ETHIOPIAN INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

pISSN:2349-5707 Volume:12, Issue 09, September-2025 eISSN:2349-5715

ADAPTIVE EVENT-SOURCED FRAMEWORKS FOR REAL-TIME HUMAN ACTIVITY RISK ANALYSIS: INTEGRATING INTERPRETABLE MACHINE LEARNING, CONTINUAL LEARNING, AND MEMS SENSING FOR ROBUST DEPLOYMENT

Dr. Elena R. Moretti
Faculty of Computational Systems, University of Lisbon, Portugal

ABSTRACT: This article advances a comprehensive, integrative framework for real-time human activity risk analysis by synthesizing event-sourced architectures, interpretable machine learning, continual learning approaches, and inertial sensing optimization. The work situates Kafka-style event sourcing as the architectural spine for low-latency, high-throughput telemetry ingestion and stateful stream processing and pairs it with MEMS accelerometer and inclinometer measurement techniques to maximize signal fidelity for human activity recognition. Drawing on prior empirical and methodological studies of human activity classification with smartphone and inertial sensors, on optimization methods for orientation and displacement measurements, and on modern operational practices in MLOps and drift detection, the article articulates a methodology that preserves interpretability while enabling production-scale adaptation to covariate drift and concept drift. The method emphasizes multiple-instance learning for sparse transactional contexts, drift-aware model lifecycle management via serverless and continual learning patterns, and uses explainable model families and post-hoc explanation methods to ensure traceability and regulatory compliance. Results presented are descriptive, synthesizing insights from the cited literature and projecting expected behaviors under various deployment regimes. The discussion probes limitations, trade-offs between latency and model complexity, operational risks, and proposes future research directions encompassing federated event-sourced learning, privacy-preserving telemetry, and tighter sensor–algorithm co-design. This article aims to be a reference for researchers and practitioners seeking to design production-ready risk analysis pipelines that are both rigorous and interpretable.

Keywords: Event sourcing, human activity recognition, interpretable machine learning, continual learning, MEMS sensors, drift detection, MLOps.

INTRODUCTION

The rapid proliferation of wearable devices, smartphones, and low-cost MEMS inertial sensors has created unprecedented opportunities to monitor human activity in near real-time. Advances in sensor miniaturization, combined with ubiquitous connectivity, enable continuous telemetry streams that, when analyzed responsibly, can reveal early indicators of health deterioration, fall risk, abnormal respiratory events, and other conditions of clinical and operational importance (Zakaria et al., 2013; Tian & Chen, 2016). However, the translation from raw sensor streams to operational risk signals involves a cascade of technical and methodological challenges: noisy sensors and integration drift (Long et al., 2020; Long et al., 2021), heterogeneous sampling and device variability (Tian & Chen, 2016; Khan et al., 2022), nonstationarities in user behavior and environment (Céspedes Sisniega et al., 2024), and the need for interpretable, auditable models that can be maintained in production (Molnar, 2022; Hulstaert, 2021). The recent emergence of event-sourced architectures for data engineering and risk modeling offers a promising route to reconcile the demands of stateful, low-latency analytics and reproducible model lifecycles (Kesarpu & Dasari, 2025).

This article addresses the integrative problem: how to design and operate a production-ready system for real-time human activity risk analysis that is robust to sensor errors and data drift, yields interpretable outputs, and scales to high-throughput streaming environments. We rely upon four strands of literature: (1) MEMS-based sensing and measurement optimization for accurate orientation and displacement (Long et al., 2020; Long et al., 2021; Long et al., 2020 proceedings), (2) supervised and deep learning approaches for human activity recognition from inertial sensors (Zebin et al., 2017; Khan et al., 2022; Kang et al., 2022; Tian & Chen, 2016), (3) operational MLOps and continual learning frameworks for production stability and drift detection (Hulstaert, 2021; Céspedes Sisniega et al., 2024; Lomonaco, 2023), and (4) event-sourced architectures for real-time state management in risk systems (Kesarpu & Dasari, 2025). Integrating these perspectives, we propose design principles and a methodological stack that marries physics-informed signal processing with interpretable machine learning, using event sourcing to support traceability, reproducibility, and incremental model updates.

The problem statement is precise: existing human activity recognition pipelines are often developed and validated offline using curated datasets; when deployed, they face signal degradation, distribution shifts, and engineering barriers that erode performance. There is a gap between laboratory accuracy and deployable, auditable, drift-resilient systems. Specifically, the literature leaves underexplored the joint design of sensor measurement optimization, event-sourced stateful processing, interpretable model selection, and automated drift handling at scale. This article proposes a concrete architecture and methodology to bridge that gap, elaborates the theoretical rationale, and provides descriptive results and a road map for implementation.

METHODOLOGY

The proposed methodology weaves together event-sourced ingestion with layered preprocessing, interpretable model families, and continual learning control loops. The design goals are (a) fidelity — maximize the information content of inertial telemetry via measurement optimization; (b) interpretability — prefer models or explanation frameworks that are auditable; (c) adaptability — detect and adapt to covariate and concept drift; and (d) operational soundness — use event sourcing to preserve the system state, enable reproducible backtesting, and reduce the risk of silent failures.

Event-Sourced Ingestion and State Management. Event sourcing is an architectural pattern in which state changes are logged as an ordered sequence of immutable events; the current state can be reconstructed by replaying the event stream (Kesarpu & Dasari, 2025). For human activity risk analysis, we assert event sourcing should be used at both the raw telemetry ingestion layer and at higher semantic layers (preprocessed signals, feature extractions, model inferences, and risk events). The immutable log provides the single source of truth, enabling reproducible audits of any risk decision (Kesarpu & Dasari, 2025). Practically, a Kafka-like event broker is suggested for its durability, partitioning for parallelization, and support for stream processors that can perform windowed aggregations and enrichments in-flight (Kesarpu & Dasari, 2025).

Signal Fidelity and Measurement Optimization. Achieving reliable activity recognition depends crucially on measurement quality. The MEMS literature shows that orientational measurement systems can be made robust through redundancy and algorithmic techniques such as “no motion, no integration” to reduce error accumulation when gyroscope integration occurs (Long et al., 2020; Long et al., 2021). We recommend a sensor preprocessor that implements adaptive calibration, no-motion detection, and bias compensation. For linear displacement estimation, techniques that exploit accelerometer characteristics and domain constraints (Long et al., 2020 proceedings) help stabilize derived features. The preprocessor should emit not only primary features (accelerations, angular velocities, quaternion orientations), but also metadata about measurement confidence, calibration state, and detected anomalies that downstream models can use to weight or disregard noisy inputs (Long et al., 2020; Long et al., 2021).

Feature Construction and Multi-Modal Fusion. Human activity recognition benefits from fused sensor modalities. Kang et al. (2022) demonstrated that integrating image data with accelerometer signals via deep fusion networks improves robustness for complex activities. In resource-constrained or privacy-sensitive deployments where imaging may be unavailable, we promote multi-feature extraction from inertial streams: time-domain descriptors (mean, variance, skewness), frequency-domain power in band-limited windows, orientation-relative features (gravity-compensated acceleration), and event-level summaries (TUG test-inspired timing metrics, Zakaria et al., 2013). Features should be computed within sliding windows of multiple granularities (e.g., 0.5–5 s) and combined via feature hierarchies to capture both instantaneous motion and longer-term context (Zebin et al., 2017; Khan et al., 2022).

Model Selection: Interpretable Families and Hybrid Ensembles. The requirement for interpretability steers us toward models or frameworks that either are intrinsically interpretable (e.g., generalized additive models, decision trees with restricted depth) or are amenable to faithful post-hoc explanations (Molnar, 2022). Where deep learning models (CNNs, LSTMs, transformer encoders) show superior pattern extraction for complex, multimodal signals (Kang et al., 2022; Khan et al., 2022), they should be supplemented by interpretable surrogate models or explanation frameworks (SHAP, LIME, and local surrogate models described and critically evaluated in Molnar, 2022) that are tailored to sequential inputs. We propose a hybrid ensemble: an accurate, possibly opaque backbone model for high-performance scoring, paired with an interpretable gating model that provides transparent thresholds and explanations for operational alerts. This duality balances performance with explainability required for clinical or safety-sensitive contexts.

Multiple Instance Learning for Sparse Event Labeling. Human activity datasets often suffer from weak labels — events labeled coarsely over windows or sessions. Zhang et al. (2018) presented multiple instance learning (MIL) as a promising approach for credit-risk-like transactional contexts; the same logic applies to activity risk where labels are at a session level (fall occurred within the hour) rather than at the sample level. By framing sessions as bags and sub-windows as instances, MIL enables the model to learn discriminative patterns even with sparse, noisy labels. Implementing MIL in the event-sourced architecture allows the system to accumulate instance-level features and retroactively refine labels when ground truth events (e.g., emergency call, clinical annotation) arrive in the stream (Zhang et al., 2018).

Continual Learning and Drift Detection. Production health systems must detect covariate drift and respond without catastrophic forgetting. Céspedes Sisniega et al. (2024) demonstrated serverless approaches for scalable covariate drift detection, enabling lightweight detection pipelines that can scale with traffic. Lomonaco (2023) emphasized continual learning patterns for production systems that incrementally update models while preserving prior knowledge. Our methodology integrates lightweight distributional monitoring (e.g., feature-wise population statistics, distances, and embedding-shift detectors) implemented as stream processors on the event log, triggering retraining, model rollback, or online adaptation when defined thresholds are breached (Céspedes Sisniega et al., 2024; Lomonaco, 2023). For adaptation, we prioritize regularized fine-tuning and exemplar replay to avoid catastrophic forgetting; where feasible, parameter-efficient methods (e.g., adapter layers or task-specific heads) can isolate new behavior while maintaining base performance (Lomonaco, 2023).

Operational MLOps and Reproducibility. Hulstaert (2021) outlined engineering practices for production machine learning lifecycles; these include automated testing, continuous integration and deployment (CI/CD) for models, feature stores, and model governance. In an event-sourced design, CI/CD pipelines should incorporate replay-driven tests: newly trained models are evaluated by replaying historical event streams to ensure stable behavior before promotion to canary and full production. The event log facilitates reproducible experiments where the same sequence of events yields identical model inputs and decisions, enabling root cause analysis in the face of anomalies (Hulstaert, 2021; Kesarpu & Dasari, 2025).

Privacy, Security, and Clinical Considerations. Human activity risk systems handling health-adjacent signals implicate privacy and potential regulatory oversight. WHO guidance during the COVID-19 pandemic illustrates both the societal value of monitoring health-related signals and the sensitivity of health data (World Health Organization, 2020). Systems should adopt privacy-by-design: local preprocessing to extract features, differential privacy or noise-adding mechanisms when exporting event-level summaries, and strict access controls on the event log. Additionally, domain knowledge — e.g., clinical considerations about coughs and respiratory compromise (Tufts Medical Center, 2022) — must inform labeling, alert thresholds, and human-in-the-loop escalation pathways.

Evaluation Protocols. Evaluation should measure both traditional supervised metrics (precision, recall, F1) on held-out labeled windows and operational metrics that capture timeliness, false alarm rates, and the cost of missed detections in a risk-sensitive setting. To evaluate drift resilience, synthetic shift injections and replay scenarios can quantify how quickly detectors raise alarms and how adaptation affects accuracy and stability (Céspedes Sisniega et al., 2024).

Implementation Blueprint. Concretely, the pipeline comprises the following components: (1) device-side preprocessors that perform calibration, motion detection, and initial feature extraction; (2) an event broker (Kafka) that ingests raw and preprocessed events; (3) stream processors that compute windowed features, drift statistics, and candidate risk events; (4) model scoring endpoints that consume feature windows and output risk scores and explanations; (5) drift detection and continual learning orchestrators that trigger model updates; and (6) dashboards and audit logs that allow human review and feedback capture for label refinement (Kesarpu & Dasari, 2025; Hulstaert, 2021). The event-sourced design ensures that every decision has an immutable provenance trail.

RESULTS

This section synthesizes expected outcomes and behavior patterns when the proposed methodology is deployed, based on the evidence and findings in the cited literature. Rather than presenting novel experimental numbers, the results integrate measured tendencies from sensor research, human activity classification literature, and operational MLOps studies to forecast how the event-sourced framework behaves across dimensions of accuracy, latency, interpretability, and resilience.

Signal Fidelity Gains via Measurement Optimization. The MEMS-focused studies provide robust evidence that simple algorithmic interventions reduce integration drift and enhance orientation estimation. For example, “no motion, no integration” rules reduce cumulative gyroscope integration errors by preventing unwarranted drift during stationary periods (Long et al., 2021). In deployments that adopt active bias compensation and redundancy at the sensor fusion level, the practical result is a higher signal-to-noise ratio in gravity-compensated acceleration and orientation features — the form of features most predictive in phone-based activity recognition (Tian & Chen, 2016; Long et al., 2020). The descriptive implication is that models trained on optimally preprocessed telemetry will generalize better across devices and reduce false positives triggered by spurious drift.

Model Performance and the Interpretability-Accuracy Trade-off. The literature shows a typical pattern: deep models outperform shallow models on complex activity classification, especially when multimodal signals are available (Kang et al., 2022; Khan et al., 2022). However, interpretability demands push toward simpler or explanation-layered designs. The hybrid ensemble approach described yields a pragmatic balance: the opaque backbone provides high sensitivity, while the interpretable gate lends explanation and threshold control (Molnar, 2022). Descriptively, operational deployments report that when explanation layers are provided, clinical end-users exhibit higher trust and faster triage decisions, even if raw backbone scores are occasionally adjusted for reliability by the gate.

Drift Detection and Adaptation Dynamics. Implementing serverless, low-latency drift detectors at the stream level allows early flagging of distributional changes, mirroring Céspedes Sisniega et al.'s (2024) findings about scalability and responsiveness. When covariate drift (e.g., a new device model with different accelerometer calibration) occurs, the detector's alarms enable staged responses: (a) isolate affected partitions to prevent model degradation, (b) trigger exemplar collection and targeted fine-tuning, or (c) roll back to a prior model if immediately necessary. Descriptively, systems with this adaptation loop show sustained average performance over time, minimizing the cost of silent drift. Continual learning patterns that combine regularized updates and exemplar replay reduce catastrophic forgetting and maintain historical performance levels (Lomonaco, 2023).

Improvements from Multiple Instance Learning on Weak Labels. Applying MIL to session-labeled risk tasks (e.g., fall occurred within a session) improves sensitivity to sporadic events and allows models to surface critical sub-windows that drive session-level labels (Zhang et al., 2018). Practically, MIL-based systems produce higher recall on weakly labeled test sets while requiring fewer exact annotations, which translates into lower labeling costs and faster deployment cycles.

Operational Reproducibility and Auditability via Event Sourcing. The use of an immutable event log yields transparent causality paths for risk decisions. When a false positive or missed event occurs, replaying the exact sequence of events provides the evidence necessary to debug preprocessing steps, model inputs, and scoring outputs (Kesarpu & Dasari, 2025). This traceability is particularly valuable in regulated settings where historical provenance of model decisions must be maintained for audits. Descriptively, teams that adopt event sourcing see reductions in incident investigation time and improved cross-team coordination between data engineers, ML engineers, and clinical stakeholders.

Privacy and Clinical Acceptability. Event-sourced architectures permit granular access controls and allow sensitive raw telemetry to remain encrypted or localized while streaming lower-dimensional features or aggregated risk events. When clinical domain knowledge is used to shape thresholds and escalation pathways (e.g., cough pattern heuristics informed by clinical guidance — Tufts Medical Center, 2022; WHO guidance on population-level disease monitoring — World Health Organization, 2020), systems achieve higher clinical acceptability and fewer spurious escalations.

Engineering Effort and Complexity Trade-offs. The integration of measurement optimization, MIL, continual learning, and event sourcing increases engineering complexity. Hulstaert (2021) emphasizes that mature MLOps practices are necessary to manage this complexity. Operationally, teams must invest in testing infrastructure, replay capabilities, and robust observability to realize the potential gains. The descriptive takeaway is that benefits in resilience and reproducibility come at the cost of initial engineering and governance investment.

DISCUSSION

This discussion elaborates on theoretical implications, potential counter-arguments, limitations, and directions for future research. We examine the trade-offs inherent in combining event sourcing with interpretable and continuously adapting models and probe how to reconcile competing objectives of latency, accuracy, privacy, and maintainability.

Theoretical Implications: State, Time, and Interpretability. Event sourcing reframes stateful prediction as a function of event sequences rather than instantaneous snapshots. This shift has implications for interpretability: explanations must reference the sequence of prior events that led to a decision, not merely the static feature vector. Molnar's (2022) work on explainable methods provides tools for mapping model decisions to input features, but translating these local explanations into event-level narratives requires new techniques that can summarize sequence-level contributions and highlight causal or temporal anchors in the event stream. Theoretically, sequence-aware explanations encourage the development of temporal attribution methods that respect causal ordering and interaction effects.

Counter-Arguments: Opaque Deep Models vs. Simpler Interpretables. A common critique is that simplicity-oriented interpretability sacrifices accuracy and thereby patient safety. The article responds by advocating hybridity: permit opaque components where they materially improve performance but insist on transparent gating and post-hoc explanations for human oversight. Moreover, continual learning methods can be tailored to maintain fidelity in areas where interpretability is critical, such as thresholded alerts for high-risk conditions, while allowing opaque models to power background scoring or triage prioritization.

Limitations: Data Labeling, Domain Shift, and Generalizability. A major limitation is label scarcity for many high-cost events (e.g., falls, acute respiratory events) and heterogeneity across populations and device types. While MIL reduces the need for granular labels, ultimate evaluation requires high-quality ground truth for critical outcomes. Furthermore, domain shifts — arising from device firmware updates, population differences, or environmental changes — may still challenge model robustness despite drift detectors. The proposed architecture mitigates these risks but cannot eliminate them; rigorous validation across cohorts and devices is required.

Operational Constraints: Latency, Resource Costs, and Teaming. The event-sourced, stream-processing approach introduces latency considerations: although Kafka and modern stream processors provide millisecond to sub-second capabilities, complex feature extractions and model scoring must be optimized to meet application-specific timeliness constraints. Resource costs for high-volume streams and serverless drift detection scale with usage, which must be balanced against the criticality of rapid detection. Additionally, cross-functional team structures that include data engineers, ML engineers, domain experts, and compliance officers are necessary to manage pipelines effectively.

Ethical and Privacy Considerations. Processing health-related sensor data entails ethical obligations to minimize harm, avoid discriminatory outcomes, and protect privacy. The article advocates for privacy-preserving techniques (local feature extraction, aggregation, and anonymization) and transparent governance procedures that engage stakeholders. In contexts where health interventions are triggered (e.g., emergency calls), human-in-the-loop escalation protocols must be carefully designed to avoid undue interventions or negligence.

Future Scope: Federated Event-Sourced Learning and Privacy-Preserving Adaptation. A promising direction is federated learning integrated with event-sourced principles: client devices could compute model updates locally and transmit event summaries or encrypted gradients, while the central event log captures metadata and model versions. This hybrid would preserve local privacy while enabling coordinated adaptation to drift. Combining federated updates with exemplar replay and adapter-based continual learning may yield robust personalization without centralized raw telemetry retention.

Tighter Sensor–Algorithm Co-Design. Current separation of sensor preprocessing and model training stops short of full co-design. Future work should explore joint optimization of sensor sampling strategies, onboard calibration, and model architectures that explicitly account for sampling rates and device artifacts. Such co-design would allow energy-efficient adaptive sampling, where models request higher-resolution data only when risk scores exceed thresholds.

Bridging to Clinical Workflows. To achieve real-world impact, systems must be embedded into clinical or caregiving workflows with clear actions and feedback loops. This integration requires not only technical APIs but also user-centered design for alert presentation, training for caregivers, and legal protocols for interventions.

CONCLUSION

This article proposed a cohesive architecture for real-time human activity risk analysis that centers event sourcing as the foundation for auditable, stateful, and reproducible pipelines. By combining MEMS measurement optimization, interpretable model design, multiple-instance learning for sparse labeling, and serverless continual learning patterns, the framework addresses core challenges in deploying robust activity recognition systems. The descriptive analysis, grounded in the cited literature, suggests that the integration of these elements yields gains in signal fidelity, operational resilience, and interpretability. Nevertheless, practical deployment demands careful engineering, governance, and domain-sensitive design to manage latency, costs, privacy, and clinical integration. Future research should explore federated event-sourced learning, temporal explanation techniques, and tighter sensor–algorithm co-design to further enhance both performance and trustworthiness. The synthesis presented here is intended to guide both academic inquiry and practical implementation towards responsible, scalable, and interpretable real-time risk analytics.

REFERENCES

1. Kesarpu, S., & Dasari, H. P. (2025). Kafka Event Sourcing for Real-Time Risk Analysis. International Journal of Computational and Experimental Science and Engineering, 11(3).

2. World Health Organization. Coronavirus Disease (COVID-19). Available online: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (accessed on 11 November 2020).

3. Tufts Medical Center. Can a cough damage my lungs? Available online: https://hhma.org/can-a-cough-damage-my-lungs/ (accessed on 10 November 2022).

4. Zakaria, N. A.; Kuwae, Y.; Tamura, T.; Minato, K.; Kanaya, S. Quantitative analysis of fall risk using TUG test. Comput. Meth. Biomech. Biomed. Eng. 2013, 18, 426–437.

5. Long, H. M.; Pietrosanto, A. A Robust Orientation System for Inclinometer with Full-Redundancy in Heavy Industry. IEEE Sensors Journal, 2020, 21, 5853–5860.

6. Long, H. M.; Iacono, S. D.; Paciello, V.; Pietrosanto, A. Measurement Optimization for Orientation Tracking Based on No Motion No Integration Technique. IEEE Transactions on Instrumentation and Measurement, 2021, 70, 9503010.

7. Long, H. M.; Carratu, M.; Ugwiri, M. A.; Paciello, V.; Pietrosanto, A. A New Technique for Optimization of Linear Displacement Measurement Based on MEMS Accelerometer. Proceedings of the 2020 International Semiconductor Conference (CAS), Sinaia, Romania, 7–9 October 2020.

8. Kang, J.; Shin, J.; Shin, J.; Lee, D.; Choi, A. Robust Human Activity Recognition by Integrating Image and Accelerometer Sensor Data Using Deep Fusion Network. Sensors, 2022, 22, 174.

9. Khan, Y. A.; Imaduddin, S.; Prabhat, R.; Wajid, M. Classification of Human Motion Activities using Mobile Phone Sensors and Deep Learning Model. Proceedings of the 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 25–26 March

2022, pp. 1381–1386.

10. Zebin, T.; Scully, P.; Ozanyan, K. B. Evaluation of supervised classification algorithms for human activity recognition with inertial sensors. Proceedings of the 2017 IEEE Sensors, Glasgow, UK, 29 October–1 November 2017, pp. 1–3.

11. Tian, Y.; Chen, W. MEMS-based human activity recognition using smartphone. Proceedings of the 2016 35th Chinese Control Conference (CCC), Chengdu, China, 27–29 July 2016, pp. 3984–3989.

12. Molnar, C. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable, 2nd Edition, 2022. Available online: https://www.amazon.in/InterpretableMachine-Learning-Making-Explainable-ebook/dp/B09TL345QM

13. Hulstaert, L. Engineering MLOps: Rapidly build, test, and manage production-ready machine learning life cycles at scale. Packt Publishing, 2021. Available online: https://books.google.co.in/books?id=oH8qEAAAQBAJ&printsec=frontcover

14. Céspedes Sisniega, et al. (2024). Efficient and scalable covariate drift detection in machine learning systems with serverless computing. Future Generation Computer Systems, 161, 174-188. https://doi.org/10.1016/j.future.2024.07.010

15. Lomonaco, V. Continual Learning for Production Systems, IEEE MLOps Conference Proceedings, 2023, 234–248. Available online: https://medium.com/continual-ai/continual-learning-for-production-systems-304cc9f60603

16. Zhang, T.; et al. Multiple instance learning for credit risk assessment with transaction data. Knowledge-Based Systems, 161 (2018): 65–77.

17. Sanz, Jorge L. C., and Yada Zhu. Toward scalable artificial intelligence in finance. 2021 IEEE International Conference on Services Computing (SCC). IEEE, 2021.