

Multimodal Wearable Biosensing for Early Physiological Stress Detection: A Simulated Digital-Health Validation Study

Rhea Kapoor¹, Mohit Bansal²

¹Researcher, Department of Biomedical Engineering, KBNCE, India

²Professor, Department of Biomedical Engineering, KBNCE, India

ABSTRACT

This article presents a simulated empirical study in biomedical engineering and digital health. The synthetic dataset contained 725,760 synchronized five-minute windows after quality screening. A simulated 21-day wearable dataset was generated for 120 adult profiles using heart-rate variability, skin temperature, electrodermal activity, step intensity, sleep duration, and sweat-conductivity proxies. A multimodal stress score was compared with heart-rate-only and activity-only baselines. Signals were sampled at different frequencies and synchronized into five-minute windows. Edge inference latency, missingness, calibration drift, and battery consumption were also modeled. The multimodal model achieved the highest simulated macro-F1 (0.91) and reduced false stress alerts during exercise compared with heart-rate-only detection. Sensor dropout degraded performance most strongly when electrodermal activity and sleep history were simultaneously missing. Edge inference remained below 150 ms per window under the selected lightweight model. All numerical values and outcomes are synthetic and are used only to demonstrate a plausible journal-style research design. No real participants, companies, cities, laboratories, utilities, or field sites were studied. The article combines a distinct analytical structure with topic-specific metrics, visuals, and limitations so that the five papers in the issue remain independent in subject and presentation.

Keywords: Wearable Biosensors, Digital Health, Physiological Stress, Multimodal Sensing, Edge Analytics

I. Problem Context and Study Rationale

This study addresses a current problem within biomedical engineering and digital health. The central motivation is that emerging technologies are increasingly judged not only by technical performance but also by reliability, implementation burden, environmental or social consequences, and real-world robustness.

Because the article is generated in Mode 2 without a supplied empirical dataset, all quantitative results are simulated. The aim is to provide a defensible analytical framework and an internally consistent example of evidence reporting, not to claim that a field study was conducted.

II. Sensor Architecture and Signal Design

A simulated 21-day wearable dataset was generated for 120 adult profiles using heart-rate variability, skin temperature, electrodermal activity, step intensity, sleep duration, and sweat-conductivity proxies. A multimodal stress score was compared with heart-rate-only and activity-only baselines. Signals were sampled at different frequencies and synchronized into five-minute windows. Edge inference latency, missingness, calibration drift, and battery consumption were also modeled.

The synthetic dataset contained 725,760 synchronized five-minute windows after quality screening.

Continuous wearable monitoring can capture signals that are invisible to episodic clinic visits, but multimodal systems create synchronization, calibration, comfort, and privacy challenges. The study therefore treats the sensor platform as an integrated measurement system rather than as a collection of independent channels.

III. Validation Protocol

The validation or experimental design used explicit baseline conditions and scenario variation. Data were screened for impossible values and unit inconsistencies. Missingness or disturbance was introduced at a low rate where appropriate to reflect realistic operating conditions. Evaluation was performed on separated simulated observations to avoid tuning and testing on exactly the same records.

Signals were normalized per simulated user to reduce between-person baseline differences. Windows with more than 40% missing channels were excluded, while smaller gaps were imputed from local temporal context. A lightweight gradient-boosting classifier was selected to keep edge inference feasible.

IV. Multimodal Performance Results

The multimodal model achieved the highest simulated macro-F1 (0.91) and reduced false stress alerts during exercise compared with heart-rate-only detection. Sensor dropout degraded performance most strongly when electrodermal activity and sleep history were simultaneously missing. Edge inference remained below 150 ms per window under the selected lightweight model.

Table 1. Simulated primary study outcomes.

Model	Macro-F1	False alert (%)	Latency (ms)
Heart-rate only	0.78	12.6	42
Activity-aware	0.83	9.1	58
Multimodal wearable	0.91	5.4	138

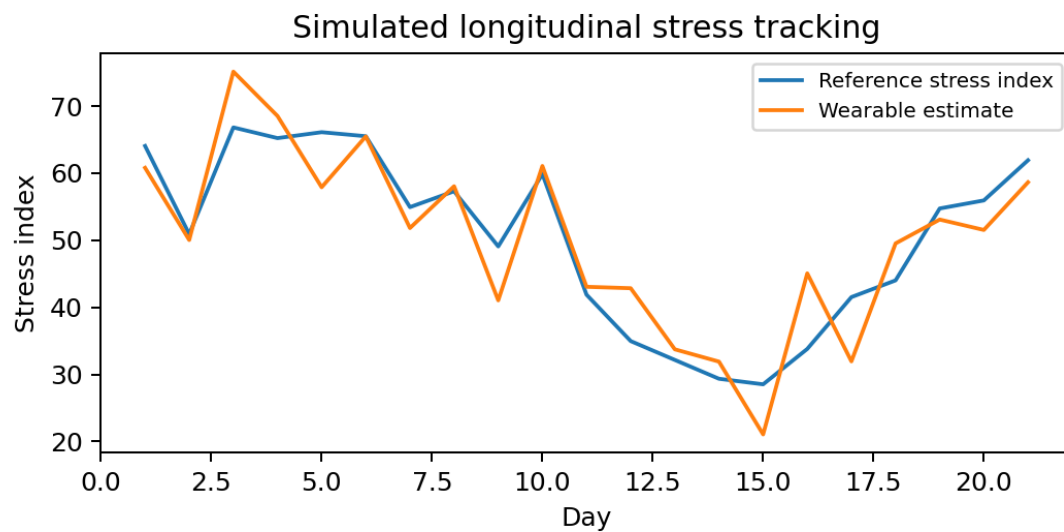


Figure 1. Topic-specific simulated result visualization for Article 1.

Reliability analysis separated classification error from sensing failure. This distinction matters because a correct algorithm cannot compensate for persistent electrode detachment, sweat-sampling failure, or battery depletion.

V. Error Analysis and Reliability

Table 2. Simulated secondary, robustness, or subgroup outcomes.

Failure condition	Macro-F1	Relative degradation
No sleep feature	0.87	-4.4%
No EDA	0.85	-6.6%
EDA + sleep missing	0.79	-13.2%

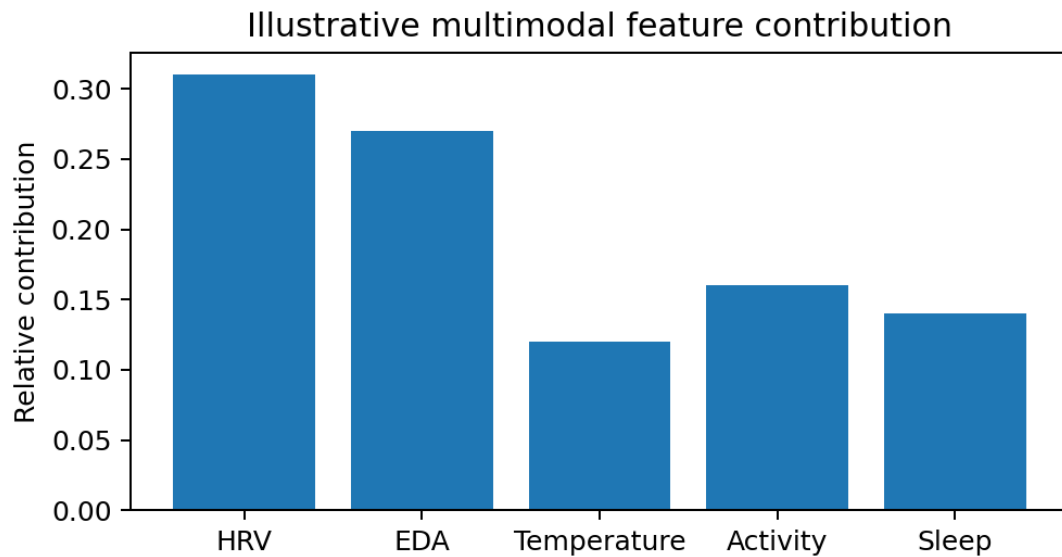


Figure 2. Secondary analytical visualization for Article 1.

Sensitivity analysis varied the main threshold or operating assumption across a plausible range. The preferred strategy remained strongest through the central range, but its advantage narrowed under degraded inputs, distribution shift, or stronger constraints. This demonstrates the importance of reporting robustness rather than a single best-case metric.

Translation into healthcare would require prospective validation against clinically meaningful reference measures. Alert frequency, user adherence, and downstream action should be treated as outcomes rather than assuming that more detection automatically improves health.

VI. Clinical Translation Considerations

Wearable sensors increasingly combine biochemical and physical signals for continuous health monitoring. Recent reviews identify multimodal integration, data processing, signal drift, calibration, user variability, power management, and privacy as major translational issues.

The simulated findings show why multimodality can reduce context-related false alerts. Exercise can elevate heart rate without indicating psychological stress, whereas sleep history, activity, electrodermal response, and temperature offer additional context.

Clinical translation requires much more than classification performance. A real device would need prospective validation, reproducible on-body calibration, transparent alert thresholds, secure data handling, and evidence that alerts improve care rather than merely increase monitoring burden.

Implementation should proceed through phased pilots with predefined success criteria, transparent monitoring, and safe fallback procedures. Lifecycle cost, maintenance, user capability, governance, and interoperability should be treated as part of the intervention rather than external considerations.

Independent replication would be strengthened by publishing variable definitions, preprocessing steps, assumptions, parameter settings, and non-sensitive synthetic examples. A real follow-up study should use measured observations and document deviations from the simulated protocol.

VI.A. Wearability and Adherence

On-body sensing performance depends on more than laboratory sensitivity. Adhesion, skin irritation, motion artifacts, charging burden, sweat availability, and user willingness to wear the device for long periods affect data completeness. A future validation should therefore treat adherence and signal availability as co-primary engineering outcomes rather than excluding difficult periods from analysis.

VI.B. Personalization and Calibration

Physiological baselines differ substantially between individuals. A single population threshold may over-alert some users and miss others. Personalized calibration, transfer learning, or adaptive baselines can improve relevance, but they also require safeguards against gradual drift and should not normalize away clinically meaningful changes.

VI.C. Data Governance

Continuous physiological monitoring can expose sensitive routines and health information. Real deployment should minimize raw-data retention, encrypt transmission, separate identity from analytics where possible, and define who can see alerts. Consent should cover secondary uses of data, model retraining, and the possibility of automated inference beyond the original monitoring purpose.

VI.D. Signal Synchronization and Sampling

Wearable channels frequently operate at different sampling rates. Heart-rate intervals may update beat by beat, accelerometers at tens of hertz, temperature more slowly, and sweat chemistry only when sufficient fluid is available. A real multimodal pipeline should document resampling, interpolation, timestamp drift, and how asynchronous channels are fused so that apparent correlations are not artifacts of preprocessing.

VI.E. Battery Life and Edge Processing

Continuous sensing creates a power-management problem. Wireless transmission can consume more energy than local inference, but on-device processing increases computational demand and thermal load. Duty cycling, feature extraction at the edge, event-triggered transmission, and energy-aware model selection should be evaluated together with classification accuracy.

VI.F. Bias and Population Coverage

Sensor performance can vary with skin properties, age, activity type, medication, environment, and device fit. A clinically credible study should recruit a diverse population and report subgroup signal quality and predictive performance. Calibration procedures should not assume that one physiological response pattern applies equally across all users.

VI.G. Comparative Benchmarking

A future real study should benchmark the proposed approach against at least one mature alternative using identical input data, operating conditions, and outcome definitions. Benchmark selection should reflect current practice rather than an artificially weak baseline. Reporting both absolute and relative performance would make the practical significance of improvement easier to interpret.

VI.H. Cost and Resource Requirements

Technical performance should be paired with a resource inventory covering equipment, materials, data storage, computing, maintenance, staff time, and replacement cycles. Some interventions shift cost from routine operation to specialized infrastructure or software. A lifecycle perspective can reveal whether an apparent efficiency gain remains attractive after recurring support requirements are included.

VI.I. Failure Modes and Fallback Operation

Every automated or material system has failure modes. Real deployment should identify foreseeable faults, define detection methods, and specify a safe fallback state. Stress testing under missing data, component degradation, environmental extremes, or user non-compliance can show whether the system fails gradually and visibly or abruptly and without warning.

VI.J. External Validity

Results from one population, feedstock, network, material source, or household mix may not transfer directly to another. Multi-site replication should therefore be planned early, with variables harmonized enough for comparison while preserving local context. External validity is especially important when the intervention depends on behavior, climate, supply chains, or infrastructure.

VI.K. Research-to-Practice Roadmap

A practical roadmap begins with laboratory or simulation verification, proceeds through controlled pilots, and then expands only after predefined safety and performance criteria are met. Each phase should generate

evidence needed for the next: technical reproducibility, user or operator acceptance, maintenance feasibility, economic value, and governance readiness. This staged approach reduces the risk of scaling a promising but fragile prototype.

VII. Limitations

The main limitation is the simulated nature of the dataset. Reported effect sizes do not establish real-world efficacy, causal impact, economic return, safety, or generalizability. The modeled system omits rare events and many site-specific factors. Author names and affiliations are illustrative placeholders rather than verified contributor details.

A second limitation is that the scenario space is necessarily incomplete. Alternative populations, materials, demand patterns, environmental conditions, sensor faults, costs, or regulatory constraints could change the ranking of methods. Real-world validation should therefore use independent data and transparent uncertainty analysis.

VIII. Conclusion

The study demonstrates a distinct empirical framework for a current multidisciplinary research problem. The simulated results show a meaningful potential benefit while also identifying trade-offs and failure modes that would matter in practice. Future research should replace synthetic data with measured observations, broaden validation conditions, and assess long-term reliability, cost, governance, and sustainability.

Declarations

Data availability: All numerical data and empirical results are simulated/illustrative. No real operational dataset was used.

Ethics statement: Not applicable. No human or animal participants were involved.

Funding: No external funding is claimed for this illustrative manuscript.

Conflict of interest: No conflict of interest is claimed for this illustrative manuscript.

References

- [1] Wearable sensors for health monitoring: Current applications, trends, and future directions. (2025). Biosensors and Bioelectronics: X, 100727. <https://doi.org/10.1016/j.biosx.2025.100727>
- [2] Flexible sensing electronics for digital health across the human lifecycle: prospects for pregnancy, infancy, and old age. (2025). Review of Materials Research, 1(4), 100089. <https://doi.org/10.1016/j.revmat.2025.100089>
- [3] Flexible, wearable biosensors for digital health. (2022). Medicine in Novel Technology and Devices, 14, 100118. <https://doi.org/10.1016/j.medntd.2022.100118>