Stress and Recovery Tracking Using Consumer Wearables: A Pilot Study

1st Yuki Yang
Carnegie Mellon University
Pittsburgh, PA, USA
yukiy@andrew.cmu.edu

2nd Alfredo Romo Osorno
Carnegie Mellon University
Pittsburgh, PA, USA
aromooso@andrew.cmu.edu

3rd Fernando Ruiloba Portilla
Carnegie Mellon University
Pittsburgh, PA, USA
fernandor@andrew.cmu.edu

Abstract—

Index Terms—**mHealth, wearable sensing, stress detection, machine learning, HRV**

I. INTRODUCTION

Chronic stress and individuals' capacity for recovery play a pivotal role in both mental and physical health. Traditional stress assessment approaches — such as self-report questionnaires or occasional clinical visits — are inherently limited: they are episodic rather than continuous, rely on subjective recall, and often fail to capture daily or intra-day fluctuations in stress and recovery. As a result, subtle but persistent patterns of stress accumulation, or deficits in recovery, may go undetected. In contrast, consumer-grade wearable devices — such as smartwatches, smart rings, and fitness trackers — have become increasingly common. These devices enable continuous, unobtrusive monitoring of physiological signals (e.g., heart rate variability, resting heart rate) and behavioral / lifestyle metrics (e.g., sleep duration and quality). This technological trend offers a promising opportunity for long-term, real-world monitoring of stress and recovery dynamics, potentially bridging gaps left by traditional assessments [1,2].

In this study, we aim to investigate whether consumer-grade wearables can be used to reliably detect and predict daily stress and recovery levels in real-world settings. Specifically, we address the following research questions:

1) Can daily aggregated features derived from wearable data (e.g., HRV, sleep metrics) reliably classify days into "high-stress" vs "low/normal-stress"?
2) Beyond binary classification, can wearable-derived features support prediction of *continuous* stress levels (i.e., a regression approach)?
3) Does the integration of multiple data sources — physiological (HRV), behavioral (sleep), and subjective self-report — enhance prediction performance, compared to using HRV alone?

The main contributions of this work include:

- Construction of a unified daily-scale dataset combining multi-device, multi-modal data (HRV, sleep, self-report) from consumer wearables.

- Implementation of both *classification* (stress-day detection) and *regression* (continuous stress prediction) models, providing more granular and practical utility for stress monitoring in everyday life.
- Use of daily-scale aggregation over longitudinal periods — offering real-world ecological validity, rather than short-term or lab-based snapshots common in prior stress studies.
- Exploration of the *feasibility and limitations* of consumer-grade wearables for continuous stress and recovery tracking, providing empirical evidence for both strengths and constraints.

The remainder of this report is organized as follows. In Section 2, we review relevant literature on wearable-based stress monitoring, including sensor validity, common methodologies, and limitations. Section 3 describes the system design, data collection procedures, and feature engineering pipeline. Section 4 presents our experimental results and evaluation, assessing both classification and regression performance. Finally, we discuss implications, limitations, and future directions.

II. RELATED WORK

A. Wearable Sensor–Based Stress and Health Monitoring: Overview

Recent years have witnessed a growing body of research leveraging wearable sensors to monitor physiological and behavioral signals for stress, mental health, and general well-being tracking. A recent systematic review synthesized findings across over 60 peer-reviewed studies from 2016 to 2025, covering a variety of wearable devices (smartwatches, wristbands, rings) and sensor modalities such as heart rate, heart rate variability (HRV), electrodermal activity (EDA), skin temperature, and sleep metrics [1]. These studies commonly target applications such as stress detection, anxiety monitoring, resilience tracking, and general wellness management.

B. Validity and Reliability of Wearable-Derived Physiological Signals (HRV, PPG, etc.)

For wearable-based stress monitoring to be meaningful, the physiological signals measured by consumer devices must be

sufficiently accurate. Several studies have validated HRV (and related metrics) derived from photoplethysmography (PPG) or wearable sensors against gold-standard electrocardiogram (ECG)-based metrics. For instance, a recent investigation demonstrated that certain wearables produce resting HRV measurements that correlate with clinical health and well-being indicators, particularly when recordings are taken during sleep or upon waking — conditions that minimize motion and external confounders [3]. Another validation study comparing multiple popular wearables found that HR and HRV measurements show variable agreement levels across devices: devices like finger-worn rings often outperform wrist-worn smartwatches under nocturnal or resting conditions, while wrist devices may show degraded reliability when users are active or during daytime activities [4,5].

These findings suggest that while consumer wearables can produce HRV metrics of reasonable fidelity under resting or sleep conditions, the accuracy for HRV during daily living — especially when motion or activities are involved — remains a concern [6]. Such variability raises caution in interpreting wearable-derived HRV as a direct proxy for autonomic nervous system (ANS) state without careful context consideration.

C. Empirical Studies: Stress Detection / Prediction Using Wearables + Machine Learning

Beyond physiological measurement validation, a number of empirical studies have attempted to detect or predict stress (or stress-related states) using wearable-derived data combined with machine learning (ML) or statistical modeling. For example, recent research using wearable biosensors in a large cohort (hundreds of participants) applied recurrent neural networks (e.g., LSTM) to HRV time series and demonstrated associations between wearable HRV and self-reported stress, anxiety, and general health status [7]. Another study specifically designed a PPG-based wearable stress detection device: after validating the PPG-derived RR intervals against ECG references, the authors trained classification models that achieved promising stress-detection performance under controlled conditions [8].

More recently, a survey of wearable-based stress detection techniques identified HRV, PPG, EDA as the most commonly used signals, and machine learning models (ranging from random forests to deep neural networks) as mainstream modeling approaches [9]. However, many of these studies suffer from limitations such as small sample sizes, short monitoring periods, and reliance on lab-based stress induction protocols, which limit their ecological validity.

D. Multi-Modal Longitudinal / Real-World Monitoring: Sleep, HRV, EMA, and Recovery

Recognizing the limitations of lab-based studies, a subset of work has focused on real-world, longitudinal monitoring by combining physiological data (HRV, HR), behavioral data (sleep), and ecological momentary assessment (EMA) or self-reported stress. Such multi-modal, real-life studies offer higher ecological validity and capture the day-to-day variability of stress and recovery. For example, one study demonstrated that wearable-derived nocturnal HRV and resting heart rate (RHR) correlate with certain mental health and behavioral indicators over extended monitoring periods — suggesting potential of wearables as general health biomarkers [3]. However, variability across individuals, missing data, and inconsistent associations (e.g., sleep quality correlates only weakly with subjective stress) remain common challenges.

E. Challenges, Limitations, and Gaps in Current Literature

Despite progress, existing literature reveals several recurrent limitations:

- Many validation studies restrict consideration to resting or sleep conditions; accuracy degrades in daily living with motion and activity.
- Sample sizes are often small, and monitoring periods short — limiting longitudinal and large-scale inference.
- Heterogeneity in devices, sensor modalities, data preprocessing, and HRV computation undermines comparability across studies.
- Stress labeling is often inconsistent: some studies rely on self-report (subjective), others on physiological proxies — complicating interpretation.
- Few studies integrate multi-modal data (physiology + sleep + self-report) over long-term real-world monitoring and adopt both classification and continuous prediction approaches.

F. Gap Assessment Positioning of Current Study

Given the limitations and heterogeneity in prior work, there is a clear need for studies that:

- Build *unified, multi-device, multi-modal datasets* covering HRV, sleep, and self-report over longitudinal daily-scale monitoring.
- Evaluate both *classification* (stress-day detection) and *regression* (continuous stress level prediction) to support flexible applications.
- Examine real-world feasibility — including accuracy, noise, missing data, and device variability — rather than idealized lab conditions.

Our current study aims precisely to address these gaps, offering empirical evidence for the feasibility and limitations of wearable-based stress and recovery monitoring in real-life contexts.

III. METHODS

IV. RESULTS

V. DISCUSSION

VI. CONCLUSION

REFERENCES

[1] Kapogianni, N.-A., Sideraki, A., & Anagnostopoulos, C.-N. (2025). Using Smartwatches in Stress Management, Mental Health, and Well-Being: A Systematic Review. *Algorithms*, 18(7), 419. https://doi.org/10.3390/a18070419

[2] Altini, M., & Kinnunen, H. (2021). The Promise of Sleep: A Multi-Sensor Approach for Accurate Sleep Stage Detection Using the Oura Ring. *Sensors*, 21(13), 4302. https://doi.org/10.3390/s21134302

[3] Jerath, R., Syam, M., & Ahmed, S. (2023). The Future of Stress Management: Integration of Smartwatches and HRV Technology. *Sensors*, 23, 7314. https://doi.org/10.3390/s23177314

[4] Pinge, A., et al. (2024). Detection and Monitoring of Stress Using Wearables: A Systematic Review. *Frontiers in Computer Science*. https://doi.org/10.3389/fcomp.2024.1478851

[5] Herberger, S., et al. (2025). Performance of wearable finger-ring trackers for diagnostic sleep staging: A validation study. *Scientific Reports*. (online first) https://doi.org/10.1038/s41598-025-93774-z