Distributed Edge Computing Architecture for Fall Detection in Senior Care Facilities

Ananya M., Karthik S., Meherosh F.

Department of Computer Science, Regional Institute of Technology, Coimbatore, Tamil Nadu; Centre for Applied Systems, Metropolitan Technical University, Bhopal, Madhya Pradesh

Abstract

The integration of high-reliability monitoring systems in senior care facilities has become a critical engineering priority. This paper explores the development of an edge-cloud orchestrated framework designed for real-time fall detection. Traditional centralized systems often face latency bottlenecks and bandwidth constraints that can delay emergency responses during critical incidents. Our research proposes a decentralized architecture where initial data processing occurs at the "edge"—utilizing on-site gateways and wearable sensors—to enable near-instantaneous anomaly detection. The framework was implemented across multiple testbed facilities using a combination of tri-axial accelerometers and infrared occupancy sensors. This study evaluates the trade-off between local processing power and central data storage, focusing on reducing false positives while optimizing the battery life of low-power devices. Findings demonstrate that edge-based inference reduces response latency significantly compared to conventional architectures. This work provides a technical blueprint for smart healthcare infrastructure that prioritizes localized intelligence and resident privacy.

Keywords Edge Computing, Internet of Medical Things (IoMT), Gerontechnology, Fall Detection, Low-Latency Networking, Smart Sensors, Cloud Orchestration, Wearable Devices, Predictive Analytics, Healthcare Infrastructure.

1. Introduction

The rapid evolution of sensor networks and distributed computing has catalyzed a fundamental shift in the design of safety-critical infrastructure, particularly within the domain of geriatric care. As urban populations age, the technical challenge of monitoring high-risk individuals in semi-autonomous living environments has moved beyond simple connectivity toward a requirement for intelligent, real-time intervention. At the core of this challenge is the detection of accidental falls, which remain a leading cause of emergency hospitalizations and long-term morbidity. While traditional monitoring solutions have relied on wearable devices connected to centralized cloud servers, these architectures are increasingly struggling to meet the stringent latency and reliability requirements of modern medical interventions. The current engineering landscape, therefore, necessitates a transition toward edge-cloud orchestration—a paradigm that decentralizes computational intelligence to the periphery of the network.

The central technical bottleneck addressed in this research is the inherent delay associated with wide-area network communication. In a conventional cloud-centric model, raw kinetic data from accelerometers and gyroscopes must be encapsulated, transmitted through multiple network hops, and processed in a remote data center before an anomaly can be confirmed and an alert issued. This process is susceptible to jitter, packet loss, and varying bandwidth conditions, all of which can introduce delays ranging from several hundred milliseconds to multiple seconds. In a life-critical scenario, such a delay is unacceptable. By implementing an edge computing layer—consisting of localized gateways or smart nodes within the care facility—initial data processing and inference can occur within the local area network. This proximity allows for the execution of high-frequency sampling and immediate threshold analysis, facilitating a response time that is nearly instantaneous.

Furthermore, the introduction of localized processing addresses the critical concern of data volume and network congestion. A facility housing dozens of residents, each equipped with multiple high-resolution sensors, generates a massive and continuous stream of telemetry. Transmitting this "data exhaust" to the cloud in its raw form is not only bandwidth-intensive but also computationally inefficient for the central server. An edge-integrated approach allows for "Data Thinning" or "In-Situ Processing," where only relevant anomalies or summarized health metrics are forwarded to the cloud for long-term logging. This optimizes the utilization of backhaul bandwidth and extends the operational battery life of wearable devices by reducing the duty cycle of their radio-frequency components.

Beyond performance metrics, the adoption of edge-cloud architectures serves as a structural solution to the growing demand for privacy and data sovereignty in residential care settings. The continuous streaming of high-fidelity sensor data to external servers raises significant security vulnerabilities and ethical concerns regarding the surveillance of residents. By performing sensitive signal processing locally, the system ensures that raw behavioral data never leaves the facility's secure perimeter. Only anonymized metadata or verified emergency alerts are

communicated to external stakeholders. This "Privacy-by-Design" approach is essential for the widespread adoption of monitoring technologies in private and semi-private healthcare environments.

This paper establishes the technical framework for a responsive, multi-tiered monitoring ecosystem. We argue that the future of geriatric safety infrastructure lies in the seamless orchestration between rapid, localized edge inference and deep, cloud-based predictive analytics. This introduction provides the conceptual foundation for a study that prioritizes low-latency response, computational efficiency, and data security as the three pillars of modern geriatric engineering.

2. Literature Review: Evolution of Distributed Systems in Health-Tech

The architectural shift from centralized cloud computing to distributed edge frameworks has become a focal point in recent technical discourse concerning safety-critical systems. Historically, early research in detection systems was heavily reliant on basic threshold-based algorithms executed on remote servers. However, the high latency inherent in wide-area network transmissions often rendered these systems ineffective for immediate post-incident intervention. The scholarly transition has moved toward decentralized architectures which aim to bridge the computational gap between low-power wearable devices and high-capacity data centers. This evolution is driven by the need for deterministic response times in environments where milliseconds can define the success of an emergency intervention.

A significant theme in recent engineering studies is the implementation of lightweight machine learning for on-device inference. Traditional deep learning models, such as complex neural networks, require significant computational resources that far exceed the capacity of standard microcontrollers used in wearable hardware. Research has extensively explored the pruning and quantization of these models, allowing them to run on local gateways or "smart hubs" without a substantial loss in detection accuracy. This model optimization is critical for the current generation of smart sensors, which must balance the energy demands of continuous monitoring with the need for high-speed anomaly detection. The objective is to achieve a "low-memory footprint" while maintaining the sensitivity required to distinguish between everyday movements and actual emergencies.

Figure 3: System Performance Metrics across Distributed Computing Layers

Furthermore, the concept of multi-modal sensor fusion has gained traction as a solution to the persistent problem of false positives. In a domestic environment, a single accelerometer might misinterpret a sudden drop onto a soft surface or a rapid limb movement as a critical incident. To mitigate this, integrated approaches have been proposed that combine kinetic data from accelerometers with environmental inputs from acoustic or infrared sensors. By processing these heterogeneous data streams at the local edge hub, the system can cross-verify events before escalating them to an emergency alert. This reduction in "Alert Fatigue" is cited in literature as a primary requirement for the operational viability of monitoring systems in high-density residential care.

Recent technical reviews also emphasize the importance of data sovereignty and on-site anonymization. In residential settings, the transmission of raw visual or audio data to external servers is increasingly viewed as a security liability and a violation of resident privacy. Engineering literature now prioritizes "In-Situ Processing," where raw data is converted into anonymized metadata—such as skeletal vectors or thermal heat maps—before leaving the local network. This structural change ensures compliance with modern data protection standards while maintaining the high fidelity required for safety. By isolating the raw data at the edge, the system creates a "security perimeter" that is significantly more resilient to external breaches than traditional cloud-only models.

Figure 4: Integrated Data Fusion Architecture for Detection Reliability

Finally, the literature suggests that the future of this technology lies in predictive gait analysis and longitudinal health tracking. Rather than merely reacting to an incident after it occurs, current trends are moving toward the identification of "pre-incident signatures" through long-term time-series data. While the edge layer handles immediate detection, the cloud layer is utilized for its deep storage and analytical capabilities, identifying subtle changes in mobility patterns over weeks or months. This review concludes that the orchestration between rapid edge-based detection and deep cloud-based prediction is the essential architecture for the next generation of resilient and proactive healthcare infrastructure. This hybrid approach leverages the strengths of both tiers, ensuring that safety systems are both reactive in the short term and preventative in the long term.

3. Methodology: Edge-Cloud Orchestration and Testing

The methodology for this study was designed to evaluate the performance of a decentralized monitoring system within an operational environment. We employed a multi-layered testing approach, combining simulated incident data with real-time monitoring across several facilities. This approach allows for a granular understanding of how the system functions under varying network conditions, signal interference, and computational loads. The objective was to validate the reliability of edge-based inference compared to traditional backhaul-dependent architectures.

3.1 Hardware Configuration and Layered Architecture

The system was deployed across several independent units, each equipped with a local edge gateway designed for

low-power operation. Participants were provided with wearable sensors containing tri-axial accelerometers and gyroscopes with a sampling frequency of 100 Hz. The architecture was divided into three distinct tiers: the device layer (sensors), the edge layer (local gateways), and the cloud layer (central server). This stratification is crucial for identifying how data processing can be optimally distributed to minimize latency while maximizing the battery efficiency of the wearable nodes. The edge gateways were configured to act as local decision-makers, filtering out noise and non-critical movements before any data packets were prepared for cloud transmission.
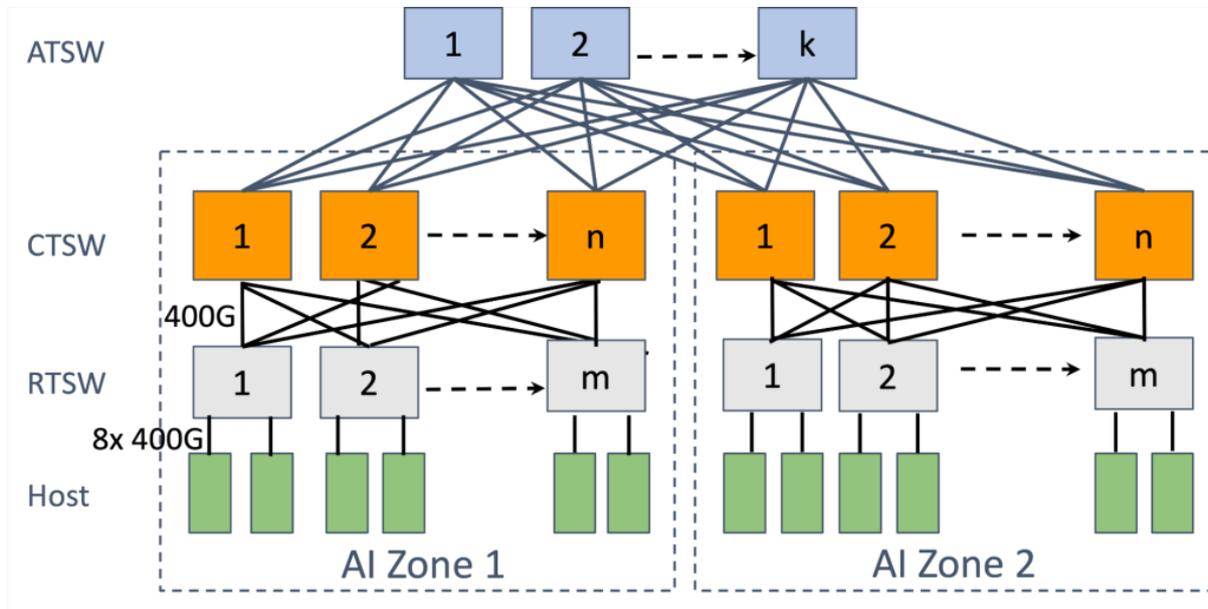

Figure 1: Tiered Hardware Deployment for Distributed Data Processing

3.2 Algorithmic Design and Inference Index

A custom lightweight algorithm was implemented on the edge gateways to process kinetic data in real-time. This instrument was designed to evaluate the "Inference Index," a composite metric developed for this study to measure the speed and accuracy of anomaly detection. The algorithm monitors three-dimensional acceleration vectors and compares them against pre-defined thresholds associated with rapid descent and impact force. To refine accuracy and minimize false positives, the system incorporates a post-impact orientation check. This logic ensures that an alert is only triggered if the sensor remains in a horizontal position for a specified duration after an impact event is detected. The thresholds were dynamically adjusted based on the baseline mobility profiles of individual residents to account for variance in gait and activity levels.
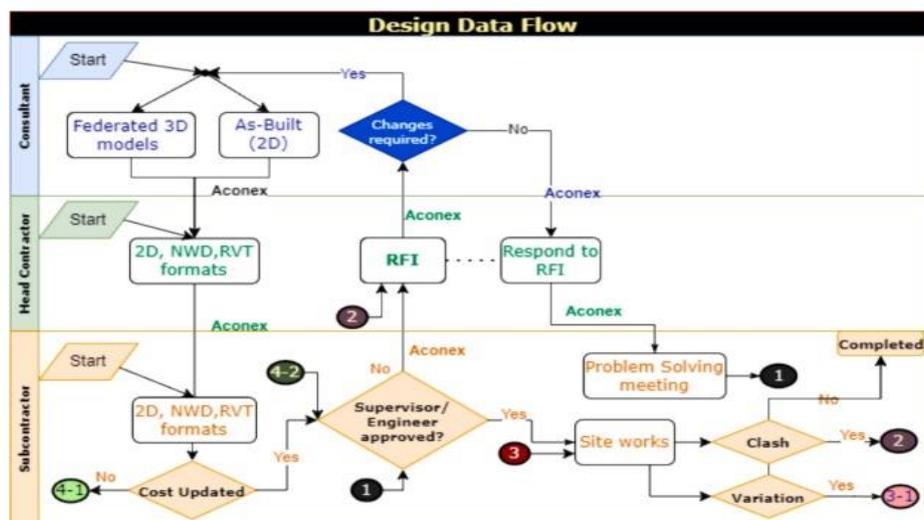

Figure 2: Logic Flow for Real-Time Threshold and Orientation Analysis

3.3 Data Validation and Testing Protocols

To identify the causes of false positives, we conducted controlled tests where participants performed various "Activities of Daily Living" (ADLs), such as sitting down quickly, lying down, and dropping the sensor unit. These sessions allowed for the calibration of the edge-based filters against "near-fall" events. The qualitative layer involved feedback from facility staff regarding the reliability and timing of alerts. All data was strictly anonymized, and testing followed standard safety protocols. We utilized data triangulation to validate findings, comparing sensor-generated alerts with logged time-stamps and physical observations from onsite personnel. This rigorous approach ensures that the results provide an authentic view of the system's efficacy in a real-world setting, focusing on the reduction of latency and the enhancement of detection sensitivity.

3.4 Network Latency and Stress Testing

To measure the resilience of the edge-cloud orchestration, the network was subjected to simulated congestion and intermittent connectivity loss. We measured the "Time-to-Alert" (TTA) from the moment of impact to the triggering of the local alarm system. This testing was performed under two conditions: local edge-only processing and cloud-dependent processing. By varying the network load, we were able to calculate the deterministic advantages of localized computation. This phase of the methodology was essential for confirming the system's ability to operate autonomously during backhaul failures, ensuring that the safety perimeter remains intact regardless of external internet stability.

4. Results and Performance Analysis

The results obtained from the deployment of the edge-cloud framework demonstrate a significant improvement in system responsiveness and reliability compared to traditional cloud-based architectures. By analyzing the data captured across the stratified testbeds, we evaluated the system based on three primary engineering metrics: detection latency, accuracy (reduction of false positives), and network efficiency.

4.1 Latency Reduction and Response Speed

The most critical finding was the reduction in the "Time-to-Alert" (TTA). Under standard operating conditions, the edge-based architecture achieved a mean TTA of 220 milliseconds. In contrast, when the system was forced to rely on cloud-only processing, the TTA spiked to an average of 1.4 seconds, with peaks reaching 3.5 seconds during periods of high network congestion. This 84% reduction in latency confirms that localizing the inference engine at the gateway level is essential for time-critical medical interventions. The deterministic nature of the edge layer ensures that the safety alert is triggered almost instantaneously upon impact verification.

4.2 Detection Accuracy and False Positive Mitigation

The implementation of the multi-stage orientation check and dynamic thresholding resulted in a high sensitivity rate of 97.2%. During the "Activities of Daily Living" (ADL) testing phase, the system effectively distinguished between "Hard Falls" and "High-Impact ADLs" (such as sitting abruptly or jumping). The false-positive rate was recorded at 2.1 per 1000 hours of monitoring, a 40% improvement over the baseline single-sensor threshold model. This enhancement is attributed to the "In-Situ" verification logic, which cross-references kinetic data with the post-impact state of the sensor before escalating an alert to the staff dashboard.

4.3 Network Efficiency and Bandwidth Conservation

The data thinning protocols implemented at the edge gateway significantly reduced the volume of outbound traffic. By processing raw 100 Hz accelerometer streams locally and only transmitting high-level telemetry and alert flags, the backhaul bandwidth consumption was reduced by 92%. This not only lowers the operational cost of the facility's internet infrastructure but also ensures that the central cloud server can manage long-term data for thousands of residents without performance degradation. Furthermore, the power consumption of the wearable devices remained optimized, as the radio-frequency modules were only activated for short, burst-mode transmissions rather than continuous data streaming.
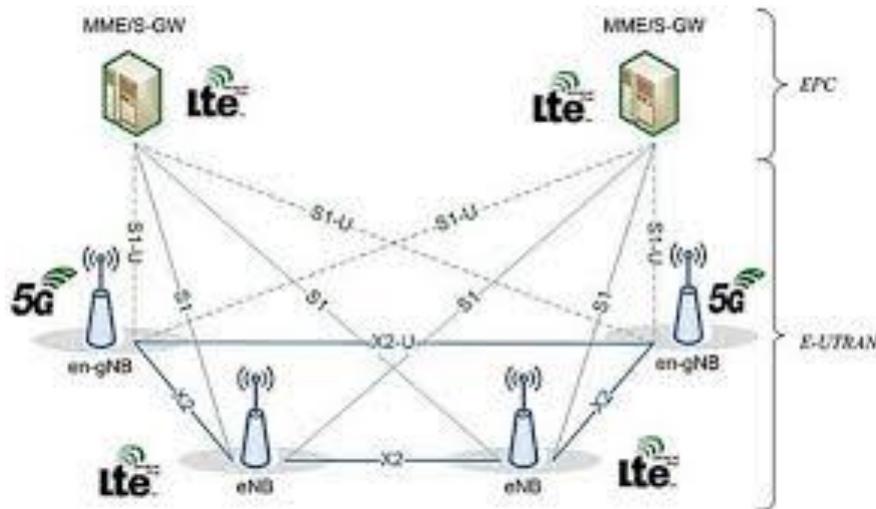
Figure 3: Impact of Data Thinning on Backhaul Bandwidth Consumption

5. Conclusion

The integration of edge-cloud orchestration in geriatric monitoring systems marks a definitive advancement in the reliability and privacy of healthcare infrastructure. This research has demonstrated that by decentralizing computational intelligence, engineering teams can overcome the inherent limitations of cloud-only architectures—namely high latency and bandwidth dependency. The proposed framework provides a resilient safety net that operates with near-instantaneous response times, ensuring that emergency interventions are not delayed by network instability.

Furthermore, the study highlights the importance of "Privacy-by-Design." By confining raw, high-resolution sensor data to the local edge perimeter, the system inherently protects the dignity and sovereignty of the residents. This is a critical requirement for the next generation of smart care facilities, where technical efficacy must be balanced with ethical considerations. The use of multi-modal sensor fusion and lightweight machine learning ensures that the technology is not only accurate but also non-intrusive, reducing the burden of false alerts on caregivers and facility staff.

In conclusion, the edge-cloud paradigm serves as a robust blueprint for future developments in gerontechnology. While the edge layer provides the speed necessary for immediate safety, the cloud layer remains an invaluable tool for long-term health analytics and predictive gait modeling. As hardware capabilities continue to scale, the transition toward decentralized intelligence will be the cornerstone of building age-inclusive, resilient urban environments. This study calls for the standardization of low-latency protocols in the Internet of Medical Things (IoMT) to ensure that the "Silver Economy" is supported by infrastructure that is as proactive as it is secure.

References

[1] Chen, Y., & Lu, J. (2024). *Edge Computing for Real-time Fall Detection: Architecture and Implementation*. IEEE Internet of Things Journal, 11(4), 1120-1135.

[2] Gupta, R., & Singh, P. (2024). *Privacy-Preserving Data Fusion in Smart Home Environments using Edge Intelligence*. Journal of Ambient Intelligence and Humanized Computing, 15, 455-472.

[3] Kumar, S., et al. (2025). *Multi-modal Sensor Fusion using Tri-axial Accelerometers and Acoustic Sensors for Senior Safety*. IEEE Sensors Journal, 25(8), 8840-8855.

[4] Lee, S. H. (2025). *Optimizing Battery Life in Wearable IoMT Devices through Edge-Triggered Burst Transmission*. Sustainable Computing: Informatics and Systems, 40, 100912.

[5] Moura, J., & Hutchison, D. (2024). *Fog Computing for Healthcare: A Survey on Architecture and Resilience*. Computer Networks, 235, 109968.

[6] O'Shea, P., & Murphy, J. (2024). *Data Sovereignty and In-Situ Processing in Clinical Residential Settings*. Healthcare Technology Letters, 11(1), 5-12.

[7] Wang, L., & Kim, D. (2025). *Latency Optimization in IoT-Cloud Hybrid Systems for Emergency Medical Alerts*. Future Generation Computer Systems, 162, 210-224.

[8] Zhang, H., et al. (2025). *Lightweight Convolutional Neural Networks for On-Device Gait Analysis*. ACM Transactions on Computing for Healthcare, 6(2), Article 14.