure, Perceived Stress Scale, Ten Item Personality Inventory) and sensor data from two wearable devices: an Empatica E4 wristband and a Polar H10 chest strap. **(B) Study Setup:** The in-lab protocol was designed to elicit a range of physiological responses by alternating between guided meditation sessions for relaxation (M1, M2) and two distinct stress-induction tasks, a socio-evaluative stressor (S1: Sing-a-Song Test) and a cognitive stressor (S2: Stroop Test). **(C) Experimental Setup:** First, the collected dataset was pre-processed to obtain time window level and user-level features. Then, it illustrates how the process data were used to systematically answer three research questions: (RQ1) identifying the physiological footprints of stress mindset, (RQ2) inferring a user's mindset from aggregated sensor features, and (RQ3) leveraging mindset to improve momentary stress detection using both feature inclusion and mindset-specific models. Sources and related content.

low [94]. Beyond this, commonly used consumer wearable devices such as Garmin, Fitbit, Apple Watch, and WHOOP also provide momentary stress detection and interventions [106]. Studies have also explored stress detection in diverse cohorts and contexts: student

populations [62, 125], work related contexts [3, 5, 51, 84], training [92], and virtual reality [127]. More recently, researchers have also discussed the dual role of stress management systems, the aim is to reduce stress in individuals, even though paradoxically those systems could cause stress [48].

2.1.1 A Data-Centric Quest for Accuracy. The conventional methodology captures physiological signals via commercial or custom wearable sensors. Devices such as the Empatica E4 wristband measure Electrodermal Activity (EDA), Blood Volume Pulse (BVP) for deriving Heart Rate (HR) and Heart Rate Variability (HRV), and Skin Temperature (ST), and are widely used in research settings [35, 82, 83]. Recent work shows that earable devices can support stress assessment with sensing modalities comparable to Empatica [54]. Across studies, the dominant pipeline moves from signal pre-processing and filtering to feature extraction and classification. It has been studied in controlled laboratory settings using standardized stressors such as the Trier Social Stress Test (TSST) [64] and mental arithmetic tasks [89, 118], and increasingly in "in-the-wild" deployments where contextual confounds can degrade performance, elevating concerns around generalization, robustness, and reproducibility [79, 114, 122, 126]. Much of the literature therefore centers on improving end-to-end accuracy through increasingly sophisticated feature design, including sliding-window statistical summaries, signal decomposition (e.g., separating EDA into tonic and phasic components), and automated feature generation [83, 114, 123]. Complementing this, studies frequently benchmark classifiers ranging from support vector machines and random forests, to deep learning models [122]. Overall, this data-centric trajectory has driven steady gains in detecting momentary stress, while highlighting the fragility of accuracy under real-world variability.

2.1.2 The Personalization Problem as a Statistical Challenge. The HCI community has long recognized that "one-size-fits-all" stress detection is impractical given substantial inter-individual differences in physiological responses [83]. Accordingly, model personalization has become a major research direction [65, 80, 105]. Yet personalization is typically framed as a data-centric, statistical challenge [44]: methods such as transfer learning, on-device fine-tuning with user feedback, and self-supervised learning (SSL) adapt models to each person's bio-signal distribution [55]. For example, Islam et al. [55] showed that SSL can learn an individual's baseline temporal dynamics from unlabeled data, reducing labeled requirements for subsequent fine-tuning. While effective at accommodating statistical variance, these approaches rarely engage with why individuals differ in the first place. This creates a personalization paradox: systems aim to be individual-sensitive, but treat the individual as a statistical black box, adapting to downstream signal patterns rather than modeling upstream drivers such as stable psychological traits that shape physiological stress responses. In parallel, much work operationalizes stress as a monolithic binary state ("stressed" vs. "not-stressed") inferred from wearable sensors alone [101]. Although some studies incorporate high-level context (e.g., activity) to improve accuracy [82], internal psychological context, including how people interpret stressors, remains largely unexplored. In particular, the role of mindset in shaping stress detection pipelines is still understudied.

2.2 The Psychology of Stress: Mindset, Eustress, and Distress

While broader computing literature (including HCI, Ubicomp, Affective Computing) has focused on the technical detection of physiological arousal, the field of psychology has long understood that

the experience of stress is not solely determined by the presence of a stressor but by an individual's interpretation of it. This perspective provides the theoretical foundation for approaching momentary stress detection with a more nuanced, psychologically-grounded understanding.

2.2.1 What is Stress Mindset? Pioneering research by Crum et al. [27] has formalized this interpretive lens into the concept of stress mindset. A mindset is defined as a mental frame that selectively organizes information, guiding an individual's understanding and response to an experience. Stress mindset exists on a continuum between two opposing beliefs: **Stress-is-Enhancing Mindset**: The belief that stress has positive consequences, leading to improved performance, productivity, health, learning, and growth; and **Stress-is-Debilitating Mindset**: The belief that stress is harmful, causing negative outcomes in these same domains and should therefore be avoided. Importantly, this mindset is not an appraisal of a single event but a stable, core belief about the nature of stress itself [58], and it can be measured using validated instruments like the Stress Mindset Measure (SMM) [27]. While it is generally stable, it can be altered with time, similar to how personality traits of individuals could get altered [26, 57].

2.2.2 Physiological and Behavioral Correlates of Mindset. Stress mindset is not merely an abstract belief; it has measurable physiological and behavioral consequences. A stress-is-enhancing mindset is associated with more adaptive physiological responses, including more moderate cortisol reactivity and sharper increases in anabolic ("growth") hormones under stress [26, 27]. It is also linked to behavioral and cognitive patterns such as a greater desire for feedback, increased cognitive flexibility, and an attentional bias toward positive stimuli, particularly when a stressor is appraised as a challenge [26]. Together, these findings motivate our central hypothesis: if stress mindset systematically shapes stress responses, then it should leave detectable signatures in the same wearable sensor data commonly used for stress sensing.

2.2.3 Connecting Mindset to Eustress and Distress. Stress mindset offers a mechanistic lens on the long-standing distinction between "eustress" and "distress" [36]. Eustress is a positive, motivating response to challenge that can enhance performance and well-being, whereas distress is a negative response associated with anxiety and impaired performance [59]. Although both can involve similar initial physiological arousal (e.g., cortisol and adrenaline release), they diverge in experienced valence and downstream outcomes. Stress mindset mediates this divergence: a stress-is-enhancing mindset biases appraisal toward challenge and thus eustress, while a stress-is-debilitating mindset biases appraisal toward threat and thus distress [36, 59, 61]. This framing suggests a hierarchy in which stable mindset shapes momentary appraisals. If we can infer this underlying mindset, we may better anticipate a user's propensity for distress across situations and design more proactive interventions than those triggered by detecting a single episode. In this paper, we address this gap by investigating whether wearable sensor data contains signatures that differentiate stress mindsets and by evaluating the feasibility of inferring an individual's mindset from those signals.

Table 1: Positioning this work in contrast with multiple research themes. Each row clarifies what is measured, over what time scale, how it is labeled, and what the typical contribution is.

Research theme	Target construct	Temporal granularity	Typical labels / instruments	What is inferred / key contribution
Momentary stress [83, 122, 126]	Physiological arousal (elevations in autonomic activity measured from the body)	Event-level, within seconds to minutes (short time windows during or around a task)	Stress induced with laboratory tasks such as the Trier Social Stress Test (TSST), the Stroop color-word task, or the Sing-a-Song test; in-the-moment self-reports (ecological momentary assessments)	Determines whether a person is currently “stressed” versus “not stressed” from wearable signals, and evaluates accuracy and across-person generalization under different modeling choices.
Eustress vs. distress [10]	Stress appraisal (whether a specific stress response is experienced as helpful/challenging eustress or harmful/threatening distress)	Episode-level, over minutes to hours (during a single activity or period)	Self-reported appraisal collected during or after the episode; affect scales such as the Self-Assessment Manikin, Circumplex Model, or the Positive and Negative Affect Schedule	Classifies the valence of a particular stress episode (challenge/eustress versus threat/distress), going beyond arousal to interpret how the stressor is perceived.
Longitudinal stress [18]	Perceived stress as an ongoing burden (how stressful life feels overall)	Days to weeks (trend over time rather than a single moment)	Periodic surveys such as the Perceived Stress Scale (PSS) or daily/weekly diaries	Forecasts past/near/future survey-based stress scores and identifies behavioral patterns associated with higher or lower perceived stress over time.
This work: Mindset about Stress [27]	Stress mindset, a relatively stable belief about whether stress is enhancing or debilitating	Stable trait observable over months to years (not tied to a single episode)	Stress Mindset Measure (SMM) questionnaire	Shows that stress mindset leaves a measurable signature in wearable signals, enables accurate inference of an individual’s mindset, and demonstrates that incorporating this trait improves momentary stress detection via mindset-aware models.

2.3 Towards Mindset-Aware Stress Sensing

The above sections point to a disciplinary divide: HCI has deep expertise in sensing and modeling physiological data, yet often abstracts away the psychological richness of stress, while psychology offers nuanced theories such as stress mindset but lacks tools for continuous, objective measurement. Our work sits at this intersection and argues that incorporating constructs like mindset is a natural next step toward truly personalized and meaningful stress sensing. In doing so, we add nuance to stress tracking and respond to calls in HCI for more transparent and informative stress measures [111]. Beyond binary stress detection, Adler et al. [4] inferred stress resilience in resident physicians using a process-oriented definition, operationalized through trajectories of mental health under prolonged stress in line with [47]. Closest to our work is computational modeling of stress appraisal [75]. For example, Awada et al. [10] classify appraisal outcomes (boredom, eustress, distress) from wearable data, advancing beyond generic stress labels. However, such approaches primarily model situational outcomes rather than the stable trait that shapes them (see Table 1). As discussed in Section 2.2.3, stress mindset is distinct from whether a particular stressor is experienced as eustress or distress; we argue that modeling mindset provides a more foundational and potentially more generalizable account of individual stress patterns. We therefore propose that the next generation of stress technologies should be *mindset-aware*. Analogous to systems that adapt to cognitive workload or attentional focus [93], a mindset-aware system would adapt to a user’s stress mindset and deliver qualitatively different support. For users with a debilitating mindset, it may prioritize reframing strategies; for users with an enhancing mindset, it may emphasize opportunities for growth. By moving beyond a simple “you are stressed” notification, mindset-aware sensing can enable

more meaningful and actionable interventions. We view this study as a foundational step in that direction.

Research Gap. Prior work has largely overlooked stress mindset as a driver of the individual differences that data-driven personalization in wearable stress detection seeks to capture. As a result, it remains unknown whether a stable trait like stress mindset leaves a discernible footprint in wearable sensor data, and whether such information can be leveraged to build more nuanced and effective stress sensing systems. This study is the first to systematically address this gap (Table 1).

3 User Study

To investigate our research questions, we designed and conducted a controlled in-lab user study.

3.1 Participants and Recruitment

We recruited 25 participants via convenience sampling from a university and a multi-national corporation, using internal mailing lists. After data quality checks, two participants’ data were excluded due to sensor signal corruption, resulting in a final sample of $N=23$ for our analysis. This sample size is consistent with prior similar studies in HCI and Ubicomp [20, 34, 63, 95]. Table 2 summarizes the demographics of our final participant group. The sample averaged 32.7 ± 6.0 years of age (range: 23–52). It included 15 males and 8 females and was ethnically diverse, though predominantly Asian ($n = 14, 61\%$). Educational attainment was high, with the majority holding a PhD or equivalent ($n = 17, 74\%$). Based on the Stress Mindset Measure, 13 participants (57%) were categorized as having an enhancing stress mindset, while 10 (43%) were categorized as having a debilitating mindset, providing a balanced distribution for our analysis. Moreover, Table 6 in Appendix co-summarizes the demographics of participants at an individual level.

Table 2: Participant demographics and scale statistics. Ages computed as 2025– birth year.

Age		Employment	
Mean \pm SD	32.7 \pm 6.0	No, I am a student	16
Range	23–52	Yes, I am employed	7
Gender identity		Ethnicity	
Male	15	Asian	14
Female	8	White, Caucasian	6
		Another ethnic group	1
Education level		Mixed/multiple ethnic groups	1
PhD or equivalent	17	Not specified	1
Master's or equivalent	3	Stress mindset	
Bachelor's or equivalent	2	Enhancing (\geq 16)	13
High school	1	Debilitating ($<$ 16)	10

Table 3: Descriptive statistics (Mean, Standard Deviation, Min, Max) for the Ten Item Personality Inventory (TIPI), Perceived Stress Scale (PSS), and Stress Mindset Measure (SMM) ($N = 23$).

Scale	Mean (SD)	Min, Max
<i>Ten Item Personality Inventory</i>		
Extraversion	3.80 (1.13)	1.50, 5.50
Agreeableness	4.89 (1.08)	3.00, 6.50
Conscientiousness	5.20 (1.21)	2.00, 7.00
Emotional Stability	4.80 (1.40)	2.00, 7.00
Openness	5.41 (0.96)	4.00, 7.00
<i>Perceived Stress Scale</i>		
Total Score	15.43 (5.79)	5.00, 31.00
<i>Stress Mindset Measure</i>		
Total Score	14.74 (6.76)	2.00, 28.00

3.2 Apparatus and Materials

The study was conducted in a quiet, sound-attenuated room. Participants were seated at a desk in front of a 24-inch monitor, which displayed a custom web-based application that guided them through the experiment. Physiological data was collected using two consumer-grade wearable devices: a Polar H10 chest strap to capture electrocardiography (ECG) for robust heart rate and HRV metrics, and an Empatica E4 wristband worn on the non-dominant wrist to capture EDA, BVP, skin temperature, and motion data via ACC. The chest ECG sensor on Polar H10 provides more precise R–R intervals for heart rate variability analysis, complementing the wristband's convenience, and was used to validate the measurements of the more noisy wrist wearable. However, after verifying the primary analysis of this paper was performed with the wrist wearable because that is the most common setup for studying stress in both research and consumer-grade wearables from companies such as Apple, Google, and Garmin.

3.3 Procedure

The study procedure was refined based on a pilot study conducted with 2 participants to ensure a seamless experience and validate

the timing of each session. Protocol was improved based on their feedback, before finalizing the larger study.

3.3.1 Pre-Study Questionnaires. A couple of days before their scheduled session, participants completed a set of online questionnaires. This was done to minimize fatigue and cognitive load on the day of the study. Questions asked in this questionnaire are as follows:

- **Stress Mindset Measure (SMM):** The 8-item SMM, developed by Crum et al. [27], assesses individuals' core beliefs regarding the effects of stress, specifically, whether they view stress as predominantly enhancing or debilitating. As a pragmatic choice for most sections, a summed score of 16 (mid-point of the scale) or greater classifies participants as having a stress-is-enhancing mindset, while scores below 16 indicate a stress-is-debilitating mindset [61]. This threshold also aligns with the median of SMM scores. Please see Section 8.3.2 for more information on this. The SMM has demonstrated good reliability and construct validity in measuring these beliefs.

- **Perceived Stress Scale (PSS):** To measure general, baseline stress levels, we administered the PSS [22], a widely validated and adopted scale in both psychology and computing research. The Perceived Stress Scale measures the degree to which individuals appraise situations in their lives as stressful, specifically focusing on their perception of unpredictability, uncontrollability, and overload during the past month. It assesses subjective stress, that is how people feel about the general stress in their lives, rather than specific life events or external stressors.

- **Ten-Item Personality Inventory (TIPI):** To account for the potential influence of individual personality differences, we employed the TIPI [43], a concise instrument designed to assess the Big Five personality traits: Extroversion, Agreeableness, Conscientiousness, Emotional Stability, and Openness to Experience. The TIPI was selected for its efficiency, making it well-suited for contexts where time is limited, and for its established validity and acceptable test–retest reliability as demonstrated both in general psychology and HCI research

- **Demographics and Consent:** We collected demographic information (age, gender, education) and confirmed that participants had no prior diagnosis of any major mental health conditions. Participants provided explicit informed consent for their data to be used in the study and to be shared publicly in an anonymized format. Finally, Table 3 provides a summary of responses obtained in pre-study questionnaires.

3.3.2 In-Lab Protocol. Upon arrival, participants were briefed on the procedure and fitted with the wearable sensors. To ensure comfort and proper placement, researchers of the same gender as the participant were available to assist. The in-lab session, lasting approximately one hour, was guided by the on-screen instructions and consisted of four sequential parts (Figure 1).

- **Meditation Session 1 (M1):** To establish a physiological baseline, participants began with a 5-minute guided audio meditation. This is a common technique used in prior work to induce a state of relaxation [17, 108].

- **Stress Session 1 (S1) – Sing-a-Song Stress Test:** To induce socio-evaluative stress, we employed the Sing-a-Song Stress Test (SSST) [112], a paradigm designed to elicit mental stress in a controlled

manner. The SSST serves as a brief, low-cost alternative to more time and labor-intensive protocols like the Trier Social Stress Test (TSST) [64], while still reliably inducing social-evaluative stress. In our implementation, participants were informed that their vocal performance would be evaluated by the in-person researcher and a remote two-person panel observing via video conference. In this procedure, participants are first presented with neutral messages on the screen and wait in short intervals of one minute. Then, the task began with a another one-minute timer, during which participants were asked to choose a song and prepare to sing. Immediately following this preparation period, they were instructed to sing their chosen song aloud for 90 seconds. This sudden demand to perform under social evaluation is designed to elicit self-conscious stress, triggering measurable physiological responses such as increases in heart rate and skin conductance.

- **Meditation Session 2 (M2):** Following the first stressor, a second, longer 10-minute guided meditation was used as a recovery period. The extended duration was a direct result of our pilot studies, which revealed that participants required around 3 minutes to allow their physiological signals to return to a baseline state after the high-arousal S1 session.
- **Stress Session 2 (S2) — Stroop Color-Word Test:** To induce cognitive stress, participants completed a computerized version of the Stroop Color-Word Test [99, 102]. This classic task induces stress by creating cognitive interference, where participants must name the color of the ink a word is printed in while ignoring the word itself (e.g., saying "blue" for the word "RED" printed in blue ink). We used a publicly available implementation from PsyToolkit for this¹.

3.3.3 Compensation. Upon completion, participants were debriefed and compensated with an approximately \$35 USD amazon gift voucher, a rate well above the local minimum wage. Light refreshments were also provided.

4 Methodology for Data Pre-Processing

Our experimental design is structured to systematically answer our three research questions. We begin by describing the common data pre-processing and feature engineering pipeline before detailing the specific statistical and machine learning methodologies used for each research question. The overall pipeline is depicted in Figure 1C.

4.1 Dataset Pre-Processing

For each of the 23 participants, raw physiological data was read directly from the 'empatica.zip' archives, which contain separate files for electrodermal activity, heart rate related information, skin temperature, and 3-axis accelerometry. A critical first step was timestamp synchronization. As the Empatica E4 records in UTC, we applied a X-hour offset to align the sensor data with the local time of the study, ensuring accurate alignment with the ground truth event markers.

¹Publicly available test can be accessed here: https://www.psychtoolkit.org/experiment-library/experiment_stroop.html

4.2 Time-Window Level Feature Extraction

We processed the synchronized time-series data using the open-source 'flirt' toolkit [35]. This is a library commonly used in signal processing pipelines for wearable devices with modalities such as EDA, ACC, and HR [11, 12, 56, 107, 109, 116, 121], and has been widely used in stress detection tasks [14, 87, 96, 110]. To investigate the impact of temporal granularity, which has been studied as an important aspect in stress detection pipelines [85, 100, 113], we extracted features across multiple sliding window configurations: 2s, 5s, 10s, 20s, and 30s (with corresponding step sizes of 1s, 2s, 5s, 10s, and 20s). This multi-window approach allows us to balance the need for temporally precise features against the stability required for robust feature calculation (e.g., for frequency-domain HRV metrics). For each window, flirt computes a comprehensive set of time-domain, frequency-domain, and non-linear features across all sensor modalities. The resulting continuous stream of feature windows was then segmented using the ground truth timestamps from our experimental protocol. This process isolated the feature windows corresponding to each of the four sessions (M1, S1, M2, S2) for every participant. To minimize artifacts from session transitions, we trimmed the first and last 30 seconds of data from each session's feature set. A detailed summary of extracted features is provided in Appendix B, Appendix C, and D.

4.3 User-Level Feature Aggregation

To answer RQ1 and RQ2, which concern the stable, trait-like construct of stress mindset, we aggregated the time-window level features into a single, static feature profile for each participant. This is a necessary step because we hypothesize that a user's mindset is reflected not in momentary fluctuations, but in their average physiological patterns and reactivity to stressors. The aggregation process involved two steps:

- **Summarizing by Session Mean:** We first calculated the mean of every feature across all windows within each session (M1, S1, M2, S2). This condensed the time-series data into a set of representative signatures for each participant under each condition (e.g., 'hrv_lf_S1', 'eda_phasic_mean_M2').
- **Creating Contrast Features:** To explicitly model physiological reactivity, we engineered a set of contrast features by calculating the difference between the aggregated session means, in line with prior work [33, 68, 86, 119]. These included direct contrasts (e.g., 'S1 - M1') and a general contrast comparing the average of the stress sessions to the average of the relaxation sessions ('Stress_mean - Rest_mean'). These delta features are powerful as they normalize for baseline individual differences.

This pipeline produced a final, 1188 static feature vector for each of the 23 participants, representing their unique physiological response profile for subsequent analysis.

5 Do self-report and wearable features discriminate stress is enhancing versus debilitating mindsets? (RQ1)

5.1 Methods

To answer our first research question, we began with a descriptive analysis to visually inspect for initial differences between stress is

enhancing and debilitating groups and rule out obvious confounds like personality or baseline stress levels (Section 5.1.1). Following this, we conducted a statistical analysis using Generalized Estimating Equations to formally test for significant associations between specific physiological features and stress mindset (Section 5.1.2).

5.1.1 Descriptive Analysis. First, we examined how stress mindsets differ for participants' personality traits. Using provide TIPI questionnaire answers, we derived personality trait scores for each participant, along the traits Extroversion, Agreeableness, Conscientiousness, Emotional Stability and Openness. Then, we separated participants with stress is enhancing and debilitating mindsets into two groups, and plotted their personality trait scores along the two axis in the radar plot provided in Figure 2. Then, we used the PSS questionnaire and derived the baseline stress level of participants and binned them into low, moderate, and high using scores calculated based on [2, 7, 66]². Then, we again separated the participants into stress is enhancing and debilitating groups, and plot the distribution of participants for different baseline PSS scores, as a bar plot in Figure 3. Finally, using the feature extraction process mentioned above, we obtained multiple datasets corresponding to different time windows. While it is not possible to show all data distributions across different features, in Figure 4, we provide sensor signal distributions for four features, stratified by stress mindset.

5.1.2 Statistical Analysis. In the descriptive analysis above, we observed that for certain sensor features, the distributions of sensor values for the two stress mindsets could be slightly different. We then quantified how different these sensor-data distributions are. Specifically, we analyzed each session's data to test whether physiological features derived from heart activity, electrodermal activity, and accelerometry were associated with a binary stress mindset label: Enhancing vs. Debilitating. Hence, to statistically quantify the difference in sensor distributions between the two mindset groups, we utilized Generalized Estimating Equations (GEE) [67] with an exchangeable working correlation [45]. This approach is specifically designed for repeated measures, allowing us to account for the dependencies inherent in multiple time windows collected from the same participant. We utilized a within/between decomposition method to separate transient fluctuations from stable traits. Since stress mindset is a constant attribute for each individual, our analysis focused on the between-person effects, testing whether a participant's average physiological response was significantly associated with an Enhancing or Debilitating mindset. We report these associations as Odds Ratios (OR) and applied the Benjamini-Hochberg correction to control for false discovery rates arising from multiple hypothesis testing. While models were fitted for all sessions, we report results primarily for the Sing-a-Song stress task (S1), which yielded the most discriminating features. For the complete mathematical formulation regarding the GEE model and variance decomposition, please refer to Appendix E.

5.2 Results

Figures 2 and 3 establish that stress mindset is a distinct psychological construct, independent of both Big Five personality traits and baseline perceived stress. As shown in Figure 2, the personality

²These bands are conventional, not diagnostic, and are provided to aid interpretation.

profiles of the Enhancing (N=13) and Debilitating (N=10) groups are broadly similar. While the Debilitating group trends higher on Openness (+0.45 on a 1-7 scale) and the Enhancing group is slightly higher on Conscientiousness (+0.31), these differences are small to moderate, not statistically significant, and the profiles largely overlap. In practice, this could mean that even a detailed personality profile is not a reliable shortcut for inferring how an individual will interpret stress; a system cannot simply use a personality test to personalize its stress interventions. Later in Table 5, we indeed test this by using personality traits to infer stress mindset, as a baseline. Similarly, Figure 3 shows no statistically significant difference in the distribution of baseline PSS scores between the two mindset groups. Practically, this suggests that a person's core belief about stress is separate from their current level of perceived stress. A system cannot assume that a user with high chronic stress automatically has a debilitating mindset, or vice versa. Together, these findings underscore the unique value of measuring stress mindset directly. They confirm, in line with prior work [21, 25, 81], that mindset is not redundant with other common psychological measures, justifying the need for the novel physiological sensing approach we explore in this paper.

Our statistical analysis, detailed in Table 4 reveals that stress mindset leaves a discernible physiological footprint during the socio-evaluative stress task. The clearest signal emerged from Heart Rate Variability (HRV), where Low Frequency (LF) power exhibited a robust and statistically significant association with a debilitating mindset (FDR $q = 0.033$). While the interpretation of HRV LF is physiologically complex, reflecting modulations from both sympathetic and parasympathetic branches [38, 49], this result points to a systematic difference in autonomic regulation. It suggests that a debilitating mindset is associated with a distinct, and potentially less efficient, pattern of autonomic modulation when facing a stressor. Beyond HRV, we observed suggestive patterns in Electrodermal Activity (EDA), though these did not withstand strict FDR correction. Several EDA tonic statistics (e.g., mean, sum) showed a positive association with an enhancing mindset (unadjusted $p \approx 0.04$). Practically, this hints that an enhancing mindset is not characterized by a lack of arousal, but rather by a state of engaged, alert readiness. Visualizations of some of these feature distributions are provided in Figure 4. Overall, these findings provide the first evidence that the stable psychological trait of stress mindset is reflected in measurable wearable bio-signals, motivating our next step: determining if these signals allow for machine learning-based inference.

6 How accurately can we infer an individual's stress mindset from user-level wearable features? (RQ2)

6.1 Methods

In RQ2, we ask whether physiological signals are strong and consistent enough to infer stress mindset from wearable data at the user level. For each participant, we derived feature vectors from HR, EDA, and related streams across four blocks (two stress-elicitation: S1, S2; two neutral/mindset: M1, M2), and also construct an ALL representation that concatenates baseline and reactivity information into a single feature vector. Sensor data are summarized in short, overlapping windows and then aggregated with robust statistics,

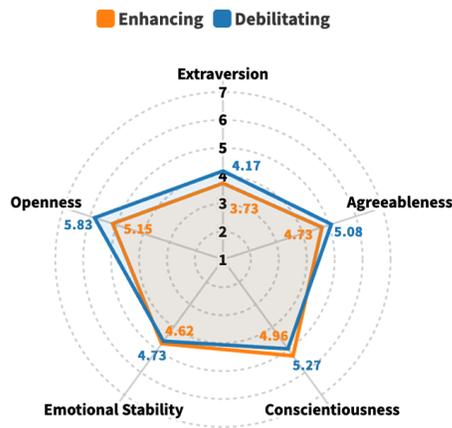


Figure 2: Radar plot of Big Five personality traits derived using TIPI questionnaire, stratified by the stress mindset.

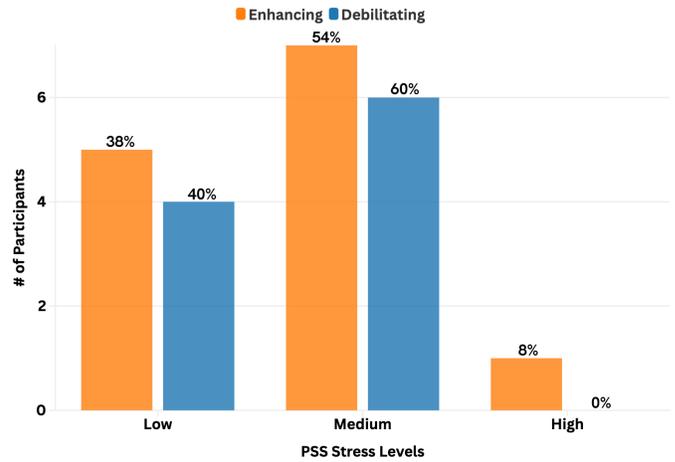


Figure 3: Bar plot different levels of Perceived Stress Scale (PSS) stratified by the stress mindset. Categories for perceived stress scale were derived based on widely accepted thresholds: Low: 0–13, Moderate: 14–26, High: 27–40. The percentage of participants within enhancing and debilitating groups, for low, medium, and high levels too are shown on top of the bars.

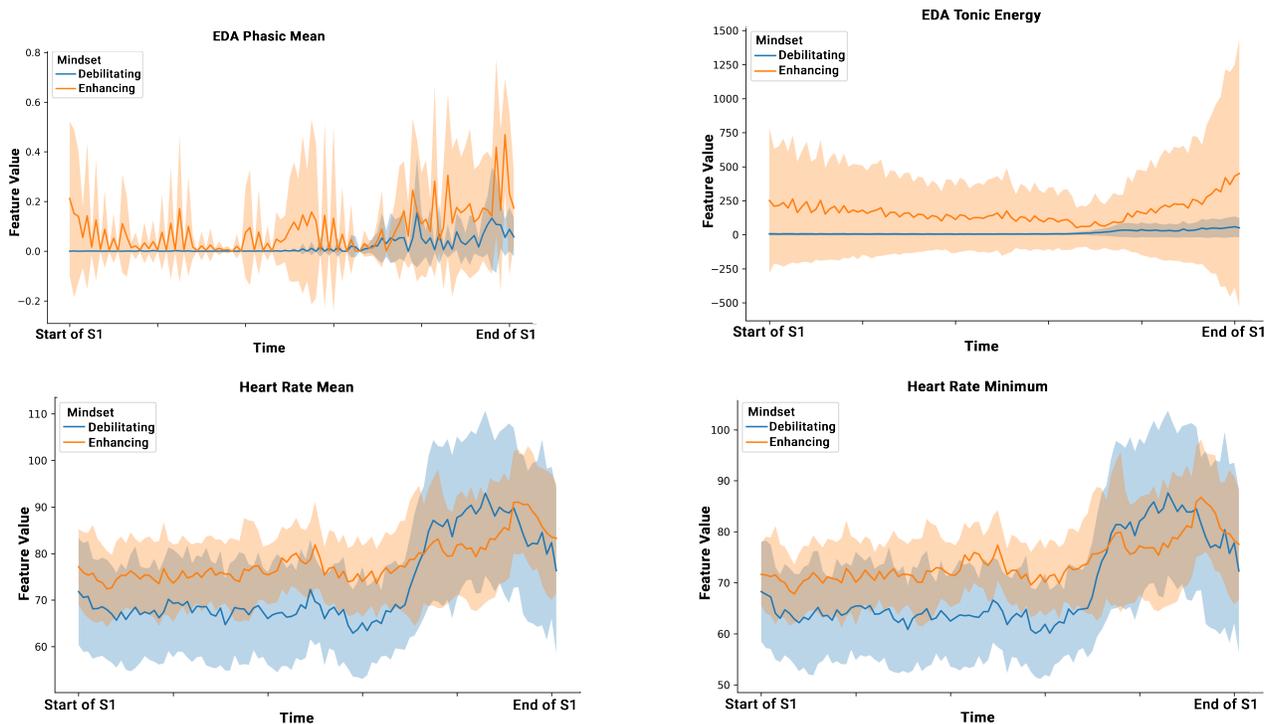


Figure 4: Four figures providing the distribution of sensor features for stress is enhancing and debilitating mindsets during S1 session. Since timelines for participants were different for the session, participant timelines were normalized to the mean time taken by participants. Top two figures show two EDA features whereas the bottom two figures show two HR features.

Table 4: Between-person associations with stress mindset (Session 1). Odds ratios (OR) > 1 indicate higher odds of an Enhancing mindset. Columns: OR = odds ratio for a one-unit increase in the feature (if predictors were standardized, interpret as per 1 SD); CI low/high = Wald 95% confidence limits; z = Wald z statistic; p = two-sided p -value; q = Benjamini–Hochberg FDR-adjusted p across tests in this table; Sig. = significance flag where * denotes $q < 0.05$. GEE (logit, exchangeable) with within/between decomposition; interpretation focuses on the between-person term. NNI corresponds to Normal-to-Normal Intervals.

Feature	OR	CI low	CI high	z	p	q	Sig.
Electrodermal Activity (EDA)							
EDA Tonic Minimum	3.07694	1.02859	9.20441	2.010	0.0444	0.29	
EDA Tonic 5th Percentile	3.08159	1.02786	9.23880	2.009	0.0445	0.29	
EDA Tonic Sum	3.14965	1.01588	9.76523	1.987	0.0469	0.29	
EDA Tonic Mean	3.14965	1.01588	9.76523	1.987	0.0469	0.29	
EDA Tonic RMS	3.15386	1.01510	9.79885	1.986	0.0470	0.29	
Heart Rate (HR)							
HRV Low Frequency Power	0.00125	0.000037	0.04164	-3.736	0.000187	0.0325	*
HRV standard deviation of NN intervals	0.11113	0.02636	0.46843	-2.993	0.00276	0.156	
HRV Standard Deviation	0.10601	0.02416	0.46507	-2.975	0.00293	0.156	
HRV NNI between successive R-waves	0.09637	0.01921	0.48355	-2.843	0.00447	0.156	
HRV Peak to Peak Intervals	0.09637	0.01921	0.48355	-2.843	0.00447	0.156	
Accelerometer (ACC)							
ACC X 95–5 Percentile Range	0.03951	0.00359	0.43514	-2.640	0.00830	0.179	
ACC X Standard Deviation	0.04849	0.00496	0.47366	-2.603	0.00925	0.179	
ACC X Interquartile Range	0.03243	0.00231	0.45550	-2.543	0.0110	0.189	
ACC Magnitude 95–5 Percentile Range	0.09048	0.01292	0.63370	-2.419	0.0155	0.225	
ACC X Sign Flips	0.00106	2.79e–06	0.40243	-2.260	0.0238	0.29	

alongside within-subject contrast features (for example, stress–rest differences), yielding one high-dimensional feature vector per participant and session that is suitable for user-level modeling. We assign user-level labels from the Stress Mindset Measure (0–32) using both a binary split at 16 (low vs. high mindset, matching the canonical debilitating vs. enhancing framing) and a three-class formulation based on score terciles (low vs. mid. vs. high mindset) to probe whether the “middle” group is intrinsically harder to separate. While stress mindset is commonly interpreted with the binary setup, we explored the three-class setup for comprehensiveness, because prior work in mental health have explored two and three class formulations for different constructs [77]. To avoid information leakage, all pre-processing steps (feature filtering, median imputation, and z-scoring) are fit only on the training data within each cross-validation fold and then applied to the corresponding test data. We compare four standard classifiers (logistic regression, SVM, random forest, XGBoost) using grouped 5-fold cross-validation with participants as groups, and report macro AUC (binary) and macro one-vs-rest AUC (three-class) as mean \pm standard deviation across folds to capture variability over repeated train–test splits, in line with standard practice. The results are presented in Table 5. The best-performing configuration is an ALL, 5 s window logistic regression model, whose cross-fold coefficients we later analyze to identify the most mindset-informative physiological features, shown in Figure 5. Full feature definitions, model settings, and additional metrics are provided in the Appendix F.

6.2 Results

A few consistent patterns emerge from Table 5. First, linear models are surprisingly strong in the binary setting. LR achieves the top two-class performance in most rows, peaking at when using all sessions (ALL) using 5s window with an AUC of 0.88 ± 0.15 , and remaining competitive across other windows (e.g., ALL, 2s: 0.82 ± 0.11 ; ALL, 30s: 0.77 ± 0.20). This suggests that, after user-level aggregation, stress mindset is largely linearly separable in the engineered HR, EDA, and ACC feature space. In low- N settings, linear margins with regularization often outperform flexible learners that overfit noise, aligning with prior findings in physiological computing that simple linear models are strong baselines for cross-subject generalization [104]. Second, three-class inference is harder but margin-based SVC excelled in specific contexts. AUCs are generally lower than two-class AUCs, as expected. Nonetheless, SVC attains 0.81 ± 0.11 when using S1 session data at 2s time window, the highest three-class performance in the table, and reaches 0.70 ± 0.25 for ALL, 5s. This indicates that when class structure is three-class, a margin-maximizing nonlinear boundary can delineate terciles better than a purely linear boundary in LR. That SVC leads in S1 (but not consistently in S2/M-blocks) may reflect that S1 session, with sing-a-song stress test, captures clearer and sharper autonomic responses with cleaner signal-to-noise for three-class separation, as compared using including M1, M2, or S2 data. Further, short windows (2s–5s) are often optimal for stress mindset inference. The best binary AUCs occur at 2s–5s windows. Longer windows (20s–30s) frequently dilute transient responses and can

Table 5: Model performance (AUC) for stress mindset inference across sessions and time windows. Two-class (left) and Three-class (right) results presented. Values are mean±SD over grouped CV folds where groups = participant. Color in purple and magenta indicate the best model per row for the respective two or three class setting, respectively. Underlined are the best models across all session and window combinations, for two and three class settings.

Session	Window	Two-Class (AUC)				Three-Class (AUC)			
		LR	SVC	RF	XGB	LR	SVC	RF	XGB
Baseline1		.50 ± .00	.50 ± .00	.50 ± .00	.50 ± .00	.50 ± .00	.50 ± .00	.50 ± .00	.50 ± .00
Baseline2: Personality		.33 ± .27	.56 ± .13	.35 ± .27	.41 ± .31	.65 ± .08	.57 ± .09	.47 ± .16	.55 ± .05
ALL	2 s	.82 ± .11	.47 ± .12	.27 ± .17	.53 ± .24	.52 ± .07	.60 ± .17	.50 ± .14	.55 ± .05
	5 s	.88 ± .15	.63 ± .31	.32 ± .24	.38 ± .35	.63 ± .07	.70 ± .25	.51 ± .20	.66 ± .09
	10 s	.82 ± .26	.41 ± .17	.47 ± .29	.45 ± .28	.70 ± .16	.49 ± .16	.55 ± .26	.68 ± .09
	20 s	.70 ± .19	.50 ± .11	.55 ± .27	.23 ± .25	.56 ± .09	.63 ± .09	.44 ± .20	.53 ± .12
	30 s	.77 ± .20	.17 ± .21	.70 ± .16	.47 ± .40	.50 ± .00	.63 ± .08	.52 ± .16	.46 ± .15
S1	2 s	.73 ± .17	.33 ± .11	.67 ± .24	.70 ± .19	.66 ± .18	.81 ± .11	.58 ± .06	.66 ± .11
	5 s	.73 ± .17	.38 ± .18	.72 ± .31	.67 ± .24	.69 ± .14	.56 ± .15	.57 ± .21	.65 ± .15
	10 s	.62 ± .29	.43 ± .17	.55 ± .31	.42 ± .21	.71 ± .15	.52 ± .27	.55 ± .10	.62 ± .14
	20 s	.50 ± .37	.43 ± .36	.35 ± .27	.27 ± .31	.71 ± .15	.35 ± .09	.48 ± .14	.57 ± .11
	30 s	.67 ± .42	.47 ± .31	.62 ± .29	.35 ± .23	.48 ± .16	.44 ± .28	.42 ± .13	.45 ± .11
S2	2 s	.72 ± .16	.38 ± .23	.47 ± .32	.53 ± .24	.63 ± .11	.59 ± .08	.56 ± .16	.60 ± .11
	5 s	.62 ± .33	.63 ± .19	.50 ± .26	.57 ± .34	.63 ± .12	.61 ± .07	.56 ± .11	.53 ± .09
	10 s	.60 ± .34	.28 ± .30	.27 ± .13	.43 ± .36	.60 ± .11	.62 ± .17	.65 ± .09	.59 ± .12
	20 s	.58 ± .27	.28 ± .27	.50 ± .33	.50 ± .37	.66 ± .14	.69 ± .06	.61 ± .10	.60 ± .13
	30 s	.75 ± .32	.40 ± .34	.70 ± .40	.50 ± .33	.68 ± .07	.67 ± .12	.62 ± .18	.65 ± .16
M1	2 s	.65 ± .23	.33 ± .35	.58 ± .34	.53 ± .40	.44 ± .12	.59 ± .12	.28 ± .12	.50 ± .07
	5 s	.70 ± .27	.60 ± .31	.70 ± .27	.75 ± .22	.69 ± .01	.61 ± .14	.46 ± .12	.60 ± .08
	10 s	.77 ± .20	.37 ± .40	.77 ± .17	.60 ± .08	.57 ± .13	.52 ± .09	.51 ± .12	.65 ± .17
	20 s	.73 ± .23	.52 ± .41	.63 ± .31	.57 ± .08	.60 ± .08	.67 ± .12	.52 ± .03	.40 ± .14
	30 s	.77 ± .20	.40 ± .34	.73 ± .23	.43 ± .34	.61 ± .08	.55 ± .07	.49 ± .12	.45 ± .07
M2	2 s	.63 ± .22	.38 ± .23	.53 ± .32	.42 ± .33	.53 ± .14	.61 ± .05	.51 ± .13	.67 ± .19
	5 s	.48 ± .31	.65 ± .26	.60 ± .27	.53 ± .32	.48 ± .03	.46 ± .14	.47 ± .11	.51 ± .12
	10 s	.47 ± .19	.16 ± .15	.67 ± .24	.40 ± .34	.39 ± .16	.44 ± .08	.41 ± .17	.63 ± .25
	20 s	.53 ± .29	.28 ± .23	.50 ± .33	.47 ± .34	.38 ± .10	.48 ± .17	.29 ± .12	.51 ± .11
	30 s	.67 ± .24	.43 ± .18	.55 ± .27	.63 ± .34	.44 ± .05	.58 ± .16	.25 ± .07	.41 ± .11

blur condition-specific dynamics. Interestingly, it is worth noting that most previous stress detection studies [32, 37, 50, 63, 82], suggests using larger windows of around 20-30s or above.

The feature analysis in Figure 5 reveals that stress mindset leaves a distinct multimodal footprint. We found that an enhancing mindset is associated with regulated autonomic responses (via HRV contrast features) and sustained engagement (via phasic EDA), consistent with the theory that adaptive appraisals mobilize energy efficiently [30, 40]. Conversely, a debilitating mindset is characterized by irregular body movement (accelerometer features), potentially reflecting behavioral restlessness or agitation [8, 74]. The cross-fold stability of these features confirms that these are not random fluctuations, but consistent physiological signatures of how individuals internalize stress. For more details on this, please refer to the Appendix G.

7 Does including an individual’s stress mindset improve prediction of momentary stress from wearables? (RQ3)

7.1 Methods

For RQ3, we examine whether knowing a person’s stress mindset improves momentary stress detection from wearable sensors, a

core task in prior stress inference work [82, 83, 122, 126]. Using the same windowed HR, EDA, and ACC features from our stress pipeline, we compare two ways of incorporating mindset. First, we treat mindset as a static covariate by adding the scalar stress mindset score to the sensor feature space and asking whether this questionnaire-derived prior improves within-window stress discrimination. Second, we train separate “mindset-specific” models by splitting participants into enhancing versus debilitating groups and fitting dedicated stress detectors for each group, alongside a one-size-fits-all baseline. In both settings, we use a Random Forest classifier, selected after preliminary comparisons with XGBoost, multilayer perceptrons, and support vector machines, as RFs are strong tabular time-series baselines [53, 70, 77]. To simulate deployment on unseen users and avoid subject leakage, we adopt a nested StratifiedGroupKFold design with grouped 5-fold outer and 3-fold inner cross-validation, stratifying by the binary stress label and grouping by user. The inner loop performs randomized hyperparameter search, and the best model is evaluated on the held-out outer-test users. We report per-user macro AUC to capture both overall discrimination and cross-user consistency, and we repeat the entire outer cross-validation procedure across multiple random seeds. Final performance is therefore summarized as mean ± standard deviation across all folds, reflecting variability

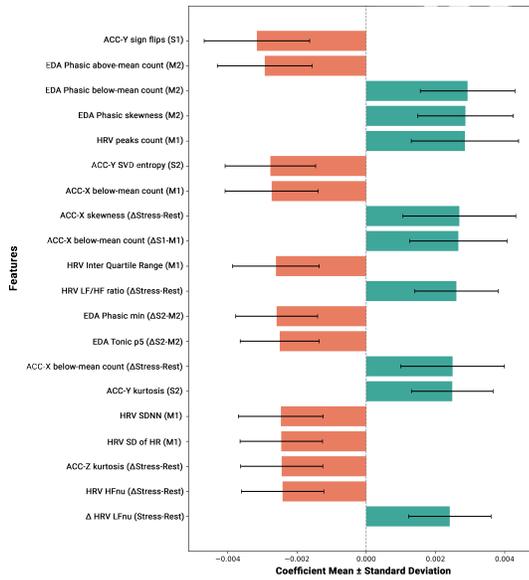


Figure 5: Top 20 standardized logistic-regression coefficients for the best model (ALL sessions, 5s window, two-class). Bars show the cross-fold mean coefficient; error bars indicate \pm SD across 5 participant-grouped folds. Features are ranked by $|\text{mean}|$; colors denote sign (teal: higher probability of an enhancing mindset; coral: higher probability of debilitating mindset). Feature names include session tags (S1/S2/M1/M2) and within-person contrasts (Δ Stress–Rest), (Δ S1–M1), and (Δ S2–M2).

over repeated train–test splits. The results for “mindset as covariate” and “mindset-specific models” across window lengths are presented in Figure 6 and Figure 7 respectively, with detailed modeling and validation details provided in the Appendix H.

7.2 Results

7.2.1 Stress Mindset as a Covariate. We quantified the utility of including a stress is enhancing mindset indicator in Random Forest models used for momentary stress detection, in Figure 6. These models are trained for time series features captured with time-window level feature extraction, as given in Section 4.2. First, even though stress mindset inference performed better with shorter-mid sized time window based feature extraction, we see that momentary stress detection works the best with larger time windows, consistent with prior work [32, 37, 50, 63, 82]. Further, including the mindset feature yields equal or better mean AUCs at most windows, with the clearest benefit at longer windows. This is consistent when using data from all sessions and also when using data from only the first two sessions. However, results must be interpreted with caution as error bars overlap. The average change across windows is only small but positive ($\Delta \approx 0.2\%$ AUC). The largest gain occurs at 20s ($\Delta \approx 2\%$) for ALL sessions; modest improvements appear at 5s ($\Delta \approx 0.5\%$), while effects are negligible or slightly negative at 2s and 30s. Benefits are more consistent for socio-evaluative stress segments, that is when using data only from M1 and S1

sessions. AUC increases at 10-30s with $\Delta \approx 1\%$ (10s), 1.1% (20s), and 1% (30s). Changes are essentially neutral at 2s and slightly negative at 5s. In summary, across both configurations, even though adding a participant’s stress mindset feature seems beneficial many times based on mean AUCs, none of the performance increases are statistically significant and error bars for w/ and w/o mindset groups are overlapping across all windows.

7.2.2 Stress Mindset-Specific Models. As given in Figure 7, across windows, the specialized model trained on the enhancing mindset group achieved the strongest AUCs at short-to-mid windows peaking at 10s with $\text{AUC} = 0.912 \pm 0.065$, and remaining above the one-size-fits-all baseline at 5–20s. This curve is inverse U-shaped, suggesting that there is a sweet spot time window where having stress mindset is beneficial for modeling momentary stress. The pooled “All” model exhibited a steady increase with larger windows and reached $\text{AUC} = 0.825 \pm 0.076$ at 30 s. The debilitating mindset model improved with window length and matched/exceeded the pooled model only at 30 s (0.827 ± 0.095). These trends suggest that (i) mindset-specific stress dynamics exist and can be captured by separate models, and (ii) the enhancing mindset group yields more discriminative short-to-mid temporal patterns than the debilitating mindset group, for which longer aggregation appears beneficial. Hence, specializing models by stress mindset can outperform a single pooled model, particularly for the enhancing group at 5–20 s, supporting our premise that stress mindset is not only a psychological construct but also a useful modeling prior for wearable stress inference. Practically, this motivates either mindset-aware personalization or a gating strategy that selects a model conditioned on an individual’s mindset profile. For a momentary stress detection model aimed at someone with a positive mindset, this could potentially yield a considerable increase in performance, if a mindset-specific model is used.

Another particularly insightful finding is the difference in optimal temporal dynamics between the mindset-specific models. The model for the enhancing group achieved its peak performance ($\text{AUC} = 0.91$) at a 10s window, while the debilitating and all models continued to improve up to a 30s window. This may suggest a potential difference in the nature of the stress response itself: individuals with an enhancing mindset may exhibit a more efficient, phasic response characterized by rapid onset and recovery, which is best captured at a shorter timescale. This finding demonstrates that mindset may not just alter the magnitude of the stress response, but its very temporal structure. This also challenges prior work that suggested momentary stress detection always work better with longer time windows [32, 37, 50, 63, 82], warranting further investigation in the future.

8 Discussion

In this section, we summarize our key results (Section 8.1), discuss their theoretical and practical implications (Section 8.2), and outline the limitations of our study and avenues for future work (Section 8.3).

8.1 Main Findings

Our study provides evidence that integrating the psychological concept of stress mindset into the technical pipeline of momentary

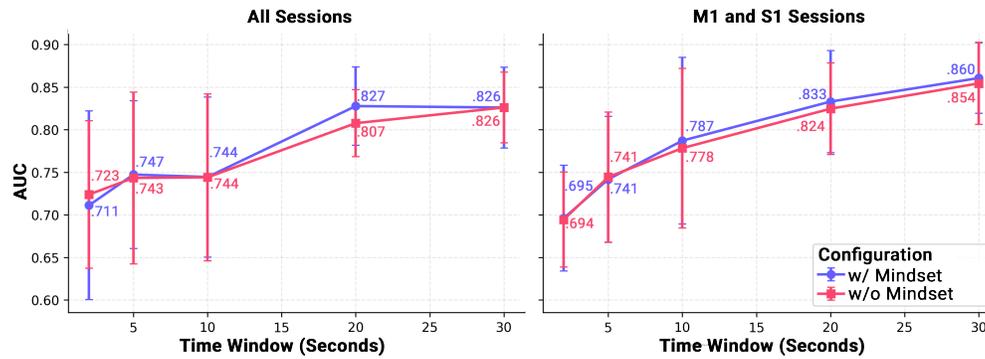


Figure 6: Random-Forest stress classification performance as a function of aggregation window (2–30 s) for a pooled model trained with just sensor features (pink) and stress mindset included as a covariate (purple). Performance given in AUC. Two configurations used— on the left, results when using ALL data is given; on the right, results when using M1+S1 data is given. None of the differences for models with and without mindset at a particular time window were statistically significant.

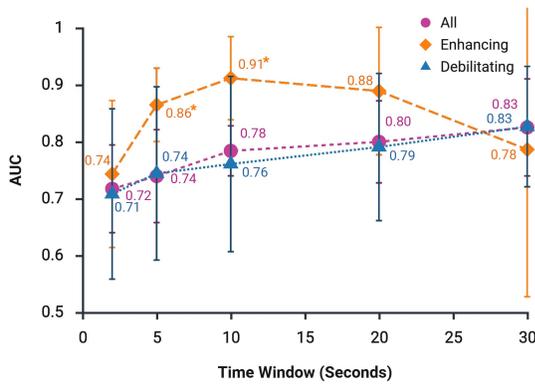


Figure 7: Random-Forest stress detection performance as a function of aggregation window (2–30 s) for a pooled model trained on all users (pink) and mindset-specialized models trained on participants with an enhancing stress mindset (orange) or a debilitating mindset (blue). Curves show mean AUC across 5 held-out folds of a StratifiedGroupKFold (grouped by participant to prevent identity leakage); error bars indicate standard deviation across folds. Statistical significance of the top performing model, in comparison to the second best performing models for the time window, is denoted with an asterisk (* : $p < 0.05$ after FDR correction).

stress sensing is both feasible and beneficial. Given below are the main findings for each research question.

RQ1: Stress mindset has a measurable physiological footprint. Our statistical analysis revealed that stable, between-person differences in stress mindset correspond to differences in physiological features. Specifically, we found that higher typical levels of ‘heart rate variability low-frequency power’ were associated with a debilitating mindset, with statistical significance. While other features

from EDA and ACC showed significance ($p < 0.05$) before multiple testing correction, they were not significant after the correction.

RQ2: Wearable Data Can Infer Stress Mindset. We demonstrated that a user’s stress mindset can be inferred with reasonable performance from their aggregated physiological profile. A regularized logistic regression model achieved an AUC of 0.88 for the binary classification of stress mindsets. Even for a more challenging three-class problem, a support vector classifier reached an AUC of 0.81. Importantly, this performance was achieved using user-level features derived from very short time windows (2–5 seconds), suggesting that the physiological signatures of mindset are present in fine-grained autonomic responses and do not require long periods of observation to detect.

RQ3: Mindset-Aware Models Outperform Generic Stress Detectors. Our results discuss two pathways for leveraging mindset to improve momentary stress detection. First, adding the mindset label as a covariate to a classifier provided only a modest increase in performance, particularly at longer time windows. However, the improvements were not statistically significant, compared to a model that does not have mindset as a feature. Second, and more impactfully, training mindset-specific models yielded a substantial increase in AUC. For instance, the model trained exclusively on participants with an enhancing mindset achieved a mean AUC of 0.91, a statistically significant improvement over the one-size-fits-all baseline model’s mean AUC of approximately 0.78, at 20s. This provides evidence that the physiological patterns of stress are systematically different between the two mindset groups, and that accounting for this heterogeneity can improve model performance for stress detection. We also find that time window required for enhancing and debilitating groups were different, where in a U-shaped curve, performance for enhancing group peaked at 10s, where as for debilitating group, a longer window was required, similar to what is suggested in prior work for generic stress detection [32, 37, 50, 63, 82]. This may suggest a potential difference in the nature of the stress response itself between the enhancing and debilitating mindset groups.

8.2 Implications

8.2.1 Bridging Stress Sensing and Stress Psychology. Our findings bridge a critical gap between the data-driven approach of affective computing and the theoretical richness of stress psychology [13] (Section 2.3). We provide empirical support for the premise that psychological constructs such as mindset have tangible, measurable physiological correlates. By showing that mindset can be inferred from sensor data, we motivate a new class of computational models that are not only statistically personalized, but also grounded in established psychological theory. This shifts individual differences from being treated as statistical "noise" to being treated as meaningful, structured signals. More broadly, our results suggest that the next frontier in personalization is not merely adapting to a user's data, but understanding and adapting to the user's psychology.

8.2.2 Designing the Next Generation of Well-being Technology. Our results show that mindset-specific models outperform a one-size-fits-all approach (AUC 0.91 at 10s vs. 0.83 at 30s), implying that physiological stress signatures differ across mindset groups and that interventions should be tailored accordingly. This aligns with recent work calling for stress interventions that go beyond generic, scheduled prompts and instead provide support in more meaningful ways [97, 98]. Rather than alerts that simply state *'Your stress level is high'*, a mindset-aware system can act as a gated triage, interpreting the same physiological data through a psychological lens and routing users to the support most likely to help them. For users with a stress-is-enhancing mindset, feedback can frame arousal as a resource and offer tools that sustain focus and performance, e.g., *'Your body is in focus mode for the upcoming deadline. This is energy you can use to perform at your best. Let's channel it'*. For users with a stress-is-debilitating mindset, the system can prioritize down-regulation and cognitive re-framing, suggesting calming activities such as brief guided breathing or cognitive behavior therapy-informed prompts, e.g., *'It looks like this project is feeling overwhelming. Let's take a 3-minute break to reset with a guided breathing exercise before you continue'*. This way, inferring stress mindset can shift a device from a passive quantifier to an active and empathetic coach, delivering support that is better matched to the user and moment [60, 73, 91].

Importantly, this mapping is not a rigid rule. The context and intensity of the stressor still matter. Even an individual with a 'stress-is-enhancing' mindset may be overwhelmed by an extreme or unmanageable event (e.g., a family emergency or a major work deadline) and benefit from down-regulation. The value of a mindset-aware system is therefore not to hard-code an intervention, but to personalize its framing. A 'debilitating' user may need messaging oriented toward safety and recovery, whereas an 'enhancing' user, even when down-regulating, may respond better to framing around resource optimization. Future work should therefore study how stable mindset can serve as a lens for tailoring the delivery, tone, and timing of interventions.

8.2.3 Navigating the Regulatory Landscape. Our work has direct implications for the emerging regulatory landscape for Artificial Intelligence such as the European Union's AI Act [1]. As AI systems become embedded in daily life, regulations increasingly focus on the risks of processing sensitive personal data to infer traits about

individuals [16, 23]. A system that infers a psychological trait like stress mindset from passive sensor data would likely be classified as a "high-risk" AI system under the EU AI Act, particularly in sensitive contexts such as workplace wellness or health applications [24]. This classification raises requirements for transparency, robustness, accuracy, and human oversight. Meeting these obligations will require rigorous validation and documentation, explicit disclosure that psychological traits are being inferred from passive data, calibrated performance across populations, and mechanisms for human review and contestability.

8.2.4 The Dual-Use Nature of Psychological Inference. While our discussion has emphasized benefits, the ability to infer a stable psychological trait like stress mindset from sensor data demands a critical perspective. For the minority of users who actively adopt wearables for self-regulation and behavior change [29], this capability may be welcome. However, for the larger population who neither intend nor consent to such uses, the possibility that routine sensor data could be repurposed to infer psychological traits raises concerns about autonomy and informational self-determination. These concerns are amplified by where computation occurs. On-device processing with user control can be reassuring, whereas cloud-based aggregation expands the attack surface and increases the risk of secondary uses [28]. The same pipelines that can empower individuals also enable harmful practices, including algorithmic management, insurance or employment screening, and targeted persuasion, culminating in forms of 'deep' psychological profiling beyond a user's control or expectation³. Our responsibility is therefore not only to advance the technology, but to embed ethical guardrails into its design. This includes enforcing purpose limitation and explicit consent, defaulting to on-device processing, and providing users with transparent, contestable, and revocable inferences, especially in non-volitional contexts like the workplace.

8.3 Limitations and Future Work

Our study provides the first evidence for the feasibility of mindset-aware stress sensing, but its contributions should be understood within the context of its methodological limitations. These limitations, inherent in any foundational study, define the boundaries of our claims and provide a clear road map for future research.

8.3.1 Sample, Statistical Power, and Generalizability. A primary limitation is our sample size (N=23), which becomes more acute after splitting participants into mindset groups (n=13 enhancing, n=10 debilitating). This reduces statistical power and limits the complexity of models we can train. Accordingly, our machine learning results should be interpreted as exploratory tests of feasibility rather than generalizable classifiers, and performance metrics (e.g., AUCs) as promising signals rather than definitive benchmarks. This study was also not intended to produce a production-ready in-the-wild model; instead, it serves as a controlled, foundational investigation into whether wearable data contains any detectable physiological signature of stress mindset. Our sample size is comparable to prior resource-intensive, controlled lab studies in broader

³See, e.g., critiques of under-the-skin surveillance and psychological profiling: <https://www.aljazeera.com/opinions/2020/5/31/surveillance-is-getting-under-our-skin-and-that-should-alarm-us>

human-computer interaction research [20, 34, 63, 95]. However, our participant pool was recruited from university and corporate environments, was predominantly highly educated (74% with a PhD or equivalent), and skewed toward lighter skin tones (61% Asian, 26% White/Caucasian), which constrains external validity. This is important because wrist-worn PPG sensing can degrade for individuals with darker skin tones [52], introducing measurement bias that can propagate through the entire pipeline. While we validated HR measurements against the ECG-based Polar H10 for our cohort, this does not resolve broader generalizability concerns. Prior work has emphasized the need to evaluate generalization, robustness, and reproducibility in sensing-based pipelines [9, 72, 77, 124, 126]. Future work should therefore test these preliminary findings in larger and more diverse samples spanning ages, educational backgrounds, cultural contexts, and skin tones, to support equitable models robust to real-world heterogeneity.

8.3.2 Binary and Ternary Formulations of Stress Mindset. We acknowledge an apparent tension in our work: while we critique the common binarization of stress state (stressed vs. not-stressed), we primarily operationalize stress mindset as a binary construct (enhancing vs. debilitating). This choice was pragmatic rather than theoretical, driven by our sample size ($N=23$). With limited data, directly predicting the continuous SMM score via regression would be underpowered and prone to overfitting. A binary formulation, consistent with core theoretical framing [61], yielded a more tractable and statistically defensible modeling task. That said, stress mindset lies on a continuum, and binarization necessarily discards nuance. It collapses meaningful differences within each category (e.g., SMM scores of 15 and 2) and can be especially problematic for individuals near the midpoint, who may hold a more ambivalent mindset but are forced into one of two classes. To partially address this, we also explored a three-class formulation (low, medium, high) to examine whether a middle group is physiologically distinct or simply ambiguous. Overall, these formulations reflect a trade-off between psychological granularity and statistical validity, mirroring common practices in early-stage mental health and sensor-based modeling research [19, 72, 77, 83, 115, 124]. Future work with larger datasets should move beyond classification, ideally using regression-based models to predict the continuous SMM score.

8.3.3 Ecological Validity and Stressor Type. Our study was conducted in a controlled laboratory setting. This environment allowed us to isolate the effects of specific stressors and collect high-fidelity physiological data with minimal noise from motion artifacts or contextual confounds. However, this internal validity comes at the cost of generalizability, a limitation that extends to our choice of stressors. The acute, performative SSST and the cognitive Stroop test, while effective for controlled stress induction, are both primarily active, performance-based tasks. Even though such tests are commonly used in both research and industry to validate stress detection models [42], we still cannot claim that our findings will generalize to other important types of stressors, such as passive stressors (e.g., viewing emotionally disturbing content) or situational/ambient stressors (e.g., driving in heavy traffic [46]). It is plausible that the physiological signatures of mindset we identified, particularly for a 'stress-is-enhancing' mindset, are most evident during an active challenge and may manifest differently, or not

at all, during a passive stress exposure. This study, focused on a controlled protocol, was necessary to prove that a detectable signal exists at all for stress mindset. The clear and vital next step is to move this research "in-the-wild". Future longitudinal studies should aim to infer and track stress mindset using data collected from wearables over weeks or months of daily life, ideally with larger samples sizes, to account for confounders. This will allow us to investigate whether the physiological signatures we identified hold true amidst the noise of daily activities and, critically, across the more heterogeneous set of stressors that individuals face in their real-world personal and professional challenges.

8.3.4 Measurement and Modeling of Mindset. In our analysis, we treated stress mindset as a stable trait, measured once via a questionnaire. This approach aligns with the foundational psychological theory, which defines mindset as a stable core belief, and follows the standard practice established by the SMM's authors [27]. However, this approach is a limitations because there is evidence that mindset can be shifted with interventions and over time [26, 57].

8.3.5 The Path from Inference to Intervention. Finally, our study is observational. We successfully demonstrated that mindset can be inferred and that mindset-specific models are more accurate. However, even though stress sensing based interventions have been studied before [41, 71, 91], we did not design or test an interactive system that uses this information to provide an intervention. Establishing the feasibility of mindset inference is the necessary prerequisite for designing interventions based on it. Our work provides this crucial foundation. The most exciting avenue for future work lies in closing the loop from inference to intervention. To this end, future research could focus on designing and building mindset-aware systems. A randomized controlled trial could, for instance, compare a generic stress management app to a mindset-aware version that provides tailored feedback and exercises based on the user's inferred mindset. This would allow us to empirically measure the real-world benefit of the personalized approach we have proposed. The public release of our dataset is intended to catalyze such efforts, providing a rich resource for the community to explore more advanced models and begin designing the next generation of psychologically-grounded well-being technologies.

9 Conclusion

This paper demonstrates that an individual's stress mindset, their core belief about whether stress is enhancing or debilitating, is not an abstract psychological concept but a tangible construct with a measurable physiological footprint. We show that this mindset can be inferred with high performance from wearable sensor data and, most importantly, that leveraging this knowledge can lead to momentary stress detection models that are significantly more accurate than their one-size-fits-all counterparts. We also discuss how knowing stress mindset could lead to personalized interventions when under stress. By bridging the technical methodologies of HCI with the theoretical depth of psychology, our work provides a clear path toward a new generation of stress management technologies that are not just personalized to a user's data, but are truly personalized to the user themselves.

References

- [1] EU Artificial Intelligence Act. 2024. The eu artificial intelligence act. *European Union* (2024).
- [2] Maheen M Adamson, Angela Phillips, Srija Seenivasan, Julian Martinez, Harlene Grewal, Xiaojian Kang, John Coetzee, Ines Luttenbacher, Ashley Jester, Odette A Harris, et al. 2020. International prevalence and correlates of psychological stress during the global COVID-19 pandemic. *International journal of environmental research and public health* 17, 24 (2020), 9248.
- [3] Daniel A Adler, Emily Tseng, Khatiya C Moon, John Q Young, John M Kane, Emanuel Moss, David C Mohr, and Tanzeem Choudhury. 2022. Burnout and the quantified workplace: tensions around personal sensing interventions for stress in resident physicians. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–48.
- [4] Daniel A Adler, Vincent W-S Tseng, Gengmo Qi, Joseph Scarpa, Srijan Sen, and Tanzeem Choudhury. 2021. Identifying mobile sensing indicators of stress-resilience. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 5, 2 (2021), 1–32.
- [5] Surely Akiri, Vasundhara Joshi, Gary Williams, Helena M Mentis, and Andrea Kleinsmith. 2025. Navigating Stress Data Sharing: Guardrails for Effective Physiological Data Sharing in High-Stress Training Environments. In *Proceedings of the 4th Annual Symposium on Human-Computer Interaction for Work*. 1–14.
- [6] Mona Alhasani and Rita Orji. 2025. Exploring Trends, Pitfalls, and Future Directions in Digital Behaviour Change Interventions for Managing Student Stress. *International Journal of Human-Computer Interaction* 41, 16 (2025), 10379–10398.
- [7] Wazzan S Aljuhani, Ziad A Aljaafri, Khalid H Alhadlaq, Abdullah M Alanazi, Abdulrahman K Alhadlaq, and Meshal K Alaqeel. 2024. Assessment of stress level and depression among orthopaedic surgeons in Saudi Arabia. *International Orthopaedics* 48, 11 (2024), 2785–2792.
- [8] LA Alker. 2019. *Distress tolerance as a mediator of the relation between stress mindset and anxiety*. B.S. thesis. University of Twente.
- [9] Karim Assi, Lakmal Meegahapola, William Droz, Peter Kun, Amalia De Götzen, Miriam Bidoglia, Sally Stares, George Gaskell, Altangerel Chagnaa, Amarsanaa Ganbold, et al. 2023. Complex daily activities, country-level diversity, and smartphone sensing: A study in denmark, italy, mongolia, paraguay, and uk. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–23.
- [10] Mohamad Awada, Burcin Becerik Gerber, Gale M Lucas, and Shawn C Roll. 2024. Stress appraisal in the workplace and its associations with productivity and mood: Insights from a multimodal machine learning analysis. *Plos one* 19, 1 (2024), e0296468.
- [11] Abeer Badawi, Samira Choudhury, Khalid Elgazzar, and Amer M Burhan. 2023. Artificial intelligence and features investigating to detect neuropsychiatric symptoms in patients with dementia: A pilot study. In *2023 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 741–746.
- [12] Abeer Badawi, Somayya Elmoghazy, Samira Choudhury, Khalid Elgazzar, and Amer Burhan. 2024. Leveraging self-training and variational autoencoder for agitation detection in people with dementia using wearable sensors. *arXiv preprint arXiv:2412.19254* (2024).
- [13] Jakob E Bardram and Aleksandar Matic. 2020. A decade of ubiquitous computing research in mental health. *IEEE Pervasive Computing* 19, 1 (2020), 62–72.
- [14] Felix Beierle and Rüdiger Pryss. 2023. Automating the Development of Stress Detection Systems. In *2023 Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE)*. IEEE, 2694–2696.
- [15] Ananya Bhattacharjee, Pan Chen, Abhijoy Mandal, Anne Hsu, Katie O’Leary, Alex Mariakakis, Joseph Jay Williams, et al. 2024. Exploring user perspectives on brief reflective questioning activities for stress management: Mixed methods study. *JMIR Formative Research* 8, 1 (2024), e47360.
- [16] Edyta Paulina Bogucka, Marios Constantinides, Julia De Miguel Velazquez, Sanja Scepanovic, Daniele Quercia, and Andrés Gvirtz. 2024. The Atlas of AI Incidents in Mobile Computing: Visualizing the Risks and Benefits of AI Gone Mobile. In *Adjunct Proceedings of the 26th International Conference on Mobile Human-Computer Interaction*. 1–6.
- [17] Erika Bondareva, Marios Constantinides, Michael S Eggleston, Ireneusz Jabłoński, Cecilia Mascolo, Zoran Radivojevic, and Sanja Šćepanović. 2022. Stress inference from abdominal sounds using machine learning. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 1985–1988.
- [18] Brandon M Booth, Hana Vrzakova, Stephen M Mattingly, Gonzalo J Martinez, Louis Faust, and Sidney K D’Mello. 2022. Toward robust stress prediction in the age of wearables: Modeling perceived stress in a longitudinal study with information workers. *IEEE Transactions on Affective Computing* 13, 4 (2022), 2201–2217.
- [19] Lindsay Brown, Bernard Grundlehner, and Julien Penders. 2011. Towards wireless emotional valence detection from EEG. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2188–2191.
- [20] Kelly Caine. 2016. Local standards for sample size at CHI. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 981–992.
- [21] Luxi Chen, Li Qu, and Ryan Y Hong. 2022. Pathways linking the big five to psychological distress: exploring the mediating roles of stress mindset and coping flexibility. *Journal of Clinical Medicine* 11, 9 (2022), 2272.
- [22] Sheldon Cohen, Tom Kamarck, and Robin Mermelstein. 1983. A global measure of perceived stress. *Journal of health and social behavior* (1983), 385–396.
- [23] Marios Constantinides, Edyta Paulina Bogucka, Sanja Scepanovic, and Daniele Quercia. 2024. Good intentions, risky inventions: A method for assessing the risks and benefits of AI in mobile and wearable uses. *Proceedings of the ACM on Human-Computer Interaction* 8, MHCI (2024), 1–28.
- [24] Marios Constantinides and Daniele Quercia. 2022. Good intentions, bad inventions: How employees judge pervasive technologies in the workplace. *IEEE Pervasive Computing* 22, 1 (2022), 69–76.
- [25] Dajana Čopec, Matea Karlović Vragolov, and Vesna Buško. 2022. Individual dispositions and situational stressors in competitive sport: The role of stress mindset in the cognitive appraisals processes. *Frontiers in psychology* 13 (2022), 829053.
- [26] Alia J Crum, Modupe Akinola, Ashley Martin, and Sean Fath. 2017. The role of stress mindset in shaping cognitive, emotional, and physiological responses to challenging and threatening stress. *Anxiety, stress, & coping* 30, 4 (2017), 379–395.
- [27] Alia J Crum, Peter Salovey, and Shawn Achor. 2013. Rethinking stress: the role of mindsets in determining the stress response. *Journal of personality and social psychology* 104, 4 (2013), 716.
- [28] Jason Peres da Silva. 2023. Privacy Data Ethics of Wearable Digital Health Technology. <https://digitalhealth.med.brown.edu/news/2023-05-04/ethics-wearables>. Accessed: 2025-09-10.
- [29] Carolina Del-Valle-Soto, Juan Carlos López-Pimentel, Javier Vázquez-Castillo, Juan Arturo Nolazco-Flores, Ramiro Velázquez, José Varela-Aldás, and Paolo Visconti. 2024. A comprehensive review of behavior change techniques in wearables and IoT: implications for health and well-being. *Sensors* 24, 8 (2024), 2429.
- [30] Allison E Diamond and Aaron J Fisher. 2017. Comparative autonomic responses to diagnostic interviewing between individuals with GAD, MDD, SAD and healthy controls. *Frontiers in human neuroscience* 10 (2017), 677.
- [31] Xianghua Ding, Shuhan Wei, Xinning Gui, Ning Gu, and Peng Zhang. 2021. Data engagement reconsidered: a study of automatic stress tracking technology in use. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–13.
- [32] Begum Egilmez, Emirhan Poyraz, Wenting Zhou, Gokhan Memik, Peter Dinda, and Nabil Alshurafa. 2017. UStress: Understanding college student subjective stress using wrist-based passive sensing. In *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, 673–678.
- [33] Stephen A Erath, Mona El-Sheikh, J Benjamin Hinnant, and E Mark Cummings. 2011. Skin conductance level reactivity moderates the association between harsh parenting and growth in child externalizing behavior. *Developmental Psychology* 47, 3 (2011), 693.
- [34] Talia Sofia Ezer, Jonathan Giron, Hadas Erel, and Oren Zuckerman. 2024. So-maesthetic Meditation Wearable: Exploring the Effect of Targeted Warmth Technology on Meditators’ Experiences. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [35] Simon Föll, Martin Maritsch, Federica Spinola, Varun Mishra, Filipe Barata, Tobias Kowatsch, Elgar Fleisch, and Felix Wortmann. 2021. FLIRT: A feature generation toolkit for wearable data. *Computer Methods and Programs in Biomedicine* 212 (2021), 106461.
- [36] Shruti Gedam and Sanchita Paul. 2021. A review on mental stress detection using wearable sensors and machine learning techniques. *IEEE Access* 9 (2021), 84045–84066.
- [37] Martin Gjoreski, Hristijan Gjoreski, Mitja Luštrek, and Matjaž Gams. 2016. Continuous stress detection using a wrist device: in laboratory and real life. In *proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing: Adjunct*. 1185–1193.
- [38] David S Goldstein, Oladi Bentho, Mee-Yeong Park, and Yehonatan Sharabi. 2011. Low-frequency power of heart rate variability is not a measure of cardiac sympathetic tone but may be a measure of modulation of cardiac autonomic outflows by baroreflexes. *Experimental physiology* 96, 12 (2011), 1255–1261.
- [39] Wuwei Gong and Susan A Geertshuis. 2023. Distress and eustress: an analysis of the stress experiences of offshore international students. *Frontiers in Psychology* 14 (2023), 1144767.
- [40] Araceli Gonzalez, Michelle Rozenman, Pauline Goger, and Sarah E Velasco. 2024. Autonomic reactivity during acute social stress: exploratory investigation of an interaction by threat interpretation bias and emotion regulation difficulties. *Anxiety, Stress, & Coping* 37, 2 (2024), 251–264.

- [41] Maria Luisa González Ramírez, Juan Pablo García Vázquez, Marcela D Rodríguez, Luis Alfredo Padilla-López, Gilberto Manuel Galindo-Aldana, and Daniel Cuevas-González. 2023. Wearables for stress management: a scoping review. In *Health-care*, Vol. 11. MDPI, 2369.
- [42] Google Blog. 2023. *How we trained Fitbit's Body Response feature to detect stress*. <https://blog.google/products/fitbit/how-we-trained-fitbits-body-response-feature-to-detect-stress/>. Accessed: 2025-12-02.
- [43] Samuel D Gosling, Peter J Rentfrow, and William B Swann Jr. 2003. A very brief measure of the Big-Five personality domains. *Journal of Research in personality* 37, 6 (2003), 504–528.
- [44] Yunjo Han, Hyemin Lee, Kobiljon E Toshnazarov, Youngtae Noh, and Uichin Lee. 2022. Stressbal: personalized just-in-time stress intervention with wearable and phone sensing. In *Adjunct Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers*. 41–43.
- [45] James A Hanley, Abdissa Negassa, Michael D deB Edwardes, and Janet E Forrester. 2003. Statistical analysis of correlated data using generalized estimating equations: an orientation. *American journal of epidemiology* 157, 4 (2003), 364–375.
- [46] Neska El Haoui, Jean-Michel Poggi, Sylvie Sevestre-Ghalila, Raja Ghozi, and Mériem Jaidane. 2018. AffectiveROAD system and database to assess driver's attention. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*. 800–803.
- [47] Odin Hjemdal, Oddgeir Friberg, Tore C Stiles, Jan H Rosenvinge, and Monica Martinussen. 2006. Resilience predicting psychiatric symptoms: A prospective study of protective factors and their role in adjustment to stressful life events. *Clinical Psychology & Psychotherapy: An International Journal of Theory & Practice* 13, 3 (2006), 194–201.
- [48] Sigrid Hoddø Bakås, Mikolaj P Woźniak, Jo Herstad, Paweł W Woźniak, and Jasmin Niess. 2025. The Dual Model for Everyday Stress Technology: Understanding the Lived Experience of Data-Driven Stress. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–18.
- [49] Melanie S Houle and George E Billman. 1999. Low-frequency component of the heart rate variability spectrum: a poor marker of sympathetic activity. *American Journal of Physiology-Heart and Circulatory Physiology* 276, 1 (1999), H215–H223.
- [50] Karen Hovsepian, Mustafa Al'Absi, Emre Ertin, Thomas Kamarck, Motohiro Nakajima, and Santosh Kumar. 2015. cStress: towards a gold standard for continuous stress assessment in the mobile environment. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*. 493–504.
- [51] Esther Howe, Jina Suh, Mehrab Bin Morshed, Daniel McDuff, Kael Rowan, Javier Hernandez, Marah Ihab Abidin, Gonzalo Ramos, Tracy Tran, and Mary P Czerwinski. 2022. Design of digital workplace stress-reduction intervention systems: Effects of intervention type and timing. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–16.
- [52] Stanley H. Hung, Kelsey Serwa, Gillian Rosenthal, and Janice J. Eng. 2025. Validity of heart rate measurements in wrist-based monitors across skin tones during exercise. *PLoS One* 20, 2 (2025), e0318724. <https://doi.org/10.1371/journal.pone.0318724>
- [53] Stephanie L Hyland, Martin Faltys, Matthias Hüser, Xinrui Lyu, Thomas Gumbsch, Cristóbal Esteban, Christian Bock, Max Horn, Michael Moor, Bastian Rieck, et al. 2020. Early prediction of circulatory failure in the intensive care unit using machine learning. *Nature medicine* 26, 3 (2020), 364–373.
- [54] Md Saiful Islam, Md Mahbubur Rahman, Mehrab Bin Morshed, David J Lin, Yunzhi Li, Hao Zhou, Wendy Berry Mendes, and Jilong Kuang. 2025. Ballistobud: Heart Rate Variability Monitoring using Earbud Accelerometry for Stress Assessment. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–21.
- [55] Tanvir Islam and Peter Washington. 2023. Individualized stress mobile sensing using self-supervised pre-training. *Applied Sciences* 13, 21 (2023), 12035.
- [56] Oussama Jlassi. 2023. *Predicting viral respiratory tract infections using wearable garment biosensors*. Ph.D. Dissertation. Université de Montréal.
- [57] Konstantinos Karampas, Christos Pezirkianidis, and Anastassios Stalikas. 2022. ReStress mindset: An internet-delivered intervention that changes university students' mindset about stress in the shadow of the COVID-19 pandemic. *Frontiers in Psychology* 13 (2022), 1036564.
- [58] Jacob J Keech and Kyra Hamilton. 2019. Stress mindset. In *Encyclopedia of behavioral medicine*. Springer, 1–4.
- [59] Elizabeth Keohan. 2023. *Distress vs. Eustress: Positive & Negative Stress Explained*. Retrieved September 6, 2025 from <https://www.talkspace.com/blog/distress-vs-eustress/>
- [60] Justin Khasentino, Anastasiya Belyaeva, Xin Liu, Zhun Yang, Nicholas A Furlotte, Chace Lee, Erik Schenck, Yojan Patel, Jian Cui, Logan Douglas Schneider, et al. 2025. A personal health large language model for sleep and fitness coaching. *Nature Medicine* (2025), 1–10.
- [61] Christopher J Kilby and Kerry A Sherman. 2016. Delineating the relationship between stress mindset and primary appraisals: preliminary findings. *Springerplus* 5, 1 (2016), 336.
- [62] Taewan Kim, Haesoo Kim, Ha Yeon Lee, Hwarang Goh, Shakhboz Abdjapporov, Mingon Jeong, Hyunsung Cho, Kyungsik Han, Youngtae Noh, Sung-Ju Lee, et al. 2022. Prediction for retrospection: Integrating algorithmic stress prediction into personal informatics systems for college students' mental health. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–20.
- [63] Zachary D King, Judith Moskowitz, Begum Egilmez, Shibo Zhang, Lida Zhang, Michael Bass, John Rogers, Roozbeh Ghaffari, Laurie Wakschlag, and Nabil Alshurafa. 2019. Micro-stress EMA: A passive sensing framework for detecting in-the-wild stress in pregnant mothers. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 3, 3 (2019), 1–22.
- [64] Clemens Kirschbaum, Karl-Martin Pirke, and Dirk H Hellhammer. 1993. The 'Trier Social Stress Test'—a tool for investigating psychological stress responses in a laboratory setting. *Neuropsychobiology* 28, 1-2 (1993), 76–81.
- [65] Rafal Kocielnik and Natalia Sidorova. 2015. Personalized stress management: enabling stress monitoring with lifelogexplorer. *KI-Künstliche Intelligenz* 29, 2 (2015), 115–122.
- [66] Shakar Kutsal, Lauri Juhani Tulkki, Tomi Sarkanen, Petra Redfors, Katarina Jood, Annika Nordanstig, Nilüfer Yeşilot, Mine Sezgin, Pauli Ylikotila, Marialuisa Zedde, et al. 2025. Association between self-perceived stress and cryptogenic ischemic stroke in young adults: a case-control study. *Neurology* 104, 6 (2025), e213369.
- [67] Liang Ky. 1986. Longitudinal data analysis using generalized linear models. *Biometrika* 73 (1986), 13–22.
- [68] Feng Lin, Kathi Heffner, Mark Mapstone, Ding-Geng Din Chen, and Anton Porsteisson. 2014. Frequency of mentally stimulating activities modifies the relationship between cardiovascular reactivity and executive function in old age. *The American Journal of Geriatric Psychiatry* 22, 11 (2014), 1210–1221.
- [69] Yunfei Luo, Iman Deznabi, Abhinav Shaw, Natcha Simsiri, Tauhidur Rahman, and Madalina Fiterau. 2024. Dynamic clustering via branched deep learning enhances personalization of stress prediction from mobile sensor data. *Scientific Reports* 14, 1 (2024), 6631.
- [70] Xinrui Lyu, Bowen Fan, Matthias Hüser, Philip Hartout, Thomas Gumbsch, Martin Faltys, Tobias M Merz, Gunnar Rättsch, and Karsten Borgwardt. 2024. An empirical study on KDIGO-defined acute kidney injury prediction in the intensive care unit. *Bioinformatics* 40, Supplement_1 (2024), i247–i256.
- [71] Diana MacLean, Asta Roseway, and Mary Czerwinski. 2013. MoodWings: a wearable biofeedback device for real-time stress intervention. In *Proceedings of the 6th international conference on Pervasive Technologies Related to Assistive Environments*. 1–8.
- [72] Aurel Ruben Mäder, Lakmal Meegahapola, and Daniel Gatica-Perez. 2024. Learning about social context from smartphone data: Generalization across countries and daily Life moments. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–18.
- [73] Matteo Malgaroli and Daniel McDuff. 2024. An overview of diagnostics and therapeutics using large language models. *Journal of Traumatic Stress* 37, 5 (2024), 754–760.
- [74] Paul C Mansell and Martin J Turner. 2023. The mediating role of proactive coping in the relationships between stress mindset, challenge appraisal tendencies, and psychological wellbeing. *Frontiers in Psychology* 14 (2023), 1140790.
- [75] Stacy Marsella, Jonathan Gratch, Paolo Petta, et al. 2010. Computational models of emotion. *A Blueprint for Affective Computing—A sourcebook and manual* 11, 1 (2010), 21–46.
- [76] Daniel J McDuff, Javier Hernandez, Sarah Gontarek, and Rosalind W Picard. 2016. Cogcam: Contact-free measurement of cognitive stress during computer tasks with a digital camera. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 4000–4004.
- [77] Lakmal Meegahapola, William Droz, Peter Kun, Amalia De Götzen, Chaitanya Nutakki, Shyam Diwaker, Salvador Ruiz Correa, Donglei Song, Hao Xu, Miriam Bidoglia, et al. 2023. Generalization and personalization of mobile sensing-based mood inference models: an analysis of college students in eight countries. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 6, 4 (2023), 1–32.
- [78] Lakmal Meegahapola and Daniel Gatica-Perez. 2020. Smartphone sensing for the well-being of young adults: A review. *IEEE Access* 9 (2020), 3374–3399.
- [79] Lakmal Meegahapola, Hamza Hassoune, and Daniel Gatica-Perez. 2024. M3BAT: Unsupervised domain adaptation for multimodal mobile sensing with multi-branch adversarial training. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 2 (2024), 1–30.
- [80] Lakmal Meegahapola, Salvador Ruiz-Correa, Viridiana del Carmen Robledo-Valero, Emilio Ernesto Hernandez-Huerfano, Leonardo Alvarez-Rivera, Ronald Chenu-Abente, and Daniel Gatica-Perez. 2021. One more bite? Inferring food consumption level of college students using smartphone sensing and self-reports. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (2021), 1–28.
- [81] Dorota Mierzejewska-Floreani, Mateusz Banaszekiewicz, and Ewa Gruszczynska. 2022. Psychometric properties of the Stress Mindset Measure (SMM) in the Polish population. *Plos one* 17, 3 (2022), e0264853.

- [82] Varun Mishra, Tian Hao, Si Sun, Kimberly N Walter, Marion J Ball, Ching-Hua Chen, and Xinxin Zhu. 2018. Investigating the role of context in perceived stress detection in the wild. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*. 1708–1716.
- [83] Varun Mishra, Sougata Sen, Grace Chen, Tian Hao, Jeffrey Rogers, Ching-Hua Chen, and David Kotz. 2020. Evaluating the reproducibility of physiological stress detection models. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 4, 4 (2020), 1–29.
- [84] Mehrab Bin Morshed, Javier Hernandez, Daniel McDuff, Jina Suh, Esther Howe, Kael Rowan, Marah Abdin, Gonzalo Ramos, Tracy Tran, and Mary Czerwinski. 2022. Advancing the understanding and measurement of workplace stress in remote information workers from passive sensors and behavioral data. In *2022 10th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 1–8.
- [85] Martin Karl Moser, Maximilian Ehrhart, and Bernd Resch. 2024. An explainable deep learning approach for stress detection in wearable sensor measurements. *Sensors* 24, 16 (2024), 5085.
- [86] Luis Moya-Albiol, Alicia Salvador, Raquel Costa, Sonia Martínez-Sanchis, and Esperanza González-Bono. 2003. Psychophysiological responses to acute stress in two groups of healthy women differing in fitness. *Psicothema* (2003), 563–568.
- [87] Md Sakibul Hasan Nahid, Tahrim Zaman Tila, and Turuna S Seecharan. 2024. Detecting emotional arousal and aggressive driving using neural networks: A pilot study involving young drivers in Duluth. *Sensors* 24, 22 (2024), 7109.
- [88] Sameer Neupane, Mithun Saha, Nasir Ali, Timothy Hnat, Shahin Alan Samiei, Anandathirtha Nandugudi, David M Almeida, and Santosh Kumar. 2024. Momentary stressor logging and reflective visualizations: Implications for stress management with wearables. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–19.
- [89] Matthias Norden, Amin Gerard Hofmann, Martin Meier, Felix Balzer, Oliver T Wolf, Erwin Böttinger, and Hanna Drimalla. 2022. Inducing and recording acute stress responses on a large scale with the digital stress test (DST): development and evaluation study. *Journal of medical internet research* 24, 7 (2022), e32280.
- [90] Aleksandr Ometov, Anzhelika Mezina, Hsiao-Chun Lin, Otso Arponen, Radim Burget, and Jari Nurmi. 2025. Stress and Emotion Open Access Data: A Review on Datasets, Modalities, Methods, Challenges, and Future Research Perspectives. *Journal of Healthcare Informatics Research* (2025), 1–33.
- [91] Joseph Orji, Gerry Chan, and Rita Orji. 2025. Revitalizing Wellbeing: App Design for Stress Reduction through Artificial Intelligence and Persuasive Technology. *International Journal of Human-Computer Studies* (2025), 103600.
- [92] JunSeo Park, Yechan Yang, Gerard Jeoungyun Kim, and Hanseob Kim. 2025. The Impact of Observer Presence on Trainees' Mental States and Performance in Remote Military Training with Virtual Humans in Mixed Reality Environment. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–19.
- [93] Evan M Peck, Emily Carlin, and Robert Jacob. 2015. Designing brain-computer interfaces for attention-aware systems. *Computer* 48, 10 (2015), 34–42.
- [94] Anuja Pinge, Vinaya Gad, Dheriyta Jaisighani, Surjya Ghosh, and Sougata Sen. 2024. Detection and monitoring of stress using wearables: a systematic review. *Frontiers in Computer Science* 6 (2024), 1478851.
- [95] Dirk Queck, Iannis Albert, Nicole Burkard, Philipp Zimmer, Georg Volkmar, Bastian Dänekas, Rainer Malaka, and Marc Herrlich. 2022. Spiderclip: towards an open source system for wearable device simulation in virtual reality. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–7.
- [96] Houda Rafi, Yannick Benezeth, Philippe Reynaud, Emmanuel Arnoux, Fan Yang Song, and Cedric Demonceaux. 2022. Personalization of AI models based on federated learning for driver stress monitoring. In *European Conference on Computer Vision*. Springer, 575–585.
- [97] Pranav Rao, Maryam Taj, Alex Mariakakis, Joseph Jay Williams, and Ananya Bhattacharjee. 2025. Fitting the Message to the Moment: Designing Calendar-Aware Stress Messaging with Large Language Models. *arXiv preprint arXiv:2505.23997* (2025).
- [98] Pranav Rao, Sarah Yi Xu, Ananya Bhattacharjee, Yuchen Zeng, Alex Mariakakis, and Joseph Jay Williams. 2024. Integrating digital calendars with large language models for stress management interventions. (2024).
- [99] Patrice Renaud and Jean-Pierre Blondin. 1997. The stress of Stroop performance: Physiological and emotional responses to color–word interference, task pacing, and pacing speed. *International Journal of Psychophysiology* 27, 2 (1997), 87–97.
- [100] Vincenzo Ronca, Ana C Martinez-Levy, Alessia Vozzi, Andrea Giorgi, Pietro Aricò, Rossella Capotorto, Gianluca Borghini, Fabio Babiloni, and Gianluca Di Flumeri. 2023. Wearable technologies for electrodermal and cardiac activity measurements: a comparison between fitbit sense, empatica E4 and shimmer GSR3+. *Sensors* 23, 13 (2023), 5847.
- [101] Virginia Sandulescu, Sally Andrews, David Ellis, Nicola Bellotto, and Oscar Martinez Mozos. 2015. Stress detection using wearable physiological sensors. In *International work-conference on the interplay between natural and artificial computation*. Springer, 526–532.
- [102] Federica Scarpina and Sofia Tagini. 2017. The stroop color and word test. *Frontiers in psychology* 8 (2017), 557.
- [103] Sanja Šćepanović, Marios Constantinides, Daniele Quercia, and Seunghyun Kim. 2023. Quantifying the impact of positive stress on companies from online employee reviews. *Scientific Reports* 13, 1 (2023), 1603.
- [104] Mhd Saeed Sharif, Madhav Raj Theeng Tamang, Cynthia Fu, Ahmed Ibrahim Alzahrani, and Fahad Alblehai. 2025. Innovative computation to detect stress in working people based on mode of commute. *Journal of Transport & Health* 41 (2025), 101979.
- [105] Abhinav Shaw, Natcha Simsiri, Iman Deznaby, Madalina Fiterau, and Tauhidur Rahman. 2019. Personalized student stress prediction with deep multitask network. *arXiv preprint arXiv:1906.11356* (2019).
- [106] Pragma Singh, Ankush Gupta, Mohan Kumar, and Pushpendra Singh. 2025. AnnoSense: A Framework for Physiological Emotion Data Collection in Everyday Settings for AI. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 9, 3 (2025), 1–47.
- [107] Žiga Stržinar, Araceli Sanchis, Agapito Ledezma, Oscar Sipele, Boštjan Pregelj, and Igor Škrjanc. 2023. Stress detection using frequency spectrum analysis of wrist-measured electrodermal activity. *Sensors* 23, 2 (2023), 963.
- [108] Yi-Yuan Tang, Britta K Hölzel, and Michael I Posner. 2015. The neuroscience of mindfulness meditation. *Nature reviews neuroscience* 16, 4 (2015), 213–225.
- [109] Tahrim Zaman Tila. 2023. *Exploring Relationship Between Electrodermal Activity and Driving Behavior*. Master's thesis. University of Minnesota.
- [110] Tahrim Zaman Tila and Turuna S Seecharan. 2022. Using wearable sensors to form a relationship between driver stress and aggressive driving habits. In *World Congress on Engineering Asset Management*. Springer, 333–342.
- [111] Merel KN van den Berg, Armağan Karahanoğlu, Matthijs L Noordzij, Els LM Maeckelberghe, and Geke DS Ludden. 2025. Facilitators and Barriers of Wearable Stress Management Technology: A Narrative Review of User Perspectives. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–24.
- [112] DJ Van der Mee, Q Duivestein, MJ Gevonden, JHDM Westerink, and EJC de Geus. 2020. The short Sing-a-Song Stress Test: A practical and valid test of autonomic responses induced by social-evaluative stress. *Autonomic Neuroscience* 224 (2020), 102612.
- [113] Gideon Vos, Maryam Ebrahimipour, Liza van Eijk, Zoltan Sarnyai, and Mostafa Rahimi Azghadi. 2025. Stress monitoring using low-cost electroencephalogram devices: A systematic literature review. *International Journal of Medical Informatics* (2025), 105859.
- [114] Gideon Vos, Kelly Trinh, Zoltan Sarnyai, and Mostafa Rahimi Azghadi. 2023. Generalizable machine learning for stress monitoring from wearable devices: A systematic literature review. *International Journal of Medical Informatics* 173 (2023), 105026.
- [115] Rafael Wampfler, Severin Klingler, Barbara Solenthaler, Victor R Schinazi, Markus Gross, and Christian Holz. 2022. Affective state prediction from smartphone touch and sensor data in the wild. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–14.
- [116] Yuna Watanabe, Natasha Yamane, Aarti Sathyanarayana, Varun Mishra, and Matthew S Goodwin. 2025. Beyond Motion Artifacts: Optimizing PPG Pre-processing for Accurate Pulse Rate Variability Estimation. *arXiv preprint arXiv:2510.06158* (2025).
- [117] Xiaoyan Wei, Zebang Zhang, Zijian Yue, and Hsiang-Ting Chen. 2025. Context-AI Tunes: Context-Aware AI-Generated Music for Stress Reduction. In *International Conference on Human-Computer Interaction*. Springer, 330–345.
- [118] Mark A Wetherell, Olivia Craw, and Michael A Smith. 2018. Techniques for inducing stress. *The Routledge international handbook of psychobiology* (2018), 47.
- [119] Brent D Winslow, George L Chadderdon, Sara J Dechmerowski, David L Jones, Solomon Kalkstein, Jennifer L Greene, and Philip Gehrman. 2016. Development and clinical evaluation of an mHealth application for stress management. *Frontiers in psychiatry* 7 (2016), 130.
- [120] World Health Organization. 2023. Stress. <https://www.who.int/news-room/questions-and-answers/item/stress>. <https://www.who.int/news-room/questions-and-answers/item/stress> Accessed: 2025-09-08.
- [121] Lingxi Wu, Arlene John, and Jorge Piano Simoes. 2025. Predicting Positive Psychological States using Machine Learning and Digital Biomarkers from Everyday Wearable Data. *medRxiv* (2025), 2025–04.
- [122] Yi Xiao, Harshit Sharma, Sawinder Kaur, Dessa Bergen-Cico, and Asif Salekin. 2025. Human Heterogeneity Invariant Stress Sensing. *arXiv preprint arXiv:2506.02256* (2025).
- [123] Yi Xiao, Harshit Sharma, Zhongyang Zhang, Dessa Bergen-Cico, Tauhidur Rahman, and Asif Salekin. 2024. Reading between the heat: Co-teaching body thermal signatures for non-intrusive stress detection. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 4 (2024), 1–30.
- [124] Xuhai Xu, Xin Liu, Han Zhang, Weichen Wang, Subigya Nepal, Yasaman Sefidgar, Woosuk Seo, Kevin S Kuehn, Jeremy F Huckins, Margaret E Morris, et al. 2023. Globem: Cross-dataset generalization of longitudinal human behavior

- modeling. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 4 (2023), 1–34.
- [125] Camellia Zakaria, Rajesh Balan, and Youngki Lee. 2019. StressMon: scalable detection of perceived stress and depression using passive sensing of changes in work Routines and group interactions. *Proceedings of the ACM on human-computer interaction* 3, CSCW (2019), 1–29.
- [126] Panyu Zhang, Gyuwon Jung, Jumabek Alikhanov, Uzair Ahmed, and Uichin Lee. 2024. A reproducible stress prediction pipeline with mobile sensor data. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 8, 3 (2024), 1–35.
- [127] Howe Yuan Zhu, Hsiang-Ting Chen, and Chin-Teng Lin. 2021. The effects of virtual and physical elevation on physiological stress during virtual reality height exposure. *IEEE Transactions on Visualization and Computer Graphics* 29, 4 (2021), 1937–1950.

A Participant-Level Demographics

Table 6: Participant-level demographics.

ID	Gender	Age	Ethnicity	Education	Stress Mindset
02	Male	31	White, Caucasian	PhD or Equivalent	Low (<16)
03	Female	27	Asian or Asian British	PhD or Equivalent	Low (<16)
04	Female	35	Asian or Asian British	PhD or Equivalent	Low (<16)
05	Male	52	Mixed or Multiple Ethnic Groups	PhD or Equivalent	Low (<16)
06	Male	42	White, Caucasian	Masters or Equivalent	Low (<16)
07	Male	23	White, Caucasian	High School	Low (<16)
08	Female	29	Asian or Asian British	PhD or Equivalent	High (≥ 16)
09	Female	30	Asian or Asian British	PhD or Equivalent	High (≥ 16)
10	Female	30	White, Caucasian	PhD or Equivalent	High (≥ 16)
11	Male	29	Asian or Asian British	Bachelors or Equivalent	High (≥ 16)
12	Male	34	Asian or Asian British	PhD or Equivalent	High (≥ 16)
13	Male	27	White, Caucasian	Masters or Equivalent	High (≥ 16)
14	Male	30	Not specified	Bachelors or Equivalent	High (≥ 16)
15	Male	37	White, Caucasian	PhD or Equivalent	High (≥ 16)
16	Male	37	Asian or Asian British	PhD or Equivalent	High (≥ 16)
17	Female	36	Asian or Asian British	PhD or Equivalent	High (≥ 16)
18	Female	28	Asian or Asian British	PhD or Equivalent	Low (<16)
20	Female	31	Asian or Asian British	PhD or Equivalent	High (≥ 16)
21	Female	35	Asian or Asian British	PhD or Equivalent	High (≥ 16)
22	Male	29	Asian or Asian British	Masters or Equivalent	Low (<16)
23	Male	33	Asian or Asian British	PhD or Equivalent	Low (<16)
24	Male	30	Asian or Asian British	PhD or Equivalent	High (≥ 16)
25	Male	38	Another Ethnic Group	PhD or Equivalent	Low (<16)

B Electrodermal Activity Features

Table 7: Decoding FLIRT feature names: Electrodermal activity (EDA). Features follow <component>_<statistic>[__ <condition or contrast>] and are computed on sliding windows.

EDA component	Pattern	What it means	Example from this dataset
Tonic (SCL)	eda_tonic_*	Statistics on baseline skin conductance (slow-varying level after decomposition).	eda_tonic_mean, eda_tonic_pct_5, eda_tonic_rms
Phasic (SCR)	eda_phasic_*	Statistics on the phasic/SCR component (event-related responses) after decomposition.	eda_phasic_min, eda_phasic_pct_5, eda_phasic_skewness
Examples of statistic tokens seen in EDA rows			
Mean / Min / Max	mean, min, max	Window summaries.	eda_tonic_mean, eda_phasic_min
Spread	std, iqr, iqr_5_95	Standard deviation; IQR; 95–5 percentile range.	eda_tonic_iqr, eda_tonic_iqr_5_95
Percentiles	pct_5, pct_95	5th and 95th percentile values.	eda_phasic_pct_5
Shape	skewness, kurtosis	Distribution asymmetry and tailedness.	eda_phasic_skewness
Counts	n_above_mean, n_below_mean	Sample counts above/below the window mean.	eda_phasic_n_above_mean
Energy / RMS	energy, rms	Sum of squares; root-mean-square amplitude.	eda_tonic_rms
Complexity	entropy variants	Entropy-based complexity measures.	eda_tonic_entropy

Note: Condition tags such as __M1, __M2, __S1, __S2 denote the segment/session used. Contrasts use __DELTA_A_minus_B and are computed as A–B on the base feature.

C Heart Rate Features

Table 8: Decoding FLIRT feature names: Heart and HRV (from IBI/NN intervals). Features follow <component>_<statistic>[_<condition or contrast>].

HR/HRV group	Pattern	What it means	Example from this dataset
Time-domain	hrv_sdn, hrv_rmssd, ...	Variation of beat-to-beat intervals in time.	hrv_sdn, hrv_std
Frequency-domain	hrv_lf, hrv_hf, hrv_lf_hf_ratio, hrv_lfnu	Spectral power in LF (0.04–0.15 Hz), HF (0.15–0.40 Hz); ratio and normalized units.	hrv_lf_hf_ratio, hrv_lfnu, hrv_hfnu
Statistical on IBI	hrv_ptp, hrv_iqr, hrv_peaks, ...	Generic statistics over the IBI series (peak-to-peak range, IQR, peak count).	hrv_ptp, hrv_iqr, hrv_peaks
Examples of statistic tokens seen in HR/HRV rows			
Mean / Min / Max	mean, min, max	Window summaries.	hrv_mean
Spread	std, iqr	Standard deviation; interquartile range.	hrv_iqr
Amplitude range	ptp	Peak-to-peak range = max–min of IBI.	hrv_ptp
Counts	peaks	Local-maximum count in the series.	hrv_peaks
Path length	lineintegral	Sum of absolute successive differences.	hrv_lineintegral
Complexity	entropy variants	Entropy-based complexity measures.	hrv_perm_entropy

Note: Normalized units (LFnu/HFnu) scale LF/HF by total power minus VLF. Condition tags and contrasts follow the same scheme as this table's EDA/ACC counterparts.

D Accelerometer Features

Table 9: Decoding FLIRT feature names: Accelerometry (ACC). Features follow <component>_<statistic>[_<condition or contrast>].

ACC component	Pattern	What it means	Example from this dataset
Per-axis	acc_acc_[x y z]*	Statistics on raw acceleration per axis.	acc_acc_x_iqr_5_95, acc_acc_x_std, acc_acc_x_n_sign_changes
Vector magnitude	acc_l2_*	Statistics on the ℓ_2 magnitude (vector norm).	acc_l2_iqr_5_95
Examples of statistic tokens seen in ACC rows			
Spread	std, iqr, iqr_5_95	Standard deviation; IQR; 95–5 percentile range.	acc_acc_x_iqr, acc_acc_x_iqr_5_95
Percentiles	pct_5, pct_95	5th and 95th percentile values.	acc_acc_x_pct_95
Shape	skewness, kurtosis	Distribution asymmetry and tailed ness.	acc_acc_y_kurtosis
Counts / dynamics	peaks, n_above_mean, n_below_mean, n_sign_changes	Local-maximum count; samples above/below mean; slope sign flips (oscillations).	acc_acc_x_n_below_mean, acc_acc_x_n_sign_changes
Energy / RMS	energy, rms	Sum of squares; root-mean-square amplitude.	acc_acc_x_rms
Complexity	entropy variants	Entropy-based complexity measures, e.g., SVD entropy.	acc_acc_y_svd_entropy

Note: Condition tags (__M1, __M2, __S1, __S2) and contrasts (__DELTA_A_minus_B) follow the same scheme as the EDA/HRV tables.

E Detailed Methodology Description of Statistical Analysis for Answering RQ1

To formally test whether physiological features derived from heart activity, electrodermal activity, and accelerometry were associated with a binary stress mindset label (*Enhancing* vs. *Debilitating*), we analyzed each session’s data using a marginal logistic model. We utilized Generalized Estimating Equations (GEE) [67] with a logit link and an exchangeable working correlation [45]. This framework allows for the estimation of average effects across the population while explicitly accounting for within-participant dependence across repeated time windows. To properly distinguish stable trait effects from momentary fluctuations, we employed a standard within/between decomposition. Let x_{it} denote the feature for participant i at window t . We first standardized each continuous feature globally (z-scored). We then decomposed the z-scored feature into a participant mean \bar{x}_i (representing the “between” effect) and a deviation $x_{it} - \bar{x}_i$ (representing the “within” effect). The model formulation included both terms simultaneously:

$$\text{logit}(P(Y_i = 1)) = \beta_0 + \beta_B \bar{x}_i + \beta_W (x_{it} - \bar{x}_i)$$

In this context, the between-person coefficient (β_B) estimates the association between a participant’s typical feature level and their mindset, while the within-person coefficient (β_W) captures short-term deviations. Because stress mindset is treated as a time-invariant trait within the scope of this study, our primary inferential focus is on the between-person odds ratios derived from β_B . Parameter uncertainty was quantified using robust (sandwich) standard errors. We report statistical significance based on Wald z statistics and two-sided p -values. To address the multiple comparisons problem inherent in screening many sensor features, we controlled multiplicity using the Benjamini–Hochberg False Discovery Rate (FDR) adjustment. All reported Odds Ratios (OR) correspond to a one-standard-deviation increase in the respective feature.

F Detailed Methodology Description for Answering RQ2

Having established that a physiological signal of mindset exists, our next goal was to determine if this signal was strong and consistent enough to build a machine learning model capable of inferring a user’s mindset directly from their sensor data. This required us to aggregate window-level features into a stable user-level profile (Section F.2) and test a suite of classifiers in a robust evaluation protocol (Section F.5).

F.1 Participants, Sessions, and Signals

We analyze the data across four task blocks: two stress-elicitation blocks (S1, S2) and two mindset/neutral blocks (M1, M2). We also use an ALL representation that concatenates features from all four blocks to capture both baseline and reactivity in a single vector. Downstream feature names follow common conventions for HR (e.g., `hrv_1f_hf_ratio`, `hrv_1fnu`, `hrv_hfnu`, entropy-based non-linearity measures) and EDA (e.g., tonic/phasic statistics, phasic entropy). For more details about the features generated using different modalities, refer to Appendix (Table 7, Table 8, and Table 9).

F.2 Windowing and Feature Construction

To capture short-lived autonomic responses while maintaining stable estimates, we extract features in fixed-length, partially overlapping windows of {2, 5, 10, 20, 30} s. Step sizes of {1, 2, 5, 5, 10} s, respectively, keep adjacent windows informative without excessive redundancy. The detailed user-level feature extraction process was described in Section 4.3, and summarized in Figure 1C. Within each window we compute standard time-, frequency-, and nonlinearity-based descriptors for sensor data. We then form user-level summaries at the session level by aggregating the window distributions with robust statistics. This yields one high-dimensional feature vector per participant and session (e.g., ~1200 features for ALL), suitable for user-level modeling while preserving the multi-session structure. As described in Section 4.3, we also include difference (contrast) features such as S1–M1, S2–M2, and a generic Stress–Rest. These encode within-subject deltas that can reduce sensitivity to individual baselines and sensor idiosyncrasies.

F.3 Label Construction: Binary and Ternary Formulations

The SMM produces a 0–32 total score. We convert this continuous score into categorical labels using two schemes: (i) Two-class. A threshold at 16 splits participants into “low” vs. “high” mindset. This matches the canonical stress-is-debilitating vs. stress-is-enhancing framing; and (ii) Three-class. Quantile-based terciles (lower, middle, upper) enable a three-way classification. While stress mindset is commonly interpreted with the binary setup, we explored the three-class setup for comprehensive, because prior work in mental health/well-being have explored two and three class setups for different constructs [77]. This also helps to test whether sensors separate the extremes and whether the middle group is intrinsically ambiguous when later binarized. Labels are assigned at the user level and are therefore constant across sessions for a given participant.

F.4 Pre-processing and Leakage Control

Due to the high dimensionality of the feature vector, we adopt a lean, leakage-safe pipeline executed strictly inside each training fold: Column filtering: drop features with no observed values across the training participants (the same mask is applied to the test fold); Imputation: feature-wise median imputation on the training fold and then apply the learned imputers to the test fold; Standardization: z-score using training-fold mean and variance and apply to train/test. Standardization is essential for margin-based models (e.g., LR, SVC). All steps use `scikit-learn` and are confined to the training data within each fold to prevent information leakage.

F.5 Models and Model Selection Rationale

We compare four widely used classifiers that span linear and non-linear decision rules while remaining appropriate for small samples: Logistic Regression (LR): Interpretable linear baseline that typically performs well in low- N regimes with regularization; Support Vector Classifier (SVC): Captures smooth nonlinear boundaries; probability calibration enables ROC-based metrics; Random Forest (RF): Non-parametric, interaction-capable trees that do not require feature

scaling; Extreme Gradient Boosting (XGB): High-capacity gradient-boosted trees competitive on tabular data. Given dataset size, we intentionally keep hyperparameter tuning light to curb overfitting and preserve out-of-sample validity. Our goal is to test whether separability is mainly linear, whether nonlinear capacity helps, and how window length and session context affect performance.

F.6 Evaluation Protocol and Metrics

We use grouped 5-fold cross-validation with participant ID as the grouping key so that all data from a participant appear exclusively in either train or test within a fold (no subject leakage). For each (session, window, model, labeling) configuration, we train on $K-1$ folds and evaluate on the held-out fold. For two-class classification we report macro AUC; for three-class classification we report macro one-vs-rest AUC. Both metrics are threshold-free and stable in small samples. Results are summarized as mean \pm SD across the five grouped folds in Table 5. Pipelines are implemented in Python using pandas (data handling), scikit-learn (pre-processing, LR/SVC/RF, metrics), and xgboost (XGB). Random seeds are fixed where applicable to reduce variance across runs. All transformations are fit on training folds and applied to test folds via the learned parameters.

F.7 Obtaining Coefficients For the Best Performing Model

Based on mean AUC across grouped folds, the ALL, 5s LR configuration is the top-performing binary model and is used for coefficient-based interpretation. After fitting LR on each training fold, we export the fold-specific coefficient vector to CSV (feature name, coefficient). We then outer-join the five CSVs by feature, and compute each feature’s cross-fold mean and standard deviation (SD). Features are ranked by mean coefficient; the top $K=20$ are visualized as horizontal bars with \pm SD error bars in Figure 5. Bars to the right (positive mean) indicate a higher probability of an enhancing mindset; bars to the left (negative mean) indicate the opposite, with a vertical line at 0 for reference.

G Detailed Result Description for Figure 5

The coefficient profile reveals a mixture of within-session and contrast features from EDA, HR, and ACC. Because all predictors were standardized, larger absolute means with small SDs indicate stronger and more stable associations with the probability of an above-midpoint mindset. Several HRV contrast features, including (Δ Stress – Rest) variants of LF/HF and normalized HF/LF powers, exhibit positive mean coefficients. This implies that windows in which stress-period HRV increases relative to rest on these indices are more likely to be associated with above-midpoint mindset. Interpreted cautiously, this pattern aligns with reports that more adaptive stress appraisals are accompanied by more regulated autonomic responses during challenge (e.g., relatively preserved vagal activity or efficient sympathetic–parasympathetic coupling), which manifest as favorable frequency-domain HRV changes under stress [30, 40]. EDA phasic statistics, such as counts above/below mean and low-percentile/minimum values, including (Δ S2 – M2) and (Δ Stress – Rest) terms, feature prominently with both positive and negative signs. Positive coefficients for phasic counts/percentiles indicate that more frequent or sustained phasic

activity (or larger phasic baselines in stress vs. rest) tends to align with positive stress mindset, consistent with engaged yet regulated arousal. Negative coefficients for minima or counts below mean suggest that suppressed phasic activity or dips relative to baseline are linked to stress is debilitating mindset. Multiple accelerometer summaries, sign-change counts, skewness, kurtosis, and counts below mean across sessions/contrasts, show negative mean coefficients. Given that residual motion can both confound EDA or HR, and reflect restlessness, these negative associations could mean that more irregular or bursty movement is linked to debilitating mindset, while comparatively calmer posture aligns with enhancing mindset, which aligns with prior work [8, 74]. Error bars for the most influential features are modest, indicating that the direction and magnitude of their associations are reasonably stable under participant-grouped cross validation, despite the sample size. This cross-fold stability strengthens confidence in the observed patterns.

H Detailed Methodology Description for Answering RQ3

The final and most critical step in our investigation was to assess the practical utility of mindset knowledge. We hypothesized that incorporating mindset into the standard stress detection pipeline would improve performance. To test this, we designed two experiments: one treating mindset as a simple feature added to a generic model (Section H.1), and a second experiment that tested whether training entirely separate models for each mindset group would yield even greater performance gains (Section ??). The main objective in this setting is to examine whether knowing stress mindset, like we did in Section 6.1, would be beneficial in improving the momentary stress detection performance, which is the typical task performed in most of the prior stress inference studies [82, 83, 122, 126].

H.1 Stress Mindset as a Covariate

From the windowed features we construct sensor-based feature spaces by using features from different modalities (prefix/sub-string rules in `_select_feature_columns`): HR (heart rate / inter-beat interval / blood volume pulse), EDA (tonic/phasic electrodermal activity), ACC (tri-axial accelerometer). Then, we evaluate with and without adding the scalar `StressMindset` as an additional covariate to the model. Including `StressMindset` as a static user covariate tests whether coarse, questionnaire-derived priors improve within-window stress discrimination and quantifies the incremental value of psychosocial context at deployment (e.g., via a brief intake of SMM or inferred as shown in Section 6.1).

For each configuration {setup, window length W } we train a Random Forest (RF) classifier using a leakage-safe, nested cross-validation design⁴. The outer loop uses `StratifiedGroupKFold` with $K=5$ folds, stratifying on the binary stress label while grouping by User so that no participant appears in both train and test in any fold. We repeat the outer CV across five random seeds (`outer_seeds = {42, 52, 62, 72, 82}`), yielding up to $5 \times 5=25$ outer evaluations per configuration. This makes the results more robust.

⁴It is worth noting that, we trained multiple model types for this setup, including random forest, XGBoost, multilayer perceptron neural networks, and support vector classifiers. We only report results for random forests to keep the paper under page and word limits, since it is the best performing model for momentary stress detection.

Within each outer training split, we perform a randomized hyperparameter search (20 trials) over the following RF space: $n_estimators \in \{200, 300, 400, 600, 800\}$, $max_depth \in \{None, 6, 8, 10, 12, 16, 20, 24, 30\}$, $min_samples_split \in \{2, 5, 10, 20\}$, $min_samples_leaf \in \{1, 2, 4, 8\}$, $max_features \in \{\text{"sqrt"}, \text{"log2"}, 0.3, 0.5, 0.7, None\}$, $bootstrap \in \{True, False\}$. The estimator is wrapped in a pipeline with median imputation (`SimpleImputer(strategy="median")`) and class weighting (`class_weight="balanced"`). The inner loop uses 3-fold StratifiedGroupKFold (same grouping by User). The inner objective is AUC on held-out inner folds; the best hyper-parameters are refit on the full outer-train split and evaluated on the outer-test split. We report Macro AUC which first groups outer-test instances by User, computes a per-user AUC when both classes are present, and averages those AUCs across users in that split:

$$AUC_{macro} = \frac{1}{|\mathcal{U}_{test}|} \sum_{u \in \mathcal{U}_{test}} AUC(y^{(u)}, \hat{p}^{(u)}),$$

where \mathcal{U}_{test} is the set of users in the outer-test fold with both classes. Finally, we aggregate across repeats and folds by reporting the mean \pm standard deviation of each metric (`macro_mean`, `macro_std`). It is worth noting that StratifiedGroupKFold prevents subject leakage (grouping by User) while preserving class balance across folds (stratification), which is important under potential stress/non-stress imbalance. $K=5$ outer folds provide a good bias-variance trade-off with $N=23$ participants (roughly 4–5 users per test fold) and are computationally lighter than leave-one-user-out while reducing estimate variance via repeated seeds (5 \times). The inner $K=3$ balances hyperparameter reliability against runtime: nested CV already multiplies compute by outer folds and repeats; three inner folds are a pragmatic choice that still guards against over-tuning. Further, RF model non-linear interactions among heterogeneous features (HR/EDA/ACC), are robust to monotonic transformations and outliers and, with `class_weight="balanced"`, handle class imbalance without explicit resampling; median imputation is simple and compatible with tree splits. RF models have shown to be good performers for time series tabular data in diverse domains [53, 70, 77], hence a good choice for our problem setting. Moreover, RandomizedSearchCV with 20 trials efficiently explores a relatively wide hyperparameter space and empirically approaches the performance of exhaustive search at a fraction of the cost. Further, AUC is threshold-independent and appropriate for imbalanced data; reporting per-user averaged (macro) AUC distinguishes overall discrimination from cross-user consistency, which is critical for subject-independent deployment. Finally, repeating the outer loop across multiple seeds stabilizes estimates against a single unfavorable fold partition and yields more reliable macro means and standard deviations. The results for this analysis is provided in Figure 6.

H.2 Stress Mindset-Specific Models

We examined whether training separate stress classifiers for participants with an enhancing vs. debilitating stress mindset improves performance over a single, one-size-fits-all model. Using the windowed feature set produced by our pipeline, we first assigned each participant to a mindset group Enhancing or Debilitating. We then

created three analysis partitions: all users (one-size-fits-all baseline), enhancing only, and debilitating only. For each partition and each time window $w \in \{2, 5, 10, 20, 30\}$ seconds, we trained a RF classifier inside a nested grouped cross-validation. Experimental setup was similar to the previous section with 5 fold outer and 3 fold inner cross validation. Grouped CV reflects the deployment setting, generalization to unseen users, and avoids overly optimistic estimates caused by subject overlap. Finally, the range of window sizes probes the temporal resolution at which stress signatures are most discriminative in wrist-worn signals. The results for performance for the three user groups (All, Enhancing, Debilitating) for different time windows is provided in Figure 7.