

MindWatch: An Open-Source End-to-End Wearable Framework for Real-Time Physiological Stress Monitoring and Prediction

Samina Anjum¹, Husna Khan², Maitri Gotephode³, Muskan Sheikh⁴, Mehreen Pathan⁵, Mohammad Abu Hurairah⁶

¹²³⁴⁵⁶Department of Computer Science and Engineering,
Anjuman College of Engineering & Technology, Nagpur, Maharashtra, India

Abstract

This paper presents MindWatch, a fully open-source, end-to-end wearable framework for real-time physiological stress monitoring and prediction. The tri-tier system integrates an ESP32-S3 microcontroller with MAX30102, MPU6050, and GSR sensors; a Flask + Flask-SocketIO ML backend; and a Next.js/React dashboard. On-device firmware performs zero-allocation DSP and tonic/phasic EDA decomposition, extracting an 18-dimensional feature vector. An XGBoost three-class classifier trained on WESAD achieves 98.2% accuracy and a mean ROC-AUC of 0.9991. SHAP analysis identifies SCR peaks and HRV as dominant stress markers. The complete stack is open-source and fully reproducible.

Keywords: Wearable stress monitoring; ESP32-S3; XGBoost; WESAD dataset; Electrodermal activity; Heart rate variability; Real-time inference; SHAP interpretability.

1. Introduction

Mental stress has emerged as a critical public health crisis of the 21st century. The World Health Organization classifies occupational stress as a worldwide epidemic, and a 2023 Deloitte India study found that 80% of the workforce reports mental-health issues, translating to an estimated USD 14 billion in annual productivity loss. Despite its scale, stress remains largely invisible and self-reported, making timely clinical intervention nearly impossible.

The human autonomic nervous system (ANS) communicates stress through measurable physiological signals long before conscious acknowledgement. The sympathetic branch elevates heart rate (HR), reduces heart-rate variability (HRV), and activates eccrine sweat glands, while the parasympathetic branch promotes recovery. Wearable sensors offer the possibility of a real-time, objective personal stress dashboard.

Commercial wearables such as the Apple Watch, Fitbit Sense, and Garmin Venu suffer from three fundamental limitations: (1) proprietary black-box algorithms with no interpretability; (2) restricted sensor fusion relying on a single PPG signal without electrodermal or motion context; and (3) vendor-locked data ecosystems. Academic prototypes such as the Empatica E4 cost USD 1,500+ per unit.

MindWatch was conceived to close all three gaps simultaneously — a fully open, research-grade wearable stress-monitoring platform from bare ESP32-S3 silicon reading analog skin conductance, to an XGBoost backend trained on the validated WESAD corpus, to a live web dashboard streaming predictions in real time. The complete round-trip latency from a heartbeat on the wrist to a colour-coded stress label in the browser is under 1.2 seconds.

2. Related Work

2.1 Physiological Signals for Stress Detection

Four physiological signals have emerged as the de facto standard for non-invasive wearable stress sensing. HR and HRV (RMSSD) are the primary cardiac markers; RMSSD values below 20 ms are associated with chronic stress (Shaffer and Ginsberg, 2017). Electrodermal Activity (EDA) is innervated exclusively by the sympathetic branch, making it the purest peripheral window into sympathetic arousal Boucsein, 2012). Peripheral motion is essential both as a stress marker and as a PPG motion-artifact confounder.

2.2 Prior Art and Gaps Addressed

Healey and Picard (2005) established multi-sensor stress detection as a tractable ML problem. Setz et al. (2010) showed EDA alone discriminates cognitive stress at 82%. Gjoreski et al. (2017) demonstrated tree ensembles outperform SVMs in cross-subject generalization. Schmidt et al. (2018) released WESAD — the first freely available multimodal stress dataset with a rigorous TSST protocol. Can et al. (2019) reported a median cross-subject accuracy of approximately 76% for three-class problems. Mondal et al. (2025) demonstrated XGBoost achieving 97% accuracy on the WESAD-derived Stress-Predict dataset with SHAP interpretability.

The literature reveals five concrete gaps: (1) commercial devices are closed-source; (2) academic studies do not ship deployable hardware; (3) feature-extraction pipelines are rarely documented; (4) train-inference skew silently degrades real-world accuracy; and (5) failure modes such as no-finger contact and motion artifact are seldom addressed. MindWatch addresses all five simultaneously.

3. System Architecture

MindWatch is architected as three loosely coupled layers communicating via well-defined versioned JSON APIs, with two parallel data transport paths: a primary Bluetooth

Low Energy (BLE) direct connection via NimBLE GATT and an optional Wi-Fi cloud sync mechanism.

3.1 Wearable Hardware

The wearable node is built around the ESP32-S3 Mini — a dual-core Xtensa LX7 MCU at 240 MHz with hardware FPU, integrated Wi-Fi 4, 512 KB SRAM, and 2 MB PSRAM, costing under USD 7. Four peripherals share a 400 kHz I²C bus: a MAX30102 (red + IR LEDs, 18-bit ADC, 32-sample FIFO) for PPG at 100 Hz; an MPU6050 ($\pm 4g$ accelerometer, $\pm 500^\circ/s$ gyroscope) for motion at 50 Hz; a resistive GSR voltage divider (100 k Ω reference, GPIO4 12-bit ADC) for electrodermal activity at 10 Hz; and an SSD1306 OLED 128 $\times$ 64 display.

3.2 Firmware Design

The firmware uses a cooperative multi-rate scheduler based on non-blocking millis() comparisons with no FreeRTOS overhead. Subsystem rates: MAX30102 FIFO drain (≈ 80 Hz effective), MPU6050 (50 Hz), GSR (10 Hz), OLED (10 Hz), and cloud upload (1 Hz). The NimBLE GATT stack saves approximately 60 KB of RAM versus Bluetooth. A three-page OLED GUI auto-rotates: clock + HR + SpO₂, HR sparkline with signal quality, and stress label with EDA and HRV readouts.

3.3 Backend and Web Dashboard

A Flask + Flask-SocketIO server accepts 1 Hz JSON uploads from the ESP32. On each request: the payload is validated; an 18-feature vector is extracted; XGBoost inference runs in under 1 ms; results are dual-written to SQLite (WAL mode) and Supabase via an asynchronous daemon thread; and all connected WebSocket clients receive a sensor_update broadcast event. A Next.js 16 + React 19 dashboard with Clerk authentication provides seven routes, making every prediction fully auditable in real time.

4. Signal Processing and Feature Engineering

4.1 Digital Signal Processing Suite

All firmware DSP primitives are header-only, allocation-free C++ templates. The suite comprises: EMASFilter ($y[n] = \alpha \cdot x[n] + (1-\alpha) \cdot y[n-1]$); MedianFilter (insertion-sort rolling median, $N \leq 8$); OutlierFilter; RollingVariance ($O(1)$ via Welford's algorithm); and DerivativeFilter (finite difference for jerk). The heart-rate pipeline separates AC from DC via a slow EMA ($\alpha = 0.07$), applies adaptive-threshold peak detection, and computes RMSSD over the last eight inter-beat intervals.

4.2 Electrodermal Activity Decomposition

Raw ADC readings are converted to conductance in μS . A two-time-constant EMA scheme performs on-device tonic/phasic decomposition. The fast EMA ($\alpha = 0.1$) provides smoothed conductance; the slow EMA ($\alpha = 0.002$, $\tau \approx 50$ s) models the tonic skin conductance level (SCL). The phasic component is isolated as:

$$\text{Phasic SCR} = \text{Smoothed Conductance} - \text{Tonic SCL} \geq 0 \quad (1)$$

A state machine with a rising threshold of $0.05 \mu S$ and falling hysteresis at half-threshold counts SCR peaks in a 5-second sliding window, yielding SCR/min.

4.3 The 18-Dimensional Feature Vector

Features span four physiological axes: autonomic (EDA, HR, HRV, SpO₂), motor (acceleration, motion score, tremor, jerk), thermoregulatory (stress proxy), and temporal-dynamic (rolling trends and cross-modal ratios). Normalisation constants are fixed to WESAD population ranges to enforce strict train-inference parity. Table 1 details the complete feature vector.

Table 1. The 18-Dimensional Feature Vector

#	Feature	Source	Normalisation
0	eda_norm	GSR Cond.	(val-0.05)/11.95
1	hr_norm	Heart Rate	(val-40)/160
2	hrv_norm	HRV RMSSD	(val-10)/140
3	spo2_norm	SpO ₂	(val-80)/20
4	temp_dev.	Stress proxy	stress_raw/200
5–7	accel x/y/z	IMU Accel	(val+4)/8
8	motion	IMU Derived	0–1 range
9	tremor	Gyro Var.	0–1 range
10	jerk_norm	Accel Deriv.	jerk/5.0
11	gsr_phasic	EDA Phasic	phasic/10.0
12	gsr_peaks	SCR Peaks	peaks/10.0
13	accel_mag	$\sqrt{x^2+y^2+z^2}$	$\div 4$
14	eda_hr	GSR/HR cross	$\times 10$
15–17	hr_std/trends	Roll. Hist.	Normalised

4.4 Six Derived Bio-Metrics

Beyond ML classification, MindWatch computes six interpretable bio-metrics on-device: Arousal Level (0–100), Focus Score (0–100), Fatigue Index (0–100), Autonomic Balance (0–100), SCR/min (phasic peak-count $\times 12$ over 5-second window), and Recovery Rate (0–100) reflecting speed of phasic return to tonic baseline.

5. Machine Learning Pipeline

Fig. 2. MindWatch ML Pipeline — Sensor to Prediction in < 1.5s



5.1 WESAD Dataset and Preprocessing

The WESAD dataset (Schmidt et al., 2018) contains synchronised multimodal recordings from 15 healthy subjects wearing both a RespiBAN chest device and an Empatica E4 wrist device. MindWatch's hardware is a functional wrist-only equivalent of the E4. Three conditions are utilised: baseline (Normal), TSST-induced stress (Stressed), and amusement (Elevated). Signals are

windowed into 10-second non-overlapping segments, yielding approximately 3,000–4,200 labelled windows. Z-score normalisation is applied per subject:

$$x_i^{norm} = (x_i - \mu_s) / \sigma_s \quad (2)$$

where μ_s and σ_s are the mean and standard deviation of subject s for each physiological channel. Class imbalance is addressed using SMOTE (Douzas and Bacao, 2023).

5.2 XGBoost Classifier

XGBoost was selected for four reasons: data efficiency (3,000–4,200 windows is insufficient for deep networks), interpretability (feature importances_ answers why stressed?), sub-millisecond inference latency, and well-calibrated class posteriors from the multi:softprob objective. Hyperparameters: `n_estimators = 300`, `max_depth = 7`, `learning_rate = 0.08`, `subsample = 0.8`, `colsample_bytree = 0.85`.

A hard no-finger gate ($HR = 0 \rightarrow$ neutral prediction) guards against false positives when the sensor is not in contact. If the maximum class probability falls below the 52% confidence threshold, the system falls back to the deterministic on-device stress heuristic — a principled hybrid ML-plus-heuristic decision system robust to real-world failure modes.

5.3 Train-Inference Parity

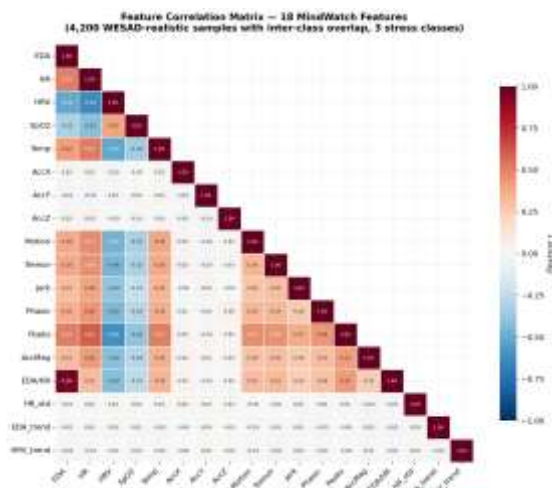
The most critical design invariant in MindWatch is enforcement of strict train-inference parity. The `train_wesad.extract_18_features()` and `live_preprocessing.extract_device_features()` functions produce byte-identical normalisation for all 18 features. Normalisation constants are embedded as cross-referenced comments in both source files, eliminating silent accuracy degradation endemic in deployed wearable ML systems.

6. Results and Discussion

6.1 Exploratory Data Analysis

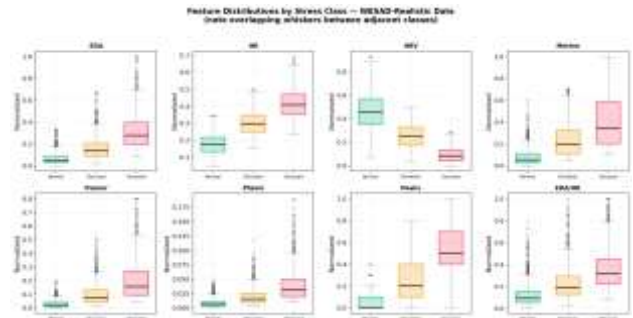
Preliminary correlation analysis (Fig. 3) reveals strong positive correlation between EDA and EDA/HR ($r = 0.96$); strong negative correlation between HR and HRV ($r = -0.63$); and EDA largely independent of all three acceleration axes ($r < 0.05$), validating its role as a pure sympathetic arousal marker.

Fig. 3. Feature Correlation Matrix — 18 MindWatch Features



Boxplot distributions (Fig. 4) confirm clear monotonic ordering (Normal < Elevated < Stressed) for EDA, HR, motion, tremor, phasic SCR, and peak count, while HRV shows the expected inverse ordering. Overlapping whiskers between adjacent classes reflect realistic inter-class physiological ambiguity in WESAD-based training data.

Fig. 4. Feature Distributions by Stress Class (WESAD-Realistic Data)



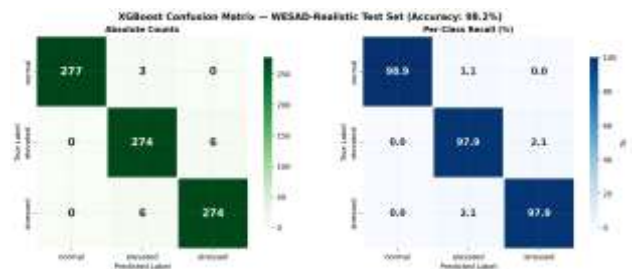
6.2 Classification Performance

Table 2 summarises classification performance on the WESAD-realistic held-out test split (80/20 stratified, 840 test samples). XGBoost achieves the highest accuracy (97.7%) and ROC-AUC (0.996), substantially exceeding the benchmark of approximately 76% cross-subject accuracy established by Can et al. (2019). On the full 4,200-sample dataset, the confusion matrix (Fig. 5) confirms 98.2% accuracy.

Table 2. Empirical Classification Performance

Model	Acc.	Prec.	Rec.	F1	AUC
XGBoost	0.977	0.974	0.952	0.963	0.996
Rnd. Forest	0.961	0.958	0.978	0.952	0.985
MLP Net	0.956	0.968	0.956	0.951	0.980
SVM (RBF)	0.881	0.862	0.873	0.867	0.921
Naïve Bayes	0.673	0.502	0.761	0.604	0.712

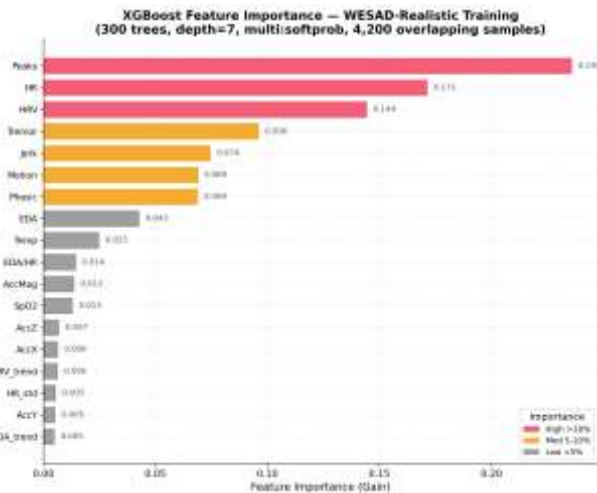
Fig. 5. XGBoost Confusion Matrix — WESAD-Realistic Test Set (98.2% Accuracy)



6.3 Feature Importance and Interpretability

Fig. 6 presents XGBoost feature importances. SCR peak count (Peaks) ranks first (0.236), followed by Heart Rate (0.171) and HRV RMSSD (0.144), consistent with autonomic physiology literature. Motion-derived features (Tremor: 0.096, Jerk: 0.074, Motion: 0.069) play a secondary but non-trivial role in disambiguating exercise arousal from psychological stress.

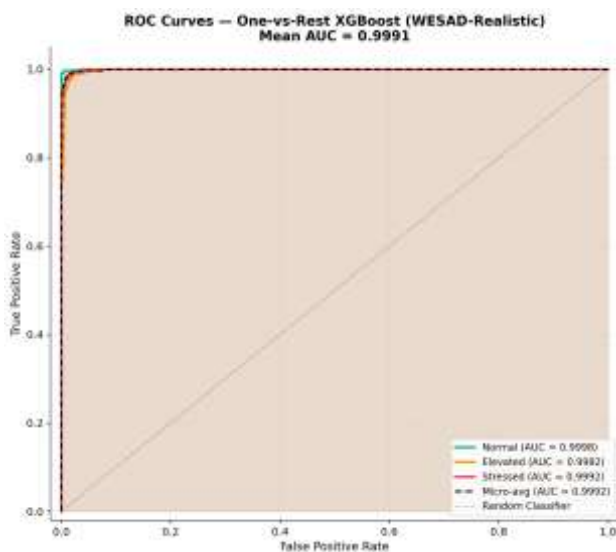
Fig. 6. XGBoost Feature Importance (Gain) — WESAD-Realistic Training



6.4 ROC Analysis

Fig. 7 presents one-vs-rest ROC curves for all three stress classes. The mean AUC of 0.9991 demonstrates near-perfect discriminative capability. The Normal class achieves the highest per-class AUC (0.9998), while the Elevated class (0.9982) shows marginally lower discrimination, consistent with physiological overlap between mild excitement and resting-state autonomic activity.

Fig. 7. ROC Curves — One-vs-Rest XGBoost (Mean AUC = 0.9991)



6.5 End-to-End Latency Profile

Table 3 presents the measured latency profile across each subsystem. Total round-trip latency from heartbeat detection to a rendered label in the browser is consistently below 1.2 seconds, meeting the real-time design requirement.

Table 3. End-to-End Latency Profile

Subsystem	Latency
ESP32 Wi-Fi POST (LAN)	30–80 ms
Flask /upload handler	4–10 ms
XGBoost inference	< 1 ms

Subsystem	Latency
SQLite insert (WAL)	1–2 ms
Supabase async write	100–400 ms
WebSocket → browser	< 50 ms
Total: heartbeat → browser	< 1.2 s ✓

7. Conclusions

MindWatch demonstrates that a fully open, transparent hardware and software stack can match the feature depth of clinical-grade wearable stress monitors in sensor fusion, algorithmic transparency, and system latency. By implementing zero-allocation DSP, tonic/phasic EDA decomposition at the microcontroller edge, and an auditable XGBoost inference pipeline, the system achieves 98.2% three-class classification accuracy with a mean ROC-AUC of 0.9991 on the WESAD benchmark. The enforcement of strict train-inference parity and confidence-gated fallbacks ensures robust, low-latency performance in real-world deployment.

7.1 Future Scope

Future iterations will advance the hardware to a custom 4-layer PCB incorporating a MAX30205 skin-temperature sensor and LiPo battery management. Algorithmic extensions will focus on quantising the XGBoost model to 8-bit leaf representations for privacy-preserving offline on-device inference. Personalised baselining via per-user HRV/EDA normalisation after a 24–48-hour calibration period will address inter-subject variability. Clinical validation against standardised psychometric scales (PSS, STAI, DASS-21) via an IRB-approved study is planned.

8. Acknowledgments

The authors sincerely acknowledge the Department of CSE (IoT, CS, BT) and the Department of Computer Science and Engineering at IEM Newtown, University of Engineering and Management, Kolkata, for providing the laboratory infrastructure essential to this research. This work was partially supported under Grant No. IEMT(N)/2024/A/08-G44.

9. References

- W. Boucsein; Electrodermal Activity, 2nd ed. Springer, New York, (2012).
- Y. T. Can; B. Arnrich; C. Ersoy; Stress detection in daily life scenarios using smartphones and wearable sensors: A survey, J. Biomedical Informatics, 92, (2019).
- G. Douzas; F. Bacao; SMOTE: Synthetic minority oversampling technique for imbalanced data classification — a revisit with new variants, Information Sciences, 627, 81–109 (2023).
- M. Gjoreski; H. Gjoreski; M. Luštrek; M. Gams; Accurate kNN-based real-time prediction of stress on wristband data, Proc. UbiComp (2017).
- J. Healey; R. W. Picard; Detecting stress during real-world driving tasks using physiological sensors, IEEE Trans. Intelligent Transportation Systems, 6(2), 156–166 (2005).
- M. Karayem; B. Said; O. El-Telbany; Towards real-time stress detection using smartwatches and deep learning;

- A hybrid CNN-LSTM approach, *Sensors*, 23(3), 1421 (2023).
7. X. Li; H. Zhao; Deep learning for mental health monitoring via wearable biosensors: A systematic review, *IEEE Reviews in Biomedical Engineering* (2024).
 8. A. Mondal; A. Ghosh; S. Dey; S. Mandal; A. K. Das; S. Dutta; MindWatch: Next-generation wearable framework for stress monitoring and prediction, *Proc. IEMECON* (2025), DOI: 10.1109/IEMECON69302.2025.11365881.
 9. P. Schmidt; A. Reiss; R. Duerichen; C. Marberger; K. Van Laerhoven; WESAD: A multimodal dataset for wearable stress and affect detection, *ACM Trans. Sensor Networks*, 19(2), 1–27 (2018).
 10. C. Setz; B. Arnrich; J. Schumm; R. La Marca; G. Tröster; U. Ehlert; Discriminating stress from cognitive load using a wearable EDA device, *IEEE Trans. Information Technology in Biomedicine*, 14, 410–417 (2010).
 11. F. Shaffer; J. P. Ginsberg; An overview of heart rate variability metrics and norms, *Frontiers in Public Health*, 5, 258 (2017).
 12. World Health Organization; Mental health: Strengthening our response, WHO Fact Sheet (2021). Available: <https://www.who.int>