 ISSN:2302-8149
e-ISSN:2540-9719

ME-SHMS: Adaptive Unsupervised Anomaly Detection for Resilient Edge-based Remote Patient Monitoring

[1]Aalaa Ahmed Mohammed, [2]Ahmed Nazar Hassan*, [3]Abdullah S. Sulaiman, [4]Mohammed Taher Ahmed Mohammed
[1]Department of Environmental Engineering, College of Engineering, University of Tikrit (Iraq)
[2]Department of Networks, College of Computer Science and Mathematics, University of Mosul (Iraq)
[3]Computer Centre, University of Mosul (Iraq)
[4]Department of Cybersecurity, College of Computer Science and Mathematics, University of Tikrit (Iraq)
*e-mail: ahmadccniit@uomosul.edu.iq

(***received***: 14 September 2026, ***revised***: 18 October 2026, ***accepted***: 19 October 2026)

Abstract

In this paper we propose MESHMS (Smart Health Monitoring System), an edge assisted RPM architecture for intermittent connectivity and scarce labeled clinical anomaly data. The four-layer architecture processes patient telemetry in real time; FIFO flash buffers at the edge gateway preserve temporal ordering during network disruptions; unsupervised conservative OR fusion combines DAE with IF without need for clinical anomaly labels for model fitting or threshold calibration. An EWMA threshold dynamically adapts the DAE decision boundary to gradual physiological baseline drift. Evaluated on the MHEALTH and synthetic multisensor datasets, the OR fused DAE and IF configuration achieved 96.71% accuracy and an F1 score of 0.9317. Deployed on an ARM Cortex A72 edge resource profile, the pipeline required 38.5 MB of RAM with an inference latency of 3.8 ms. Across five simulated network outages ranging from 5 minutes to 24 hours, zero packet loss occurred while maintaining full FIFO sequence integrity. Under an 8% simulated baseline drift, the EWMA threshold with $\alpha = 0.01$ reduced false positives by 64.5% compared to a static threshold. While these findings demonstrate technical feasibility under benchmark and simulated outage conditions, clinical readiness still requires prospective field and clinical validation.

Keywords: *remote patient monitoring, unsupervised machine learning, deep autoencoder, isolation forest, exponentially weighted moving average*

1 Introduction

Global population growth is accelerating to the point of irreversible change and the world is shifting to an aging demographic. According to the United Nations World Population Prospects, individuals aged 65 and over will surpass 2.2 billion globally by the late 2070s, with life expectancy already reaching 73.3 years in 2024 [1]. By 2050 this cohort will comprise 22% of the global population [2] and put traditional healthcare systems under unprecedented operational and financial strain [3]. The MENA region is no exception to the global demographic transition towards population aging. In 2018 non-communicable diseases (NCD) were responsible for 74% of deaths in the region; diabetes, cardiovascular disease, chronic respiratory disease, cancer and end-stage renal disease are major chronic disease burdens [4]. This growing burden of NCDs together with population aging puts pressure on health systems and strengthens the need for scalable remote monitoring strategies.

In response, Remote Patient Monitoring (RPM) leveraging the Internet of Things (IoT) and Artificial Intelligence (AI) has emerged as a mechanism for supporting continuous care beyond conventional clinical settings [5, 6]. A systematic review by Tan et al. reported that RPM interventions demonstrated positive outcomes for patient safety and adherence, improved mobility and functional status in several studies, and showed a downward trend in hospital admission or readmission risk, length of stay, outpatient visits, and non-hospitalization costs. The same review also noted that the global RPM market was projected to grow at a compound annual growth rate of 18.9% between 2021 and 2028. There is a lack of evidence regarding outcomes and quality of life in clinical settings and more implementation studies and economic evaluations are required [6].

Challenges for the implementation and effective use of RPM systems in healthcare in the MENA region, however, persist. Three key challenges are interrelated: (1) Technological, organizational,

 ISSN:2302-8149
e-ISSN:2540-9719

legal and regulatory, cultural, financial, and individual-level barriers to telemedicine in the region and across countries exist [7]. These systems will have to deal with delays in telemetry because they are based on continuous cloud connectivity. (2) Locally relevant and large enough labeled clinical anomaly datasets for the detection of rare events in elderly patients do not exist. Therefore, supervised approaches to anomaly detection using supervised learning (e.g., supervised neural networks) that require large amounts of labeled data of normal and abnormal behavior will encounter practical challenges. (3) Usability concerns, issues of privacy, and health literacy challenges to the adoption of digital health by elderly patients exist and have to be addressed by targeted health education and user-centered health IT design [8, 9]. These constraints motivate RPM architectures that can operate under intermittent connectivity, learn from unlabeled patient baselines, and minimize intrusive or difficult-to-use sensing requirements.

Existing research addresses individual facets of this problem but leaves an integration gap. MENA-focused telemedicine and digital-health reviews [10, 11] document infrastructural, organizational, privacy, and adoption barriers, yet they do not offer an experimentally evaluated, resilience-oriented RPM architecture. Edge-based elderly-monitoring systems [12, 17, 18] reduce latency and cloud dependence, but continuity of ordered telemetry during prolonged upstream outages is not their primary evaluation target. Unsupervised AE/IF frameworks [13, 14] have shown label-independent anomaly detection but mostly on IoT network-security traffic and not patient-specific physiological signals. Recent work on personalized unsupervised and self-supervised anomaly detection in real world dementia monitoring [25] is clinically relevant but addresses a different sensing and deployment problem and does not evaluate FIFO outage recovery with DAE–IF fusion and EWMA based drift adaptation. None of the reviewed studies address connectivity resilience, label independent patient specific detection, complementary abrupt and gradual events detection, baseline drift adaptation and edge resource feasibility jointly within one pipeline evaluated. ME-SHMS attempts to fill this gap. Its novelty is not that DAE or IF are used individually nor EWMA or edge computing; it is rather that they are integrated and controlled evaluation performed together in one RPM resilience oriented pipeline. The contributions of this study are:

1. Outage-resilient edge telemetry management. We present a gateway-level design where flash backed FIFO buffering and sequence aware recovery preserve temporal ordering of patient telemetry when upstream connectivity is lost: local sensing, inference and event logging continue during the outage; deferred records are synchronized in order once link is restored.
2. Label independent patient specific anomaly detection. The DAE and IF detectors are fitted to each patient's own normal baseline (no labeled clinical anomalies used for model fitting, threshold initialization or adaptive threshold updating). Anomaly labels only serve post-inference evaluation.
3. DAE–IF complementation for screening. A conservative OR-fusion rule combines reconstruction evidence (sensitive to gradual physiological deviations) with isolation evidence (sensitive to abrupt structural outliers). Recall is increased at the cost of false positives but this trade-off is suited for monitoring oriented towards screening.
4. Contamination controlled drift adaptation: EWMA updates the DAE decision threshold only with samples classified as normal, so that threshold drifts gradually with baseline drift but contamination by anomalies is controlled.
5. Hybrid technical validation. The pipeline is evaluated on the public MHEALTH benchmark and on a controlled MENA-oriented synthetic multisensor scenario.The synthetic scenario permits systematic stress testing of specified anomaly classes (tachycardia, desaturation, falls) and simulated outages of 5 minutes to 24 hours.
6. Edge-resource feasibility assessment. Inference latency (3.8 ms) and memory footprint (38.5 MB) are profiled under an ARM Cortex-A72 resource configuration. This assesses compatibility with low-resource edge deployment prior to validation on physical gateway hardware. We emphasize that the novelty lies not in the individual use of DAE, IF, EWMA, or edge computing, but in their coordinated integration and controlled evaluation within a single resilience-oriented RPM pipeline.

The remainder of this paper is organized as follows. Section 2 reviews related work on edge-assisted remote patient monitoring, elderly care, and unsupervised anomaly detection, and positions ME-SHMS against it. Section 3 presents the four-layer architecture, the anomaly-detection and threshold-adaptation methodology, the security design, and the experimental setup and datasets. Section 4 reports and discusses the results, covering calibration sensitivity, the OR-fusion ablation, per-anomaly-type performance, FIFO outage resilience, adaptive thresholding, and a controlled

 ISSN:2302-8149
e-ISSN:2540-9719

comparison with adapted baselines. It also considers the study's limitations and a staged validation roadmap. Section 5 concludes the paper and outlines directions for future research.

2 Literature Review

This section reviews the most relevant work in edge-assisted RPM, elderly monitoring and unsupervised anomaly detection and then positions ME-SHMS against capabilities that are directly relevant to the present study (infrastructure resilience, labeled data dependence, physiological anomaly detection, personalization and validation setting). IoT- and AI-based RPM has pushed more data processing functions from centralized cloud platforms to the edge. Ali et al. [12] proposed an edge-centric framework for elderly monitoring based on non wearable sensors and local anomaly analysis so as to reduce latency and privacy exposure due to continuous cloud side processing. Alasmary [17] proposed SDH (ScalableDigitalHealth) that combines IoT-based RPM with latency aware edge autoscaling. Joo et al. [18] built an edge computing patient management system around IoT pulse oximeters where biosignal data can be collected locally before longer term synchronization. These works show feasibility of edge assisted monitoring but they do not test the same combination of prolonged ordered telemetry buffering, label independent DAE/IF physiological anomaly detection and adaptive EWMA drift handling as ME-SHMS. A separate line of research has been done on continuity and anomaly detection in IoT systems but infrastructure resilience and clinical anomaly detection are often considered separately. In RPM a clinically useful edge node must distinguish local inference availability from remote communication availability: the system may still sense and detect events locally while upstream link is down but remote dashboards and caregiver notifications cannot be assumed to be reachable either. This distinction drives explicit local buffering, temporal-order preservation and post reconnection recovery testing rather than treating edge processing alone as proof of end-to-end resilience.

Privacy-sensitive sensing is also needed in older-adult monitoring. mmWave radar can support activity and fall-related sensing without collecting conventional camera imagery [19, 24], so that exposure of directly identifiable visual information is less than for camera-based monitoring. Such sensing does not by itself establish legal or regulatory compliance; patient information protection also depends on secure transport, protected local storage, authentication/authorization/auditability and deployment specific governance controls.

Hybrid unsupervised anomaly detection has been studied outside physiological RPM. EcoDefender [13] combines Autoencoder representation learning with Isolation Forest scoring for resource-conscious anomaly detection in IoT edge-gateway network traffic. Bachar et al. [14] similarly combine AE and IF for IoT cybersecurity anomaly detection and report strong performance on CIC IoT-DIAD 2024 and other network-security datasets. These works support the complementarity of AE- and IF-based detectors, but their targets are network/security anomalies rather than patient-specific physiological change. In contrast, Bijlani [25] develops personalized unsupervised and self-supervised anomaly-detection methods for sensor-based remote healthcare monitoring and reports operational validation in people living with dementia. That work provides clinically relevant evidence for personalized anomaly monitoring, while ME-SHMS focuses on a different integration problem: physiological multisensor fusion, explicit edge outage recovery, DAE/IF OR-fusion, and EWMA baseline adaptation in an infrastructure-constrained MENA-oriented setting. These differences are summarised in Table 1, which does not regard diverse research as directly equivalent benchmarks.

Table 1 Comparative Analysis of Recent RPM Frameworks and the Identified Research Gap

Reference / Study	Target	Infra. Resilience	ML Paradigm	Data Dependency	Validation
Damera et al. [22]	RPM / cardiac	Low (cloud-centric)	Supervised Transformer (TL-SAM)	Labeled UCI heart-failure data	UCI clinical dataset

 ISSN:2302-8149
e-ISSN:2540-9719

Reference / Study	Target	Infra. Resilience	ML Paradigm	Data Dependency	Validation
Khan and Alkhathami [23]	IoT healthcare security	N/A (network-security study)	Supervised RF / DNN and other ML	Labeled CICIoT2023 traffic	CICIoT2023
El-Jardali et al. [11]	MENA fragile states	N/A (review)	N/A	N/A	Scoping review
EcoDefender [13]	IoT security	Medium (edge)	Unsupervised AE + IF	Zero-label anomaly modeling	IoT network traffic
Bachar et al. [14]	IoT security	Not an RPM resilience study	Hybrid AE + IF	Unsupervised anomaly modeling	CIC IoT-DIAD 2024 + other IoT datasets
Bijlani [25]	**Remote healthcare / dementia**	**Operational; outage recovery not primary focus**	**Unsupervised / self-supervised**	**Sparse / unreliable annotations**	**Real-world dementia monitoring**
Proposed ME-SHMS	**MENA RPM**	**High (FIFO edge)**	**Unsup. DAE + IF + EWMA**	**No anomaly labels for training**	**MHEALTH + synthetic + outages**

3 Research Method

This section details the four-layer ME-SHMS architecture and its end-to-end data path. Layer 1 acquires non-intrusive physiological and activity signals; Layer 2 validates, normalizes, encrypts, and buffers telemetry at the edge when upstream communication is unavailable; Layer 3 performs patient-specific unsupervised inference locally under the proposed deployment topology; and Layer 4 synchronizes longitudinal records and exposes authenticated alert/dashboard services when remote connectivity is available.

3.1 Design Philosophy and Architecture Overview

ME-SHMS is designed around three principles. First, resilience-by-design: incoming telemetry is assigned temporal/sequence metadata and can be buffered locally so that upstream communication loss does not automatically imply loss of the monitored time series. Second, label-independence: model fitting and threshold calibration use patient-specific baseline data without requiring labeled clinical anomaly repositories. Third, deployability-oriented design: the inference pipeline is constrained to an edge-class computational envelope and is therefore profiled under an ARM Cortex-A72 resource configuration. The current study evaluates this design under a resource profile and simulated outages; it does not claim that the complete clinical system has already been validated on production hardware or in routine care.

ME-SHMS has four layers in sensor-to-edge-to-notification flow. Sensing Layer (Layer 1) collects the physiological/activity measurements and generates time stamped feature vectors; Edge Resilience Layer (Layer 2) validates, clips outliers, normalizes, assigns sequences, encrypts and handles connectivity state; under our proposed deployment topology the normalized vector is delivered locally to co-located AI Inference Layer (Layer 3), where DAE and IF run concurrently and their decisions are fused together (DAE threshold is updated by EWMA only from observations classified as normal). Notification and Logging Layer (Layer 4) receives anomaly verdicts and longitudinal records for authenticated dashboard access and caregiver workflows when upstream connectivity is available. Local sensing, inference, event logging and FIFO preservation can continue

 ISSN:2302-8149
e-ISSN:2540-9719

even if upstream goes down completely but remote notification cannot be guaranteed until communication path is restored. Adaptive feedback path therefore resides in the inference/monitoring process itself rather than on continuous availability of cloud.

3.2 Layer 1: Non-Intrusive Physiological Sensing

For Layer 1, three non-intrusive sensing modalities are considered to address usability and privacy constraints in older-adult home monitoring [9]. A 60 GHz millimeter-wave (mmWave) FMCW radar mounted in the living space provides reflection-derived information from which position, velocity, posture, gait-related features, respiration-related features, and fall-related indicators can be derived. Because conventional camera imagery is not collected, this sensing modality reduces exposure of directly identifiable visual information relative to camera-based monitoring, although privacy and regulatory compliance still depend on the complete system security and governance controls [19, 24]. A pressure-sensitive floor mat at selected high-traffic transitions records weight-distribution and step-frequency information. A Bluetooth Low Energy (BLE) wristband provides heart rate (HR), blood oxygen saturation (SpO_2), and skin temperature. At each time step t, the available modalities are concatenated into a unified physiological feature vector. The present architecture specifies BLE for the wristband; the physical bus/interface used by the radar and floor-mat modules may depend on the final hardware integration and is not treated as a standardized contribution of this study.

$$S_t = [HR_t, SpO_{2t}, Temp_t, Gait_t, Resp_t, FallFlag_t, PressurePattern_t] \in R^7$$

3.3 Layer 2: Resilient Edge Gateway and FIFO Buffering

Layer 2 is the edge-resilience part of ME-SHMS. The gateway checks incoming measurements, clips implausible outliers according to the preprocessing configuration, min-max normalizes vectors, attaches temporal/sequence metadata and stores upstream connectivity state. Connectivity probe is lightweight in the prototype used to make link-state decisions but telemetry delivery success is still a separate transport level concern. In the proposed deployment topology normalized vectors are passed locally to Layer 3 for anomaly inference rather than requiring a cloud round trip. When remote synchronization is available and FIFO queue in local node is empty records selected for synchronization are sent over TLS 1.3 protected channel. Measured packet size (in reported prototype configuration) of telemetry packets is 128 bytes. When upstream link fails incoming records are protected and appended to local FIFO flash-memory buffer with metadata needed to reconstruct temporal order (deployment security profile requires AES-256 authenticated encryption of buffered data at rest but that requirement is distinguished from performance analysis in Section 3.6). With measured packet size of 128 bytes an 8 GB eMMC device has quite large nominal capacity for single patient telemetry at 60 s sampling interval but real deployments must also account for filesystem overhead encryption metadata logging retention rules flash endurance multi-patient scaling. After reconnection gateway chunked recovers bandwidth aware while preserving FIFO order. Note that outage resilience claim is about preservation of telemetry and locally generated event records: remote SMS/push/dashboard delivery is not guaranteed in complete upstream loss and pending remote notifications are synchronized after connectivity comes back on.

3.4 Layer 3: AI Inference Engine

Layer 3 includes an AI inference engine that detects physiological anomalies using reconstruction-based error evaluation and tree-based extreme-event scoring. It is calibrated for the patient in a specific calibration phase such that normal physiological behavior is used as a baseline to detect deviations with high recall.

3.4.1 Deep Autoencoder (DAE) Calibration

The core of the proposed framework is a deep autoencoder (DAE) consisting of an encoder $f\theta$ and a decoder $g\varphi$ where θ and $\emptyset$ are trainable parameters (weights and biases) of the encoder and decoder networks respectively, which are optimized during calibration. Calibration happens only on samples recorded from the target patient under normal physiological conditions; anomalous samples are deliberately omitted from training set because abnormal patterns are not learned into the baseline

 ISSN:2302-8149
e-ISSN:2540-9719

otherwise model sensitivity to deviations at inference time would be reduced. For an input vector X∈ R^d the encoder maps it to a lower dimensional latent representation z ∈ R^N (N < d) using Equation(1):

$$z = f\theta(X) \quad (1)$$

The latent vector z summarizes the physiological state of the patient. The decoder reconstructs the input from this representation: Equation (2)

$$\hat{X} = g\varphi(z) \quad (2)$$

The encoder comprises two fully connected hidden layers with ReLU activation functions, and the decoder mirrors this structure symmetrically. The model parameters are learned by minimizing the mean squared error (MSE) between the input and its reconstruction, as defined by Equation (3):

$$L(\theta, \varphi) = \frac{1}{N}\sum_{j=1}^{N} \quad \frac{1}{d}\sum_{i=1}^{d}(Xj_i - \hat{X}j_i)^2 \quad (3)$$

where N is the number of calibration samples and d is the input dimensionality. Minimizing this objective means that DAE learns to reproduce normal physiological patterns of the patient correctly so samples away from this baseline will have higher reconstruction errors at inference time.

3.4.2 Anomaly Scoring and Threshold Determination

After calibration, each incoming sample X is passed through the trained DAE, and its anomaly score is defined as the per-sample reconstruction error, which is calculated using Equation (4):

$$S(X) = \frac{1}{d}\sum_{j=1}^{d}(Xj - \hat{X}j)^2 \quad (4)$$

3.4.3 Isolation Forest (IF) Scoring

An Isolation Forest (IF) is used in parallel with the DAE to detect abrupt or extreme outlier events. IF isolates observations by recursively selecting a feature and a split value between that feature's observed minimum and maximum values. For an event X, the anomaly score produced by a forest fitted to N observations is computed according to Equation (5):

$$S(X, N) = 2^{-\frac{E[H(X)]}{C(N)}} \quad (5)$$

where E[h(X)] is the expected path length of X across the isolation trees, and c(N) is the normalization factor for N observations. Scores near 0.5 are typical observations; values approaching 1 are stronger anomaly evidence.

3.4.4 Dynamic Thresholding and Fusion Logic

Because ME-SHMS is intended as a screening-oriented monitoring framework, Layer 3 uses a conservative logical OR-fusion rule: a physiological sample is flagged when either the DAE reconstruction error exceeds its adaptive threshold gamma or the IF anomaly score exceeds its fixed threshold theta. The DAE threshold is initialized from the 95th percentile of calibration reconstruction errors and is then updated using EWMA to follow gradual physiological drift, as governed by Equation (6):

$$\gamma new = \alpha \cdot Loss_t + (1 - \alpha) \cdot \gamma old \quad (6)$$

The conditional EWMA update is important: only the reconstruction loss of time steps classified as normal is used to update the adaptive threshold. Samples classified as anomalous are excluded from the update to reduce contamination of the normal baseline estimate.

 ISSN:2302-8149
e-ISSN:2540-9719

3.5 Layer 4: Alert and Dashboard Layer

Layer 4 converts the binary inference output into caregiver-facing services and longitudinal records. The Dynamic Alert Engine creates an event containing a pseudonymous patient identifier, timestamp, anomaly type, and the model(s) contributing to the decision (DAE, IF, or both). Remote SMS or push notifications should only contain information needed to take action (raw physiological vectors and detailed records are restricted to the authenticated dashboard); monitoring dashboard shows longitudinal physiological trends and anomaly history, Routine Data Logger stores time stamped physiological vectors reconstruction errors anomaly scores verdicts for retrospective review and local EWMA update process. If upstream link is down remote delivery is deferred while event records remain queued locally according to resilience policy of Layer 2.

3.6 Security, Authentication, and Patient Information Protection

ME-SHMS is defense-in-depth because edge processing alone does not protect health information completely. For data at rest AES-256 authenticated encryption (e.g., GCM) in AES standard and NIST guidance for GCM [26, 27] are specified; TLS 1.3 with current IETF specification [28] for gateway-to-remote synchronization is specified; cryptographic keys have to be device specific and stored in protected operating system keystore or hardware backed secure element (if available); keys cannot be stored plaintext alongside buffered patient records. These cryptographic mechanisms are design requirements of the system and were not independently penetration tested or certified in the present experimental study. Authentication/authorization are separated from encryption: each registered gateway must have its own cryptographic device identity; clinical and administrative users access dashboard via authenticated accounts (role based access control, principle of least privilege). Multi-factor authentication has to be enabled in deployment profile for privileged clinical/administrative accounts. Pseudonymous identifiers are used to refer to patient records where identity is not required for inference; longitudinal records, alert acknowledgement, administrative actions and authentication events should be logged in tamper-resistant audit logs. Deployment policies must also list credential revocation, key rotation, data retention periods, secure deletion, protected backups and incident response procedures. Because legal obligations vary by jurisdiction these controls should not be interpreted as proof of universal regulatory compliance; privacy impact assessment, cybersecurity testing and jurisdiction specific regulatory review are still prerequisites for clinical deployment.

3.7 Experimental Setup and Datasets

ME-SHMS models were implemented in Python 3.11 using Scikit-Learn 1.4 and PyTorch 2.2 on an Intel Core i7-12700K host machine. Edge-resource evaluations were conducted under an ARM Cortex-A72 resource profile configured as a 4-core, 1.5 GHz, 4 GB LPDDR4 environment; these results therefore represent profile-based feasibility measurements rather than measurements from a complete physical clinical gateway. Latency and RAM footprint are reported for this profile. CPU utilization, energy consumption, thermal behavior, and flash-endurance effects were not measured and are consequently reserved for the physical-gateway validation stage. Telemetry packet size was measured at 128 bytes per packet in the reported prototype configuration and used in the FIFO buffering-capacity analysis.

Two distinct validation environments were employed. For public benchmark validation, the open-source MHEALTH dataset [15] — comprising 23 physiological and motion sensor channels from 10 subjects performing 12 distinct physical activities — was evaluated using an 80/20 train-test split to assess anomaly-detection performance under multimodal sensor noise. The same preprocessing pipeline was applied across ME-SHMS and all adapted comparison implementations. Activities deviating more than two standard deviations from the resting baseline distribution were assigned anomaly labels for evaluation purposes only; these labels were not used during model training, threshold calibration, or adaptive-threshold updating.

For the simulated regional dataset, a MENA-oriented synthetic deployment scenario was generated, comprising records from 50 virtual elderly patients monitored at a 60-second sampling interval over 200 log cycles, resulting in 10,000 total records. We used 5% event rate and 500 total events in total, split evenly across 3 types of synthetic events: tachycardic episodes (HR > 140bpm) (n=170), desaturation events (SpO_2 < 92%) (n=165), and kinetic fall events (n=165). The generated data set for training the models contained 9,500 normal events, and the events were all un-labeled and

 ISSN:2302-8149
e-ISSN:2540-9719

used for training and calibration of the models in an unsupervised learning fashion. The DAE architecture had input layer of size d=7, 2 hidden layers in the encoder with ReLU activation of size 16 and size 8 respectively, a bottleneck latent space of size 4, and corresponding symmetric decoder. The Isolation Forest (IF) had number of estimators of 200 and contamination of 0.05. All values are averages over 5 separate runs using different random seeds.

4 Results and Analysis

This section reports the experimental results and analyzes anomaly-detection performance, component complementarity, buffering resilience, adaptive thresholding, and edge-resource feasibility under the evaluated conditions.

4.1 Individual Model and Calibration Sensitivity Analysis

We also investigated the sensitivity of the detection performance for the calibration of the monitor by increasing the number of calibration samples from N = 50 to N = 200. This increase in the number of calibration samples resulted in an increase of the accuracy from 0.9201 to 0.9642, an absolute F1-score increase of 0.0800 (8.0 percentage points; approximately 9.6% relative) from 0.8354 to 0.9154, and a decrease of the false positive rate from 11.2% to 4.1%. However, a further increase from N = 200 to N = 1,000 calibration samples only marginally increased the F1-score by 0.16% points. Thus, for the subsequent investigations on the detection performance we chose the number of 200 samples (corresponding to approximately 3.3 hours of monitoring at a sampling interval of 60 seconds) as our so-called operating calibration size. The reader has to be aware that this size of the calibration set has been investigated on the basis of the used data sets and the experimental setup, and should not be misused as a fixed clinical size for calibration.

In the previous section we have introduced a variety of unsupervised anomaly detection models. In this section we compare the performance of these models in an experimental setting. We compare their performance on two different data sets: a simulated data set and the MHEALTH data set. The results of the models are presented in Table 2. We have added the K-Means clustering model as a fast unsupervised baseline. It has the highest F1-score on the simulated data set (F1-score ≈ 0.947) and, therefore, the lowest latency (0.8 ms). On the other hand, the best performance on the MHEALTH data set was achieved by the Deep Autoencoder. It required the highest latency of 3.4–3.8 ms, but the F1-score of the Deep Autoencoder was the highest among all models (F1-score = 0.9154). The Isolation Forest took an intermediate position (F1-score = 0.9143, latency: 1.2 ms).

Table 2 Individual Unsupervised Model Performance Across Both Datasets

Model	Dataset	Acc.	Prec.	Recall	F1	Latency
Isolation Forest	Simulated	0.9910	0.960	0.8727	0.9143	1.2 ms
K-Means Clustering	Simulated	0.9944	0.994	0.9036	0.9467	0.8 ms
Deep Autoencoder	Simulated	0.9924	0.974	0.8855	0.9276	3.4 ms
Deep Autoencoder	**MHEALTH**	**0.9642**	**0.9410**	**0.8912**	**0.9154**	**3.8 ms**

4.2 OR Fusion Ablation Study

To analyze the contributions of the individual components DAE and IF as well as of their fusion, we perform an ablation study for the three different configurations DAE-only, IF-only and OR-fused DAE+IF on the two datasets, which are listed in Table 3. For the simulated dataset the best detection performance is achieved by the OR-fused DAE+IF configuration with an accuracy, recall and F1-score of 0.9961, 0.9412 and 0.9549, respectively. Here, the most significant improvement of the detection performance is achieved by the fused configuration of DAE and IF. In comparison to the DAE-only configuration, the recall of the detection is even increased by about 5.6 percentage points. Similar results are obtained for the MHEALTH dataset, where the best detection performance is also achieved by the OR-fused DAE+IF configuration with an accuracy, recall and F1-score of 0.9671, 0.9247 and 0.9317, respectively. Again, the lowest F1-score is achieved by the IF-only configuration.

 ISSN:2302-8149
e-ISSN:2540-9719

The main cost of the OR strategy is the increased False Positive Rate (FPR) of 3.4% (simulated data) and 5.8% (MHEALTH data) compared to 2.6% and 4.1% respectively. However, the design of ME-SHMS is screening oriented and thus improving the anomaly recall is more important than keeping the number of normal alerts low. This increased FPR has however to be addressed in the subsequent processing of alerts by thresholding or by using an appropriate alarm-management strategy.

Table 3 OR Fusion Ablation: DAE Only vs. IF Only vs. DAE+IF Fused

Configuration	Dataset	Acc.	Prec.	Recall	F1	FPR
DAE only	Simulated	0.9924	0.9740	0.8855	0.9276	2.6%
IF only	Simulated	0.9910	0.9600	0.8727	0.9143	3.1%
DAE+IF (OR fusion)	**Simulated**	**0.9961**	**0.9690**	**0.9412**	**0.9549**	**3.4%**
DAE only	MHEALTH	0.9642	0.9410	0.8912	0.9154	4.1%
IF only	MHEALTH	0.9510	0.9180	0.8614	0.8888	5.3%
DAE+IF (OR fusion)	**MHEALTH**	**0.9671**	**0.9388**	**0.9247**	**0.9317**	**5.8%**

4.3 Per-Anomaly-Type Detection Breakdown

The characteristics of the IF and DAE detectors for three clinically motivated synthetic anomalies that were created in a simulated dataset and are therefore synthetic are depicted in Table 4. The recall of IF was better for abrupt isolated geometric anomalies than the recall of DAE for the same type of anomalies. Specifically, the recall of IF for tachycardic episodes was 0.9640 whereas the recall of DAE was 0.8812. The recall of IF for kinetic falls was 0.9511 while the recall of DAE for the same type of anomalies was 0.8644. However, for anomalies that gradually deviate from a baseline, DAE performed better than IF. Specifically, the recall of DAE for desaturation of SpO_2 was 0.9334 while the recall of IF for the same type of anomalies was 0.8501. The OR fusion results for the three types of anomalies were 0.9721 for tachycardic episodes, 0.9489 for desaturation of SpO_2 and 0.9688 for kinetic falls. Across all simulated anomaly types, the fused configuration achieved overall recall of 0.9633 and a false negative rate of 3.7%.

Table 4 Per-Anomaly-Type Detection Recall: DAE vs. IF vs. OR Fused (Simulated Dataset)

Anomaly Type	DAE Recall	IF Recall	OR Fused Recall	FNR	Dominant Detector
Tachycardia (HR>140)	0.8812	0.9640	0.9721	2.8%	IF (abrupt spike)
SpO2 Desaturation (<92%)	0.9334	0.8501	0.9489	5.1%	DAE (gradual drift)
Kinetic Fall Event	0.8644	0.9511	0.9688	3.1%	IF (structural outlier)
Combined (all types)	**0.8930**	**0.9217**	**0.9633**	**3.7%**	**Fusion dominant**

4.4 FIFO Edge-Buffering Resilience Verification

Five simulated network outage scenarios, ranging from a 5-minute dropout to a 24-hour connectivity interruption, were used to evaluate the FIFO edge-buffering mechanism under controlled outage conditions Table 5. Across the five tested outage scenarios, zero packet loss was recorded and FIFO packet ordering was preserved. Recovery time increased with outage duration and buffered packet count for the reported configuration: 5 min outage recovered in less than 1 sec, 2 h outage recovered in 4.7 sec and 24 h outage recovered in 54.9 sec. Using the measured telemetry packet size of 128 bytes the eMMC flash usage increased predictably with number of buffered packets (reaching

 ISSN:2302-8149
e-ISSN:2540-9719

184.3 KB for 24 h outage). This means that the tested outage durations required only a small storage footprint under the reported single-patient, 60 s sampling configuration.

Table 5 FIFO Buffering Resilience — Simulated Outage Quantification

Outage Duration	Packets Buffered	Storage Used (KB)	Recovery Time (s)	Data Loss	Ordering Integrity
5 minutes	5	0.6	<1	0 packets	Preserved
30 minutes	30	3.8	1.2	0 packets	Preserved
2 hours	120	15.4	4.7	0 packets	Preserved
8 hours	480	61.4	18.3	0 packets	Preserved
24 hours	**1,440**	**184.3**	**54.9**	**0 packets**	**100% preserved**

4.5 Adaptive EWMA Threshold Analysis

A performance comparison of static versus EWMA adaptive thresholding was conducted under an 8% simulated physiological baseline drift condition. The static threshold, fixed at the 95th percentile, produced 31 false positives over 200 steps, corresponding to an FPR of 15.5% and an F1-score of 0.8774. EWMA adaptation with $\alpha = 0.05$ reduced false positives to 18, corresponding to an FPR of 9.0%, while $\alpha = 0.01$ produced the strongest observed trade-off among the tested settings, with 11 false positives, an FPR of 5.8%, and F1-score = 0.9317. Adaptation with $\alpha = 0.001$ was too slow: The threshold did not keep up with the drift in the simulation well and therefore was more precise but less sensitive. The F1-score was only 0.8990. However, the EWMA with $\alpha = 0.01$ was the better empirical setting for updating the threshold for the following experiments since it achieved the highest F1-score. Compared to the static threshold, EWMA led to a reduction of false positives by 64.5% at 8% drift in the simulation.

4.6 Controlled Comparison with Adapted IoT Healthcare Architectures

The comparison in this subsection is a controlled local protocol rather than a direct state-of-the-art benchmark. The publications as architectural/model inspirations are for different tasks, datasets and deployment goals so their original published metrics cannot be compared directly with ME-SHMS. For Table 6 the locally adapted proxy implementations were evaluated on the same MHEALTH 80/20 split and the same preprocessing pipeline used for ME-SHMS. The values in Table 6 are local experimental measurements only and should not be interpreted as exact reproductions of (or superiority claims over) the original published systems.

The controlled protocol consists of locally adapted baselines from Khan and Alkhathami [23], Damera et al. [22] and the Surrey sensor-based remote-healthcare work represented by Bijlani [25]. This qualification is particularly important because Khan and Alkhathami [23] study supervised cybersecurity anomaly detection on CICIoT2023 network traffic, Damera et al. [22] study supervised transformer-based heart-failure classification in a cloud-centric IoMT framework, and Bijlani [25] develops personalized unsupervised/self-supervised anomaly detection for home-sensor remote healthcare monitoring. Table 6 therefore evaluates local proxy implementations under a common MHEALTH protocol; it does not transfer the original authors' reported performance values into the ME-SHMS experiment. Under this local protocol, ME-SHMS achieved accuracy 0.9671 and F1-score 0.9317 with a 38.5 MB memory footprint and 3.8 ms inference latency under the ARM Cortex-A72 resource profile.

Table 6 Controlled Prototype-Level Comparison on the MHEALTH Dataset

Architecture	Acc.	Prec.	Recall	F1	Edge Mem. (MB)	Latency (ms)
Khan-inspired local baseline [23]	0.9412	0.9120	0.8640	0.8874	142.4	12.4
Damera-inspired local baseline [22]	0.9580	0.9350	0.8810	0.9072	84.1	5.1

 ISSN:2302-8149
e-ISSN:2540-9719

Architecture	Acc.	Prec.	Recall	F1	Edge Mem. (MB)	Latency (ms)
Surrey-inspired local baseline [25]	0.9510	0.9290	0.8750	0.9013	22.8	2.9
Proposed ME-SHMS	**0.9671**	**0.9388**	**0.9247**	**0.9317**	**38.5**	**3.8**

Note: All non-ME-SHMS entries are locally adapted proxy baselines evaluated by the authors on the same MHEALTH 80/20 split and preprocessing pipeline. The labels identify the publications that motivated the corresponding local baseline; the numerical values in this table are not the original published metrics and do not constitute exact reproductions of those systems. Because the source studies address heterogeneous tasks and datasets, this table is intended only as a controlled local implementation comparison.

4.7 Label-Independent Training and Calibration

The ME-SHMS training, calibration, threshold initialization and EWMA updating code does not use any information about anomaly labels. It does not use any labeled EHR records, any clinical event annotations, or any pre-trained supervised models. It does not use any labeled data for fitting the model and for threshold calibration. Anomaly labels are only used to compute evaluation measures after the system has produced an inference. The system learns the boundary for anomaly detection from a 200-sample baseline set that is used for threshold calibration and from the two different detection paradigms, i.e., from the error of a model's reconstructed input and from the isolation of an input within a subspace of its replicated inputs.

These experiments support the claim of label-independent model fitting and threshold calibration, together with patient-specific personalization. Each patient is calibrated against that individual's own physiological baseline rather than relying solely on a population-wide supervised classifier. This design reduces dependence on externally annotated anomalies that may not represent the target population, but it does not remove the need for clinician review of generated alerts or prospective clinical validation.

4.8 Limitations and Threats to Validity

This work has several limitations and should be read as technical/preclinical evaluation rather than proof of clinical readiness: First, evaluation is on the public MHEALTH benchmark and a MENA-oriented synthetic multisensor dataset (synthetic tachycardia, desaturation and fall cannot reproduce all variability in human physiology, multimorbidity, medication effects, adherence behavior, household context or real sensor artifacts). Second, anomaly labels are only used for post-inference evaluation but the assumed normal baseline used for patient specific calibration may still be imperfect in real populations (and requires clinician governed enrollment procedures). Third, outage experiments are controlled simulations under single-patient 60-second sampling configuration so they do not establish multi-patient scalability, flash endurance or performance under heterogeneous real networks. Fourth, ARM Cortex-A72 results are resource-profile measurements not measurements on a complete physical gateway and CPU utilization, energy consumption, thermal stability and long duration encrypted storage overhead were not evaluated. Fifth, Table 6 contains locally adapted proxy baselines from heterogeneous source studies and is not a direct reproduction benchmark. Finally, no prospective study with elderly patients, caregivers or clinicians has established clinical sensitivity, alert burden, usability, safety or impact on patient outcomes. Consequently, generated alerts remain decision-support signals requiring professional review rather than autonomous clinical decisions.

4.9 Future Validation Roadmap

Stage I – Physical gateway and field-engineering validation: ME-SHMS will be moved from the current resource profile to a physical low-power gateway embedded with the target sensors. Performance will be measured for sustained inference latency, CPU and memory usage, power consumption, thermal behaviour, encrypted storage overhead, flash endurance, sensor-to-gateway reliability, recovery under real network interruptions, cybersecurity failure modes and single versus multi-patient scaling. This stage will also include independent security testing of the authentication key management encryption authorization audit mechanisms described in Section 3.6 . Stage II —

 ISSN:2302-8149
e-ISSN:2540-9719

Prospective home care feasibility and user acceptance study: After ethics approval and informed consent, a small prospective deployment with elderly participants, caregivers and clinicians will test data completeness, system availability (end-to-end alert latency), anomaly-alert sensitivity, false alarm burden, clinician interpretation, technical failure modes and workflow integration. Patient, caregiver and professional acceptance will be assessed using validated usability/technology-acceptance instruments and qualitative feedback. Clinicians will remain in the decision loop as early stage clinical AI evaluation principles (e.g. DECIDE-AI [29]) require. Stage III - Longitudinal multi-site clinical validation: If feasibility and safety criteria are met a larger longitudinal study will address generalizability across sites (subgroups of patients), comorbidities, medications, sensor configurations and network conditions; stability of patient specific calibration and EWMA adaptation; alert fatigue; subgroup performance; clinically meaningful outcomes. If interventional randomized design is used trial reporting and AI intervention description follows applicable CONSORT-AI recommendations [30]. Only after these stages (and jurisdiction specific regulatory and privacy review) could deployment readiness be assessed.

5 Conclusion

ME-SHMS addresses two practical problems in remote patient monitoring in infrastructure constrained settings: intermittent upstream connectivity and limited labeled patient-specific anomaly data. FIFO edge buffering, unsupervised DAE–IF anomaly detection for each patient and EWMA adaptive thresholding are integrated into a resource conscious pipeline in our proposed architecture. On the MHEALTH evaluation OR-fused DAE/IF achieved 96.71% accuracy and F1-score of 0.9317 under evaluated conditions while in controlled synthetic analysis abrupt and gradual anomaly types were complementary behaviour of the detectors. Under simulated outages up to 24 h zero packet loss was observed and FIFO order was preserved under tested single-patient 60 s sampling configuration (full sequence recovery in 54.9 s for longest outage). EWMA with $\alpha = 0.01$ reduced number of false positives by 64.5 % compared to static threshold under 8% simulated baseline drift evaluated. The ARM Cortex-A72 resource profile required 38.5 MB RAM and an average inference latency of 3.8 ms. These results indicate technical feasibility under benchmark, synthetic, simulated-outage and resource-profile conditions but do not establish clinical efficacy, safety, regulatory compliance or readiness for routine healthcare deployment: physical gateway validation; prospective home care studies with elderly participants; clinician-in-the-loop assessment; user acceptance evaluation; cybersecurity validation and longitudinal multi-site clinical studies need to be completed before ME-SHMS can be considered clinically deployment ready.

References

[1] United Nations. *World Population Ageing 2019: Highlights*. New York: UN DESA, 2019.
[2] United Nations. *World Population Prospects 2024: Summary of Results*. New York: UN DESA, 2024.
[3] S. Khan et al., "*Public Health Challenges of Growing Ageing Populations*," Public Health Challenges, Vol. 3, No. 3, p. e213, 2024.
[4] Z. Z. McNatt, "*Addressing Noncommunicable Diseases among Urban Refugees in the Middle East and North Africa: A Scoping Review*," Conflict and Health, Vol. 14, No. 9, 2020, DOI: 10.1186/s13031-020-0255-4.
[5] M. Alshamrani, "*IoT and Artificial Intelligence Implementations for Remote Healthcare Monitoring Systems: A Survey*," J. King Saud Univ. Comput. Inf. SCI., Vol. 34, No. 6, pp. 3145–3161, 2022.
[6] S. Y. Tan et al., "*A Systematic Review of the Impacts of RPM Interventions*," npj Digital Medicine, Vol. 7, p. 192, 2024.
[7] H. Al-Samarraie, S. Ghazal, A. I. Alzahrani, and L. Moody, "*Telemedicine in Middle Eastern Countries: Progress, Barriers, and Policy Recommendations*," International Journal of Medical Informatics, Vol. 141, Art. No. 104232, 2020, DOI: 10.1016/j.ijmedinf.2020.104232.
[8] T. M. Alqarni, "*Perceived Usefulness of IoT in Healthcare in the Middle East*," Inform. Med. Unlocked, Vol. 45, p. 101153, 2024.
[9] N. Alqahtani and D. Almaghaslah, "*Attitudes of Elderly Population Toward Mobile Health Applications in Saudi Arabia*," Healthcare, Vol. 13, No. 19, p. 2464, 2025.

 ISSN:2302-8149
e-ISSN:2540-9719

[10] M. R. Abouzid, S. M. Elshafei, and I. Elkhawas, "*Applications of Telemedicine in the MENA Region: Benefits and Challenges*," Cureus, Vol. 14, No. 7, p. e26611, 2022.
[11] F. El-Jardali, L. Bou-Karroum, M. Jabbour, K. Bou-Karroum, A. Aoun, S. Salameh, P. Mecheal, and C. Sinha, "*Digital Health in Fragile States in the Middle East and North Africa (MENA) Region: A Scoping Review of the Literature*," PLOS ONE, Vol. 18, No. 4, Art. No. e0285226, 2023, DOI: 10.1371/journal.pone.0285226.
[12] A. Ali et al., "*An Innovative IoT and Edge Intelligence Framework for Monitoring Elderly People using Anomaly Detection*," Sensors, Vol. 25, No. 6, p. 1735, 2025.
[13] S. Jamshidi, F. Erfan, O. Abdul-Wahab, M. Bellaiche, and F. Khomh, "*Lightweight Autoencoder-Isolation Forest Anomaly Detection for Green IoT Edge Gateways*," arXiv:2511.18235, 2025.
[14] M. Bachar, A. Khiat, and K. El Guemmat, "*Hybrid Autoencoder and Isolation Forest for IoT Anomaly Detection with a Novel Model*," Engineering, Technology & Applied Science Research, Vol. 16, No. 1, pp. 31123-31129, 2026, DOI: 10.48084/etasr.15288.
[15] O. Banos, R. Garcia, and J. A. Holgado-Terriza, "*MHEALTH: A Novel Multi-Modal Dataset for Mobile Health*," UCI ML Repository, 2014.
[16] S. Ahmed, S. Irfan, and N. Kiran, "*Remote Health Monitoring Systems for Elderly People: A Survey*," Sensors, Vol. 23, No. 16, p. 7095, 2023.
[17] H. Alasmary, "*ScalableDigitalHealth (SDH): An IoT-based Scalable Framework for RPM*," Sensors, Vol. 24, No. 4, p. 1346, 2024.
[18] M.-I. Joo et al., "*A Patient Management System using an Edge Computing-based IoT Pulse Oximeter*," Appl. SCI., Vol. 14, No. 1, p. 414, 2024.
[19] A. Haque et al., "*Millimeter-Wave Radar-based Human Activity Recognition for Healthcare Monitoring Robot*," arXiv:2405.01882, 2024.
[20] Y. J. A. Tian et al., "*Benefits and Barriers of Smart Home Technologies for Older Adults*," BMC Geriatrics, Vol. 24, p. 186, 2024.
[21] M. Morshedi et al., "*A Comprehensive Review of DL Techniques for Anomaly Detection in IoT Networks*," Eng. Reports, 2025.
[22] V. K. Damera, R. Cheripelli, N. Putta, G. Sirisha, and D. Kalavala, "*Enhancing Remote Patient Monitoring with AI-Driven IoMT and Cloud Computing Technologies*," Scientific Reports, Vol. 15, Art. No. 24088, 2025, DOI: 10.1038/s41598-025-09727-z.
[23] M. M. Khan and M. Alkhathami, "*Anomaly Detection in IoT-based Healthcare: Machine Learning for Enhanced Security*," Scientific Reports, Vol. 14, Art. No. 5872, 2024, DOI: 10.1038/s41598-024-56126-x.
[24] H.-H. Chen et al., "*Millimeter-Wave Technology for Multi-Person Fall Detection Validated through Wearable Sensors and Real-Life Scenarios*," Scientific Reports, Vol. 16, Art. no. 8859, 2026, DOI: 10.1038/s41598-026-40330-y.
[25] N. Bijlani, "*Unsupervised Anomaly Detection in Sensor-based Remote Healthcare Monitoring*," Ph.D. Dissertation, University of Surrey, 2025, DOI: 10.15126/thesis.901800.
[26] National Institute of Standards and Technology, "*Advanced Encryption Standard (AES)*," FIPS PUB 197-upd1, May 2023.
[27] M. Dworkin, "*Recommendation for Block Cipher Modes of Operation: Galois/Counter Mode (GCM) and GMAC*," NIST Special Publication 800-38D, 2007, DOI: 10.6028/NIST.SP.800-38D.
[28] E. Rescorla, "*The Transport Layer Security (TLS) Protocol Version 1.3*," RFC 9846, Internet Engineering Task Force, July 2026.
[29] B. Vasey et al., "*Reporting Guideline for the Early-Stage Clinical Evaluation of Decision Support Systems Driven by Artificial Intelligence: DECIDE-AI*," Nature Medicine, Vol. 28, pp. 924-933, 2022, DOI: 10.1038/s41591-022-01772-9.
[30] X. Liu, S. Cruz Rivera, D. Moher, M. J. Calvert, A. K. Denniston, and the SPIRIT-AI and CONSORT-AI Working Group, "*Reporting Guidelines for Clinical Trial Reports for Interventions Involving Artificial Intelligence: The CONSORT-AI Extension*," Nature Medicine, Vol. 26, pp. 1364-1374, 2020, DOI: 10.1038/s41591-020-1034-x.