



Real-Time Behaviour-Aware Wearable Intelligence for Women's Safety and Emergency Anomaly Detection

Dhamdhare Shubhangi Tulshidas^{1*}, Sivaram Ponnusamy¹

¹Research Scholar,, Department of Computer Science and Engineering, Sandip University, Nashik, Maharashtra,
India – 422213

²Research Guide,, Department of Computer Science and Engineering, Sandip University, Nashik, Maharashtra,
India – 422213

Abstract- Personal safety technologies for women in public and semi-public spaces remain dominated by manually triggered panic-button applications, a design that depends on the wearer being conscious, unrestrained, and able to deliberately activate a device at precisely the moments when that is least likely. This paper reviews two artificial-intelligence-based wearable sensing approaches capable of detecting an emergency automatically, without a manual trigger: motion-based anomaly and fall detection, which recognizes sudden-impact or struggle-like movement using inertial sensors and deep human-activity-recognition models, and physiological stress sensing, which detects the autonomic-arousal signature of fear using heart-rate and skin-conductance signals already available on consumer wearables. The review identifies three related gaps: deployed safety products remain almost entirely manually triggered despite the maturity of automatic anomaly-sensing methods; motion-based and physiological-based detection have developed as separate research communities and are rarely fused into one behavior-risk score; and the always-on, battery- and privacy-constrained deployment context of a wearable is rarely addressed jointly with detection accuracy. To address these gaps, the paper proposes an integrated four-stage architecture, data acquisition, behavior and anomaly modeling, risk fusion and tier classification, and graduated emergency response, together with an illustrative four-tier behavior-risk scheme that replaces a binary silent-or-SOS design with escalating response actions. Following a literature-synthesis and conceptual-design methodology rather than a primary field trial involving human subjects, the results synthesize literature-reported performance for each modality and work through an illustrative, hypothetical wearable-deployment cohort to show how a graduated response would distribute across everyday use. The discussion highlights the unvalidated fusion gap and on-device privacy and battery constraints as the most consequential caveats, and the paper closes by outlining the consent-based validation studies and standardized, wearable-first benchmarks needed to move from the current, component-by-component evidence base toward a deployable, trustworthy safety wearable.

Keywords: Wearable Sensors, Human Activity Recognition, Anomaly Detection, Behavior Modeling, Emergency Alert Systems, Physiological Signal Monitoring, Edge Computing, Internet of Things, Women's Safety, Real-Time Systems, Marketing, Consumer Behaviour.



I. INTRODUCTION

Personal safety in public, semi-public, and transit spaces remains a widely documented concern for women in many regions, and has motivated a large ecosystem of personal safety mobile applications and wearable devices built around a manually triggered panic alert, such as a dedicated button or a shake gesture that sends a message and location fix to pre-registered contacts (Yarrabothu & Thota, 2015). A manually triggered design has an important structural limitation: it depends on the wearer being conscious, unrestrained, and aware enough of the unfolding situation to deliberately reach for a device and activate it, conditions that are least likely to hold during precisely the moments of greatest danger, such as a physical struggle, sudden incapacitation, or an assailant seizing the phone itself.

Wearable sensing and on-device artificial intelligence offer a complementary, behavior-aware alternative that does not depend on a conscious manual trigger. Inertial motion sensors already used for step counting and fitness tracking can, with appropriate modeling, also recognize the kind of sudden, high-impact, or irregular movement associated with a fall, a struggle, or a forced restraint, following the same signal-processing foundations developed for wearable fall detection (Igual et al., 2013; Mubashir et al., 2013). Photoplethysmography (PPG) and galvanic skin response (GSR) sensors, already embedded in many consumer smartwatches and fitness bands, can similarly detect the sharp autonomic-arousal signature, elevated heart rate, altered heart-rate variability, and increased skin conductance, associated with fear or acute stress, following the physiological stress-sensing foundations established in the affective-computing literature (Healey & Picard, 2005; Picard, 1997). This paper's scope is a systems-level literature synthesis rather than a single controlled experiment: it reviews motion-based and physiological-based wearable anomaly-detection methods that have largely developed along separate research tracks, proposes an integrated architecture that fuses them into a single behavior-risk score with a graduated emergency-response pathway, and works through an illustrative deployment scenario to make the resulting alerting logic concrete.

The remainder of the paper is organized as follows. Section 2 reviews prior work on wearable human activity recognition, motion-based anomaly and fall detection, physiological stress sensing, and the mobile and sensor-network infrastructure that any always-on safety wearable depends on. Section 3 identifies the research gap this review exposes. Section 4 sets out the paper's objectives. Section 5 describes the methodology used to synthesize the literature into a proposed multi-modal system architecture and an illustrative deployment scenario. Section 6 presents the resulting findings, including a literature-synthesis table and the illustrative scenario's outputs. Section 7 discusses the interpretation, implications, and limitations of these findings. Section 8 concludes, and Section outlines future scope.

II LITERATURE REVIEW

Human activity recognition (HAR) using body-worn inertial sensors is the technical foundation this paper builds on for motion-based anomaly sensing. An early and still widely cited study demonstrated that ordinary cell-phone accelerometers, combined with comparatively simple classifiers, could distinguish everyday activities such as walking, sitting, and climbing stairs with useful accuracy, establishing the feasibility of activity sensing on hardware people already carry (Kwapisz et al., 2011). A comprehensive tutorial on body-worn inertial-sensor HAR formalized the now-standard pipeline of windowing a continuous sensor stream, extracting statistical or frequency-domain features from each window, and classifying the resulting feature vector, a pipeline against which later deep-learning approaches are consistently compared (Bulling et al., 2014). A broader survey of wearable-sensor HAR methods concluded that sensor placement and signal preprocessing often matter more for final accuracy than the specific classifier chosen, a finding with direct relevance to where a safety wearable's sensors are physically worn (Lara & Labrador, 2013). A systematic study of segmentation-window length



further showed a direct trade-off between recognition accuracy and detection latency, longer windows improve accuracy but delay the system's response, a trade-off that becomes safety-critical rather than merely a performance metric when the goal is rapid anomaly detection rather than offline activity logging (Banos et al., 2014).

Deep learning has extended this foundation considerably. An architecture combining convolutional feature extraction with recurrent (LSTM) temporal modeling, evaluated across several multimodal wearable-sensor benchmarks, was shown to recognize short, transitional, and irregular movements more reliably than classical feature-engineering pipelines, a capability directly relevant to detecting a brief but violent movement rather than a sustained, easily characterized activity (Ordóñez & Roggen, 2016). A broader survey of deep-learning HAR architectures synthesized reported accuracy gains of convolutional, recurrent, and hybrid networks over classical pipelines across multiple public benchmark datasets, while also noting that most reported gains come from curated, activity-labeled datasets rather than from naturalistic, continuously worn deployments (Wang et al., 2019). Because a safety wearable must run continuously on constrained hardware rather than upload raw sensor data for offline analysis, on-node deep-learning inference has been specifically investigated: compact deep models can be executed directly on wearable-class processors within realistic power and latency budgets, avoiding the connectivity dependence and privacy exposure of a cloud-only pipeline (Ravi et al., 2017).

Motion-based anomaly detection for personal safety has a close technical relative in wearable fall detection, which has been studied far longer and far more systematically. A widely cited review of fall-detection system principles found that threshold-based accelerometer methods are simple and low-power but require careful, often device-specific tuning to avoid classifying vigorous ordinary movement, such as sitting down quickly or jumping, as a fall (Igual et al., 2013). A complementary survey comparing wearable, ambient, and vision-based fall-detection approaches concluded that wearable accelerometer-based systems offer the best combination of cost, mobility, and deployability, at the price of a persistent false-alarm risk from energetic but harmless activity, a trade-off this paper's proposed system inherits directly (Mubashir et al., 2013). More generally, the anomaly-detection literature outside any single application domain frames this same trade-off formally, as a balance between detection sensitivity and the false-positive rate an application and its users can tolerate (Chandola et al., 2009), and deep-learning approaches to time-series anomaly detection, including LSTM-based encoder-decoder reconstruction models and dedicated convolutional anomaly-detection architectures, have been developed specifically to flag rare, previously unseen deviations from a learned normal pattern rather than requiring every anomaly type to be labeled in advance (Malhotra et al., 2016; Munir et al., 2019).

Physiological, affect-oriented sensing developed along a substantially different track, grounded in psychophysiology rather than motion biomechanics. The foundational premise, articulated across the affective-computing literature, is that acute stress and fear produce a measurable, largely involuntary autonomic-nervous-system signature, changes in heart rate and heart-rate variability, increased skin conductance, and altered respiration, that can be sensed non-invasively and continuously (Picard, 1997). This premise was demonstrated directly in a real-world driving study, in which physiological sensors detected elevated stress during objectively more demanding driving conditions, establishing that ambulatory, non-laboratory physiological stress detection is technically feasible outside a controlled setting (Healey & Picard, 2005). More recently, a purpose-built multimodal dataset combining wrist- and chest-worn physiological sensors with self-reported affect labels was introduced specifically to support wearable stress- and affect-detection research, reflecting the field's shift toward the same kind of standardized, wearable-first benchmarking that HAR research established earlier (Schmidt et al., 2018).

A safety wearable is also a networked embedded system, and its feasibility depends on infrastructure literatures beyond sensing and modeling alone. Reviews of the Internet of Things applied to healthcare



describe the now-common architecture of a body-worn sensor node, a smartphone or gateway relay, and a cloud or edge backend, together with the connectivity and interoperability challenges that architecture introduces (Islam et al., 2015). A top-down survey of energy efficiency in wireless sensor networks catalogs the duty-cycling, radio-scheduling, and low-power-processing techniques an always-on, battery-constrained wearable must adopt to remain usable across a full day rather than requiring frequent recharging (Rault et al., 2014), and the TinyML literature on running machine-learning inference on ultra-low-power microcontrollers extends this same efficiency concern directly to the on-node anomaly-detection models this paper's architecture depends on (Warden & Situnayake, 2019).

Location reporting, a core function of any emergency-alert pathway, inherits its own well-documented accuracy limitations: assisted-GPS positioning on consumer mobile hardware has been shown to degrade measurably in urban, indoor, and obstructed environments relative to open-sky conditions (Zandbergen & Barbeau, 2011), a constraint directly relevant to how confidently a safety system can report a wearer's location during an emergency. Finally, ambient audio has been explored as a further anomaly cue in adjacent surveillance contexts, where one-class support-vector-machine classifiers trained only on normal ambient sound have been used to flag acoustically abnormal events without requiring every abnormal sound to be labeled in advance (Rabaoui et al., 2008), a technique with a plausible but largely unexplored extension to detecting a scream or other distress vocalization from a wearable microphone

III. RESEARCH GAP

remain almost entirely manually triggered (Yarrabothu & Thota, 2015), while the technically mature capability to detect an anomaly automatically, from sudden-impact motion patterns as in wearable fall Three gaps emerge from the literature reviewed above. First, deployed personal-safety applications detection (Igual et al., 2013; Mubashir et al., 2013) or from autonomic-arousal physiological signals as in affective computing (Healey & Picard, 2005; Schmidt et al., 2018), has not been integrated into a personal-safety product as a continuous, always-on complement to a manual panic trigger. Second, motion-based and physiological-based anomaly sensing have developed as largely separate research communities, evaluated on their own benchmark datasets and their own application framings, fall detection and general HAR for motion (Bulling et al., 2014; Lara & Labrador, 2013), affect and stress detection for physiology (Schmidt et al., 2018), and very few published studies fuse both evidence streams into a single behavior-risk score or examine what each modality catches that the other misses.

This matters specifically for personal safety because the two modalities are complementary in exactly the way a sudden physical assault differs from a slower-building threatening encounter: a physical struggle produces a strong, unambiguous motion signature but may or may not register physiologically in time to help, while a threatening but not yet physical encounter, being followed, cornered, or verbally threatened, may produce a clear physiological arousal signature well before any motion anomaly occurs at all. Third, the always-on, battery- and privacy-constrained deployment context of a wearable safety device, requiring on-node inference (Ravi et al., 2017; Warden & Situnayake, 2019) under strict energy budgets (Rault et al., 2014) and location reporting with known accuracy limitations (Zandbergen & Barbeau, 2011), is rarely addressed jointly with the detection-accuracy question in the reviewed literature; most HAR and affect-detection studies report accuracy on a benchmark dataset without addressing whether the same model is deployable on the constrained hardware a wearable actually uses.

Taken together, these gaps mean that while the individual technical components, motion-based anomaly classifiers, physiological arousal detectors, low-power on-node inference, and mobile emergency-alert delivery, each have a reasonably well-validated basis in their own literatures, the field lacks an integrated, deployable framework that fuses motion and physiological evidence into one



behavior-risk score, translates that score into a graduated, alert-fatigue-aware emergency response rather than a binary silent-or-SOS decision, and demonstrates, even illustratively, how such a system would behave across a realistic population of everyday wearers. This paper addresses that gap by proposing such an integrated framework and by working through an illustrative deployment scenario that shows how its outputs could be used.

Objectives

- This paper pursues the following specific objectives:
- To systematically review published methods for wearable human activity recognition, motion-based anomaly and fall detection, physiological stress and affect sensing, and the low-power embedded infrastructure a continuously worn safety device depends on, identifying where the published evidence is strong and where it is thin.
- To propose an integrated four-stage system architecture, multi-modal data acquisition, behavior and anomaly modeling, risk fusion and tier classification, and graduated emergency response, that connects motion-based and physiological-based anomaly sensing into a single deployable pipeline.
- To define an illustrative, literature-grounded behavior-risk tier scheme that maps a fused, continuous risk score onto discrete, escalating response actions rather than a single binary alert decision.
- To synthesize literature-reported performance findings for each sensing modality into a single comparative summary, without collapsing non-comparable metrics into a single misleading figure.
- To construct and analyze an illustrative, hypothetical wearable-deployment cohort that demonstrates how a fused behavior-risk score would distribute across everyday use relative to a purely manually triggered alert design.

V. METHODOLOGY

This study follows a systematic literature-synthesis and conceptual-systems-design methodology rather than a primary field trial involving human subjects, a deliberate choice given the sensitivity of collecting motion, physiological, audio, and location data from real participants in a personal-safety context. It proceeds in three stages. First, published literature on wearable human activity recognition, motion-based anomaly and fall detection, physiological stress sensing, and low-power embedded and mobile infrastructure was reviewed and organized by sensing modality, as presented in Section 2. Second, this literature was synthesized into a proposed four-stage system architecture and an accompanying behavior-risk tier scheme, described below, that together define how the reviewed components would be connected in a deployed wearable system. Third, an illustrative, hypothetical deployment scenario was constructed to demonstrate how the architecture's outputs would distribute across a realistic cohort of everyday wearers; this scenario's figures are explicitly hypothetical and are labeled as such wherever they appear. Performance figures reported in Section 6 are drawn directly from the cited original studies rather than measured independently by the authors, and the paper is explicit throughout about which findings are literature-reported and which are illustrative constructions.

Proposed System Architecture

Fig. (1) presents the four-stage pipeline this paper uses to organize the capabilities surveyed in Section 2 into a single deployable system, from raw wearable-sensor data through to a graduated emergency response.

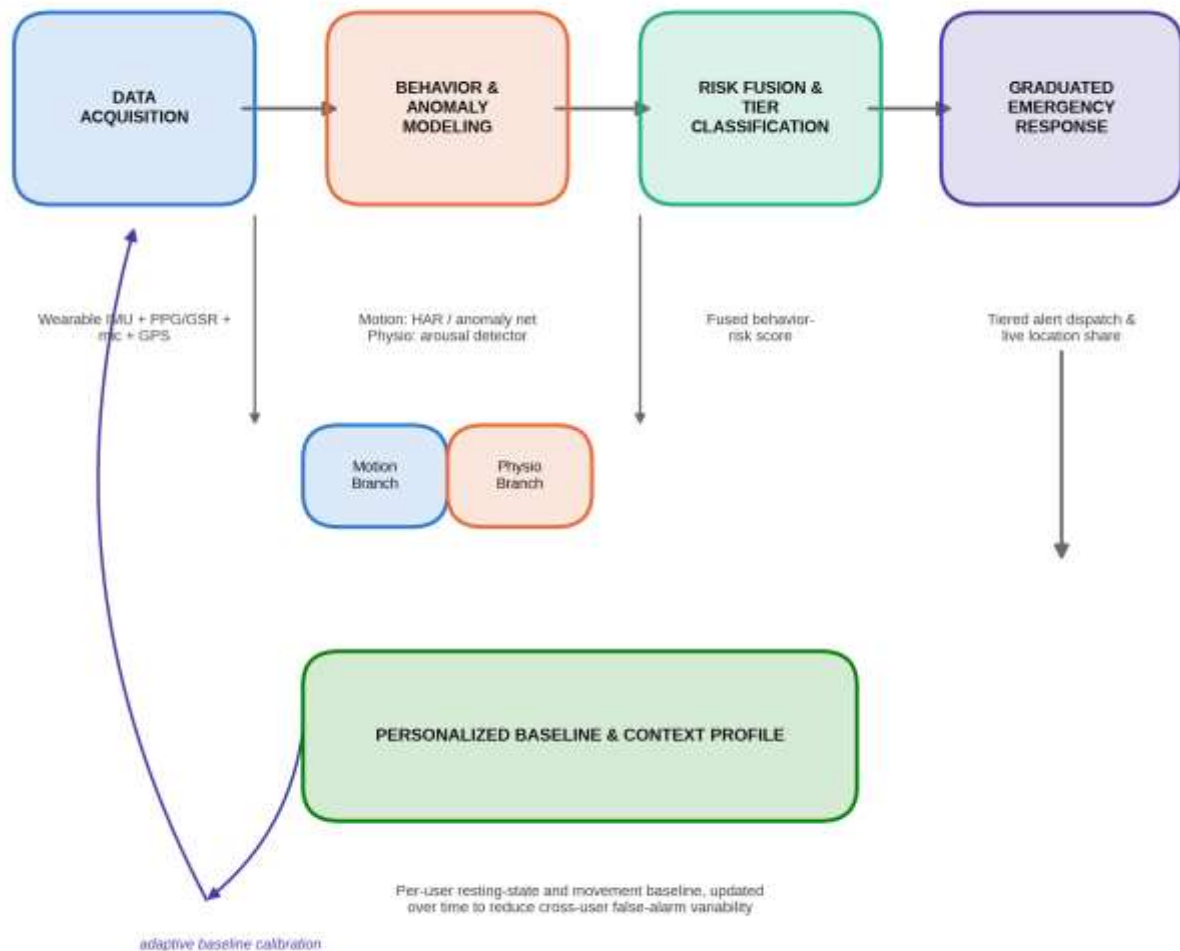


Fig. (1). Proposed pipeline from multi-modal wearable data acquisition through behavior and anomaly modeling and risk fusion to a graduated emergency response, with adaptive per-user baseline calibration.

Data acquisition combines the sensing channels surveyed in Section 2: an inertial-motion channel following the accelerometer-based HAR and fall-detection literature (Kwapisz et al., 2011; Igual et al., 2013), a physiological channel using wrist- or chest-worn PPG and GSR sensing following the affective-computing and WESAD literature (Healey & Picard, 2005; Schmidt et al., 2018), and optional microphone and GPS channels for ambient-audio and location context (Rabaoui et al., 2008; Zandbergen & Barbeau, 2011). Behavior and anomaly modeling then applies the two primary branches surveyed in Section 2 in parallel: a motion branch that recognizes sudden-impact, struggle-like, or otherwise anomalous movement using the deep HAR and anomaly-detection approaches reviewed above (Ordóñez & Roggen, 2016; Malhotra et al., 2016; Munir et al., 2019), and a physiological branch that detects an autonomic-arousal pattern consistent with acute fear or stress (Healey & Picard, 2005). Risk fusion and tier classification combines the two branches' outputs into a single, continuously updated behavior-risk score, addressing the fusion gap identified in Section 3 rather than treating either modality's alert independently. Finally, the graduated emergency response stage translates that score into one of several discrete response actions, described in Section 5.2, ranging from silent logging through to an automatic emergency alert with live location sharing.



A distinguishing feature of this architecture, shown as the feedback path in Fig. (1), is a personalized baseline and context profile that is continuously updated from each wearer's own resting-state physiology and typical movement patterns. Because resting heart rate, typical gait, and baseline skin conductance vary considerably between individuals, a fixed, population-wide threshold would necessarily be either too sensitive for some wearers or too insensitive for others; adapting the behavior-risk model's reference point to each wearer over time is intended to reduce this cross-user false-alarm variability, consistent with the sensor-placement and calibration sensitivity noted in the wearable HAR literature (Lara & Labrador, 2013; Banos et al., 2014).

Behavior-Risk Tier Scheme

Detection alone does not indicate how a system should respond, and a binary silent-or-SOS design risks either alert fatigue from frequent false positives or a dangerously narrow trigger condition that misses genuine but ambiguous distress. Fig. (2) sets out an illustrative four-tier behavior-risk scale intended to translate a continuous fused risk score into a graduated set of response actions of increasing urgency and increasing tolerance for interrupting the wearer.

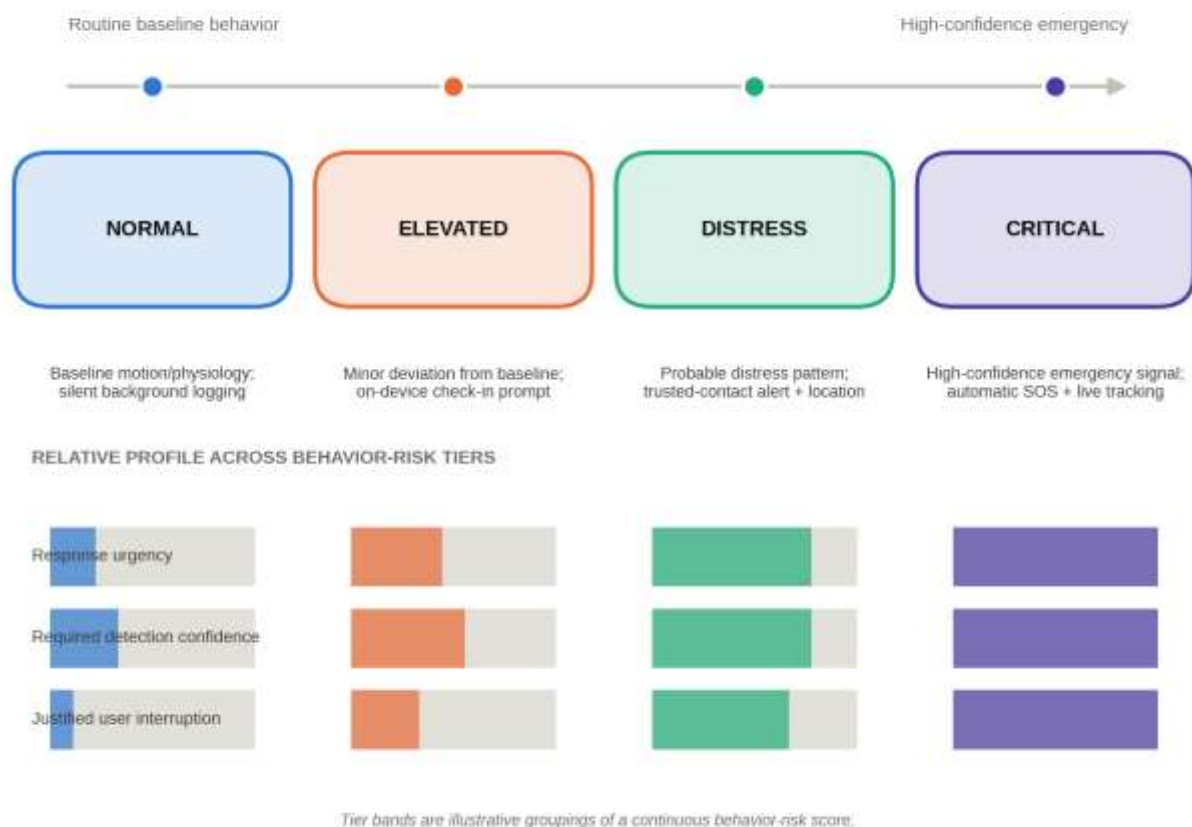


Fig. (2). An illustrative four-tier behavior-risk scale mapped to response urgency, required detection confidence, and the degree of user interruption each tier's response action justifies.

The specific tier boundaries shown in Fig. (2) are illustrative groupings of a continuous risk score rather than a fixed, universal policy; an actual deployment would calibrate these thresholds per user, following the personalized-baseline mechanism of Section 5.1, and potentially per deployment context. What the



four-tier structure is intended to capture, independent of the exact thresholds chosen, is the practical distinction between routine background monitoring that requires no user awareness at all, a low-friction check-in prompt that gives the wearer a simple way to stand down a borderline detection, and an escalating alert pathway that a system commits to only as its confidence in a genuine emergency increases.

Illustrative Deployment Scenario Design

To make the architecture of Section 5.1 concrete, an illustrative, entirely hypothetical deployment scenario was constructed rather than only citing published component-level metrics. The scenario assumes a cohort of wearers using the proposed device continuously over an extended period of ordinary daily activity, commuting, exercise, work, and leisure, interspersed with rare episodes of genuine or ambiguous distress. An illustrative distribution of monitored time across the four behavior-risk tiers of Fig. (2) was then constructed to demonstrate how a fused, continuously updated risk score would be expected to distribute in practice, and how this distribution would translate into the response workload the graduated-alert design of Section 5.2 places on the wearer and on trusted contacts; this distribution is presented and analyzed in Section 6.4, and its specific figures carry no empirical weight, a limitation discussed explicitly in Section 7.

VI. RESULTS

This section presents the paper's findings in two parts: a synthesis of literature-reported performance findings for each sensing modality, and the outputs of the illustrative deployment scenario described in Section 5.3. Fig. (3) previews both in graphical form: its left panel condenses the modality-performance comparison of Sections 6.1-6.2, and its right panel graphs the illustrative deployment-cohort tier composition of Section 6.4.

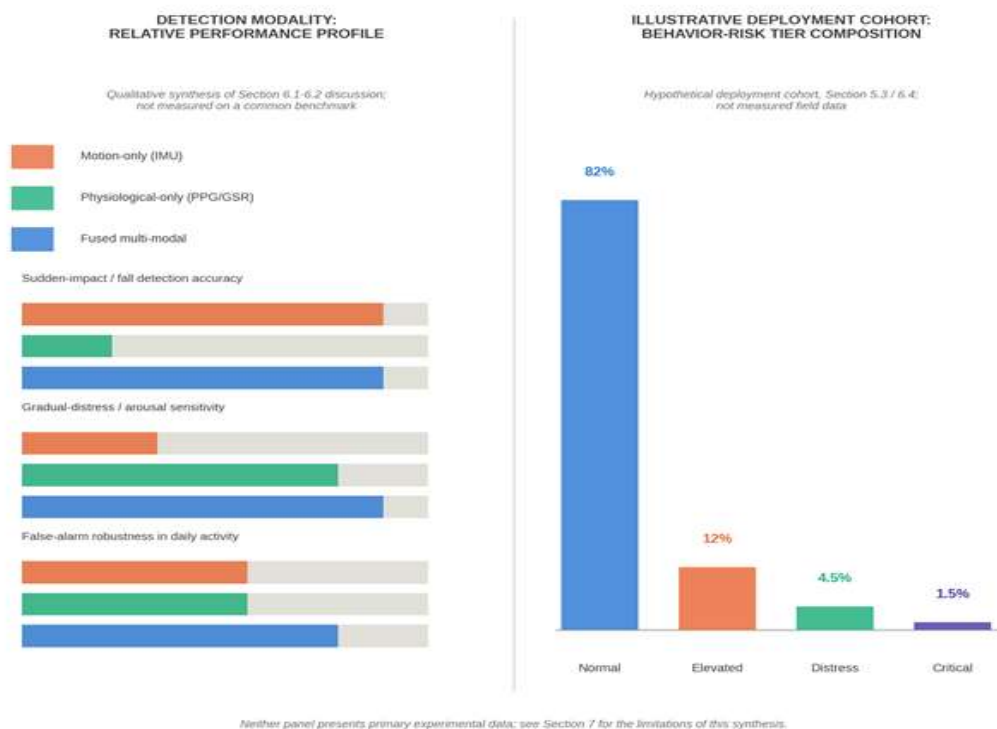


Fig. (3). Left: qualitative relative-performance synthesis comparing motion-only, physiological-only, and fused multi-modal anomaly detection across three dimensions discussed in Sections 6.1-6.3. Right:



illustrative behavior-risk tier composition of the hypothetical deployment cohort from Section 6.4 (Table 2). Neither panel is measured field data.

Motion-Based Anomaly Detection Performance

Reported motion-based recognition and anomaly-detection performance varies considerably with sensor placement, window length, and the specific comparison a given study makes, which makes a single summary number misleading. Table 1 compiles the approach and reported evaluation focus from several of the studies discussed in Section 2, without collapsing them to a single directly comparable figure, since the underlying datasets, sensors, and evaluation protocols are not uniform across studies.

Table 1. Approach and reported evaluation focus across wearable motion-based activity-recognition and anomaly/fall-detection studies discussed in Sections 2 and 6.1.

Study	Approach	Reported Evaluation Focus
Kwapisz et al. (2011)	Accelerometer-based activity recognition using simple classifiers on phone-embedded sensors	Reports classification accuracy across everyday activities using a benchmark cellphone-accelerometer dataset
Bulling et al. (2014)	Tutorial/synthesis of the standard body-worn inertial-sensor HAR pipeline (windowing, features, classification)	Establishes the pipeline against which later deep-learning HAR approaches are consistently compared
Lara & Labrador (2013)	Survey of wearable-sensor HAR methods and sensor-placement effects	Concludes sensor placement and preprocessing often matter more than classifier choice for accuracy
Banos et al. (2014)	Systematic study of segmentation-window-length impact on HAR accuracy	Reports an accuracy-versus-latency trade-off curve as a function of window length
Ordóñez & Roggen (2016)	CNN-LSTM (DeepConvLSTM) hybrid architecture on multimodal wearable-sensor data	Reports improved recognition of short, transitional, and irregular movements over feature-engineered baselines
Wang et al. (2019)	Survey of deep-learning HAR architectures (CNN, RNN, hybrid)	Synthesizes reported accuracy gains of deep methods over classical pipelines across multiple benchmark datasets
Igual et al. (2013)	Review of wearable fall-detection principles (threshold-based and machine-learning approaches)	Reports wide sensitivity/specificity variation depending on threshold tuning and sensor placement
Mubashir et al. (2013)	Survey of fall-detection approaches (wearable, ambient, vision-based)	Concludes wearable accelerometer-based approaches offer the best cost/mobility trade-off but risk false positives
Ravi et al. (2017)	On-node deep-learning sensor analytics for mobile and wearable devices	Reports feasibility of running compact deep models directly on wearable hardware within power/latency budgets

Note: figures and findings are as reported in the cited original studies; underlying datasets, sensors, populations, and evaluation protocols differ across studies and are not directly comparable.



Read across these studies, deep architectures that combine convolutional and recurrent processing consistently outperform classical feature-engineering pipelines on short, irregular, or transitional movements (Ordóñez & Roggen, 2016), the category of motion most relevant to detecting a struggle or sudden impact, while classical threshold-based methods remain attractive for their low power cost and interpretability in simpler fall-detection deployments (Igual et al., 2013). On-node inference studies (Ravi et al., 2017) indicate that this accuracy advantage need not come at the cost of cloud dependence, provided the deployed model is compact enough for the target wearable hardware.

Physiological Stress and Arousal Detection Performance

Physiological arousal detection reliably distinguishes objectively more stressful conditions from calmer baseline conditions in ambulatory, real-world settings, as demonstrated in the driving-stress literature (Healey & Picard, 2005), and the introduction of a standardized, wearable-first multimodal stress-and-affect dataset has begun to support more direct comparison across modeling approaches for this task (Schmidt et al., 2018). At the same time, physiological arousal is a comparatively non-specific signal: elevated heart rate and skin conductance accompany exercise, caffeine, ordinary excitement, and many other everyday states in addition to fear, a specificity limitation this paper's proposed architecture addresses only partially, through fusion with motion evidence and per-user baseline calibration, rather than resolving outright.

Multi-Modal Fusion Considerations

No study reviewed in Section 2 evaluates motion-based and physiological-based anomaly detection jointly on the same wearer for a personal-safety application specifically; the general anomaly-detection literature's framing of the sensitivity-versus-false-positive trade-off (Chandola et al., 2009) and the deep time-series anomaly-detection methods developed for other domains (Malhotra et al., 2016; Munir et al., 2019) are, in this paper's assessment, directly transferable to a fused motion-and-physiology risk score, but that transfer has not itself been empirically validated in the published literature. Ambient-audio anomaly cues (Rabaoui et al., 2008) represent a third, largely unexplored modality that could plausibly be fused alongside motion and physiology, particularly for detecting a vocalized distress signal that neither of the other two modalities would capture.

Illustrative Deployment Scenario Findings

Table 2 reports how the behavior-risk tiers of Fig. (2) translate into an illustrative distribution of monitored time under the scenario design of Section 5.3; the specific shares and response actions are hypothetical and are intended only to show how a graduated, tiered response reshapes the alert workload relative to a binary manually triggered design, not to predict the actual tier distribution any real wearer or deployment would produce. The right panel of Fig. (3) graphs the same illustrative distribution for a quicker visual comparison across tiers.

Table 2. Illustrative behavior-risk tier composition and graduated response workflow for the hypothetical wearable-deployment cohort of Section 5.3 / 6.4.

Behavior-Risk Tier	Illustrative Share of Monitored Time	Typical Response Action	Notified Party
Normal	~82%	Silent background logging; no user-facing alert	None
Elevated	~12%	On-device check-in prompt with stand-down option	Wearer only



Behavior-Risk Tier	Illustrative Share of Monitored Time	Typical Response Action	Notified Party
Distress	~4.5%	Alert to trusted contacts with live location share	Trusted contacts
Critical	~1.5%	Automatic SOS with continuous location tracking	Emergency contacts / services

Note: entirely illustrative figures intended to demonstrate how a graduated response reshapes the alert workload relative to a binary manually triggered design; not derived from a field trial and not a predicted tier distribution for any real wearer or population.

Under this illustration, the overwhelming majority of monitored time falls in the Normal tier, requiring no user-facing interruption at all, while the Elevated tier's low-friction check-in prompt is designed to resolve most borderline detections without escalating further; only the small Distress and Critical shares are expected to reach a trusted contact or trigger an automatic emergency alert. This graduated structure is intended to keep the false-alarm burden on the wearer and their trusted contacts low in ordinary daily use, while still providing an automatic, non-manually-triggered escalation pathway for the rare episodes that matter most.

VII. DISCUSSION

The complementary sensitivity gap between motion-only and physiological-only detection identified in Sections 6.1-6.3 is, in the assessment of this paper, the most important interpretive caveat for any system relying on a single sensing modality: a motion-only design can miss a slower-building threatening encounter that never produces an unambiguous impact signature, and a physiological-only design can miss a sudden physical assault if the wearer's autonomic response lags the event or is masked by concurrent physical exertion. Neither modality reviewed in Section 2 was evaluated as part of a fused pipeline against the other for a personal-safety application, which is itself evidence for the fusion gap identified in Section 3 rather than a settled question the literature has already answered.

Framed in terms of alert design, the case for a graduated, tiered response over a binary manually triggered alert rests on a similar logic to the biennial-versus-continuous monitoring trade-off documented in other structural-monitoring domains: a manually triggered design places the entire detection burden on a wearer who may be unable to act, while a purely automatic, low-threshold design risks enough false alarms to erode user trust and, eventually, adoption itself. The behavior-risk tier scheme of Section 5.2 is this paper's proposed resolution, reserving the most disruptive response, an automatic alert with live location sharing, for the tier at which detection confidence is highest, while still logging and, at the Elevated tier, surfacing lower-confidence anomalies for the wearer's own awareness. Adopting this kind of always-on, behavior-aware pipeline also raises considerations beyond detection accuracy alone. Continuous physiological and location monitoring is inherently privacy-sensitive, which motivates the on-node, edge-computing emphasis of this paper's proposed architecture (Ravi et al., 2017; Warden & Situnayake, 2019): performing anomaly inference directly on the wearable, rather than continuously streaming raw motion, physiological, and location data to a remote server, limits the data that must leave the device at all under normal, non-alert conditions. Battery life is an equally practical constraint; a safety wearable that requires daily or twice-daily charging is one a wearer may simply stop wearing, which would defeat the system's purpose regardless of its detection accuracy, underscoring why the energy-efficiency literature (Rault et al., 2014) is as relevant to this application as the modeling literature itself.



Challenges and Limitations

- Physiological non-specificity: elevated heart rate and skin conductance accompany exercise and ordinary excitement as well as fear, requiring careful fusion and per-user calibration to avoid frequent false positives (Healey & Picard, 2005).
- Motion false-alarm sources: vigorous ordinary activity, such as sports, dancing, or an abrupt fall unrelated to any threat, can resemble the motion signature this paper's architecture is designed to flag (Igual et al., 2013; Mubashir et al., 2013).
- On-device inference and battery constraints: continuous multi-modal sensing and always-on inference must fit within a wearable's power budget without requiring impractically frequent recharging (Rault et al., 2014; Warden & Situnayake, 2019).
- Location accuracy during emergencies: assisted-GPS positioning degrades in indoor, urban-canyon, and obstructed environments, precisely the settings in which an emergency alert's location payload may matter most (Zandbergen & Barbeau, 2011).
- Fusion validation gap: because motion-based and physiological-based anomaly detection are rarely evaluated together for this specific application, there is limited published evidence on how to weight the two modalities' evidence when they disagree.
- Adversarial and edge-case robustness: a system designed around a wearer's own baseline behavior has not, in the reviewed literature, been evaluated against an assailant's awareness of the device or deliberate attempts to prevent an alert from being sent.

None of these challenges is individually disqualifying, and several are shared with the adjacent fields of wearable fall detection and affective computing, which have continued to mature despite comparable limitations. What distinguishes a personal-safety application specifically is the asymmetric cost of a missed detection relative to a false alarm, which argues for the graduated, confidence-tiered response of Section 5.2 rather than a single fixed operating point. It is equally important to be explicit about the limitations of this paper's own synthesis method: comparing detection accuracy and false-alarm behavior across studies that use different sensors, populations, and evaluation protocols, as Table 1 does deliberately without collapsing the comparison to a single number, is inherently weaker evidence than a single controlled experiment evaluating multiple methods on one shared cohort under one shared protocol. The illustrative scenario in Section 6.4 makes the same trade-off in the opposite direction: it demonstrates a graduated-response logic clearly, but its specific numbers are invented for illustration and carry no empirical weight. Readers should treat this paper's contribution as a structured map of what the published literature currently supports and where it is silent, rather than as a validated performance guarantee for the four-stage pipeline as a whole.

VIII. CONCLUSION

This paper has argued that AI-driven personal safety for wearable devices is best understood as a four-stage, multi-modal pipeline, data acquisition, behavior and anomaly modeling, risk fusion and tier classification, and graduated emergency response, rather than as an extension of a manually triggered panic-button application alone. The literature reviewed in Section 2 and synthesized in Section 6 shows that each contributing capability individually has a credible, reasonably well-validated technical basis, motion-based anomaly and fall detection with a substantial research history, physiological stress and affect sensing grounded in affective computing, and low-power on-node inference now demonstrated on wearable-class hardware, but that these capabilities are rarely fused and evaluated together for a personal-safety application specifically, the gap identified in Section 3. Closing that gap, and translating the resulting fused behavior-risk score into a graduated response that avoids both alert fatigue and missed detections, is the practical work that stands between the capabilities surveyed in this paper and a deployable, trustworthy safety wearable. For the illustrative deployment cohort analyzed in Section 6.4, even a partial realization of this integration, reliable motion-based anomaly detection combined



with basic physiological screening, would already extend meaningful protection to the moments when a wearer is unable to manually trigger an alert at all.

FUTURE SCOPE

The gaps identified in Sections 3 and 7 point toward a research agenda focused less on incremental accuracy gains for either sensing modality alone and more on integration, validation, and responsible deployment. Five directions in particular stand out from this review.

- Ethically governed, consent-based field studies that jointly collect motion, physiological, and location data from volunteer participants under simulated or supervised distress scenarios, to validate the fused behavior-risk model against something closer to ground truth than the largely separate evaluation traditions surveyed in Section 6.
- Fusion architectures explicitly validated against confirmed distress events, to establish how much additional sensitivity a fused motion-and-physiology score provides over either modality used alone, and to quantify the false-alarm reduction the personalized-baseline mechanism of Section 5.1 actually achieves.
- On-device audio distress-cue detection, extending the ambient-audio anomaly-detection techniques used in adjacent surveillance contexts (Rabaoui et al., 2008) into a privacy-preserving, on-node scream or distress-vocalization classifier that never transmits raw audio off the device under normal conditions.
- Federated, privacy-preserving model improvement across many wearers' devices, allowing a shared anomaly-detection model to benefit from a larger and more diverse population of everyday movement and physiological patterns without any individual's raw sensor data leaving their own device.
- Standardized, wearable-first benchmark protocols specifically for personal-safety anomaly detection, analogous to the WESAD dataset's role for wearable stress and affect detection (Schmidt et al., 2018), to enable direct, reproducible comparison across future fusion architectures rather than the largely incomparable, modality-specific benchmarks available today.

REFERENCES

1. Banos, O., Galvez, J. M., Damas, M., Pomares, H., & Rojas, I. (2014). Window size impact in human activity recognition. *Sensors*, 14(4), 6474-6499. <https://doi.org/10.3390/s140406474>
2. Bulling, A., Blanke, U., & Schiele, B. (2014). A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys*, 46(3), 1-33. <https://doi.org/10.1145/2499621>
3. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, 41(3), 1-58. <https://doi.org/10.1145/1541880.1541882>
4. Healey, J. A., & Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems*, 6(2), 156-166. <https://doi.org/10.1109/TITS.2005.848368>
5. Igual, R., Medrano, C., & Plaza, I. (2013). Challenges, issues and trends in fall detection systems. *BioMedical Engineering OnLine*, 12, 66. <https://doi.org/10.1186/1475-925X-12-66>
6. Islam, S. M. R., Kwak, D., Kabir, M. H., Hossain, M., & Kwak, K. S. (2015). The internet of things for health care: A comprehensive survey. *IEEE Access*, 3, 678-708. <https://doi.org/10.1109/ACCESS.2015.2437951>
7. Kwapisz, J. R., Weiss, G. M., & Moore, S. A. (2011). Activity recognition using cell phone accelerometers. *ACM SIGKDD Explorations Newsletter*, 12(2), 74-82. <https://doi.org/10.1145/1964897.1964918>
8. Lara, O. D., & Labrador, M. A. (2013). A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys & Tutorials*, 15(3), 1192-1209. <https://doi.org/10.1109/SURV.2012.110112.00192>
9. Malhotra, P., Ramakrishnan, A., Anand, G., Vig, L., Agarwal, P., & Shroff, G. (2016). LSTM-based encoder-decoder for multi-sensor anomaly detection. *arXiv*. <https://arxiv.org/abs/1607.00148>



10. Mubashir, M., Shao, L., & Seed, L. (2013). A survey on fall detection: Principles and approaches. *Neurocomputing*, 100, 144-152. <https://doi.org/10.1016/j.neucom.2011.09.037>
11. Munir, M., Siddiqui, S. A., Dengel, A., & Ahmed, S. (2019). DeepAnT: A deep learning approach for unsupervised anomaly detection in time series. *IEEE Access*, 7, 1991-2005. <https://doi.org/10.1109/ACCESS.2018.2886457>
12. Ordóñez, F. J., & Roggen, D. (2016). Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. *Sensors*, 16(1), 115. <https://doi.org/10.3390/s16010115>
13. Picard, R. W. (1997). *Affective computing*. MIT Press.
14. Rabaoui, A., Davy, M., Rossignol, S., & Ellouze, N. (2008). Using one-class SVMs and wavelets for audio surveillance. *IEEE Transactions on Information Forensics and Security*, 3(4), 763-775. <https://doi.org/10.1109/TIFS.2008.2008216>
15. Rault, T., Bouabdallah, A., & Challal, Y. (2014). Energy efficiency in wireless sensor networks: A top-down survey. *Computer Networks*, 67, 104-122. <https://doi.org/10.1016/j.comnet.2014.03.027>
16. Ravi, D., Wong, C., Lo, B., & Yang, G. Z. (2017). A deep learning approach to on-node sensor data analytics for mobile or wearable devices. *IEEE Journal of Biomedical and Health Informatics*, 21(1), 56-64. <https://doi.org/10.1109/JBHI.2016.2633287>
17. Schmidt, P., Reiss, A., Duerichen, R., Marberger, C., & Van Laerhoven, K. (2018). Introducing WESAD, a multimodal dataset for wearable stress and affect detection. *Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI)*, 400-408. <https://doi.org/10.1145/3242969.3242985>
18. Wang, J., Chen, Y., Hao, S., Peng, X., & Hu, L. (2019). Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters*, 119, 3-11. <https://doi.org/10.1016/j.patrec.2018.02.010>
19. Warden, P., & Situnayake, D. (2019). *TinyML: Machine learning with TensorFlow Lite on Arduino and ultra-low-power microcontrollers*. O'Reilly Media.
20. Yarrabothu, R. S., & Thota, B. (2015). Abhaya: An Android app for the safety of women. In *2015 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)* (pp. 1-4). IEEE. <https://doi.org/10.1109/ANTS.2015.7413660>
21. Zandbergen, P. A., & Barbeau, S. J. (2011). Positional accuracy of assisted GPS data from high-sensitivity GPS-enabled mobile phones. *Journal of Navigation*, 64(3), 381-399. <https://doi.org/10.1017/S0373463311000051>