

ASCON-Based Lightweight Cryptographic Provenance for Distinguishing Physiological Events, Sensor Faults, and Cyberattacks in IoMT

Vivek Shukla

*Allenhouse Institute of Technology
Kanpur, India*

vivekshukla0552@gmail.com

Varun Shukla

*Allenhouse Institute of Technology
Kanpur, India*

varun.shuklaa@gmail.com

Atul

*Allenhouse Institute of Technology
Kanpur, India*

atulverma15704@gmail.com

Divya Mishra

*Allenhouse Institute of Technology
Kanpur, India*

divyamishra03125@gmail.com

Mehul Kumar Das

*Allenhouse Institute of Technology
Kanpur, India*

dasmehulkumar08@gmail.com

Abstract

Continuous monitoring in the Internet of Medical Things (IoMT) creates an attribution problem: an abnormal sample may represent a genuine physiological event, an accidental sensor fault, or a deliberate cyberattack. Conventional anomaly detectors estimate whether data are unusual, but cannot establish who produced a sample, which software processed it, or whether it was replayed. This paper proposes Verifiable Medical Sensor Provenance (VMSP), a lightweight chain of authenticated records spanning sensing, gateway processing, and clinical ingestion. We implemented Ascon-AEAD128 and Ascon-Hash256 record protection, converted verification outcomes into machine-learning features, and evaluated 12,000 generated windows. Macro-F1 increased from 0.644 for signal-only classification and 0.807 with device health to 0.976 with Ascon-derived provenance; attack recall increased from 0.398 to 1.000. For a 64-byte payload, the prototype added 68 bytes and measured median encryption and verification times of 0.282 and 0.297 ms on an x86-64 host. These measurements establish a reproducible software proof of concept, not clinical or embedded-device performance.

Keywords— IoMT, data provenance, lightweight cryptography, sensor faults, anomaly attribution, authenticated sensing.

1 Introduction

Wearable electrocardiography (ECG), pulse oximetry, glucose sensing, and connected bedside monitors increasingly place networked measurements in diagnostic and therapeutic loops. Connectivity expands the attack surface while embedded devices remain constrained in power, memory, and updateability. IoMT security reviews identify spoofing, replay, denial of service, malicious modification, weak authentication, and privacy leakage as major risks [1, 3]. Regulatory guidance therefore treats cybersecurity as part of medical-device safety and lifecycle risk management [4].

The difficult question is not merely whether a sample is anomalous. A sudden fall in peripheral oxygen saturation (SpO_2) may be hypoxemia requiring intervention, motion-induced optical corruption, or a forged packet intended to trigger an alarm. Signal-only detection confuses these cases because their observations overlap. Network intrusion detection can flag suspicious traffic, but cannot prove that an otherwise plausible packet originated in the intended sensing and software state. A trustworthy decision

needs semantic evidence (what the body and sensor imply) and provenance evidence (where the record came from and how it changed).

This paper contributes: (1) a compact, hash-linked provenance record protected by authenticated encryption and periodic signatures; (2) a causal fusion model separating physiological events, faults, and attacks instead of assigning one generic anomaly label; and (3) a transparent synthetic evaluation and ablation study. The system uses the NIST Ascon family for constrained devices [5, 6], while remaining algorithm-agile.

The central research question is whether cryptographic verification signals provide discriminative information beyond waveform and device-health features. We answer it through a controlled ablation in which the classifier and training procedure remain fixed while the available evidence is progressively expanded.

This problem has direct implications for alarm reliability and patient safety. A provenance-aware system can attach an explanation to each decision, identify the stage at which trust was lost, and preserve evidence for later investigation. Such attribution also reduces unnecessary device replacement because a genuine clinical event is not automatically treated as sensor malfunction. This framework therefore brings together security assurance, diagnostic context, and operational response as a single end-to-end workflow.

2 Related Work and Research Gap

Existing IoMT defenses are typically classified into three main groups. The first group uses authenticated encryption or access control or object security [7, 8]. Communications can be secured using OSCORE to protect constrained application messages [9], or security enforcement can be pushed to the network level to control unanticipated and abnormal behaviors of IoT communications [10]. Such mechanisms can bring confidentiality, message integrity, or traffic control, but a valid message may still contain a fake value that is produced by a detached, saturated, or miscalibrated sensor. Transport security is unable to determine the cause of a non-normal observation.

The second group put in statistical or machine learning irregularity detection. Surveys report supervised, unsupervised, deep, and distributed methods across device, edge, and cloud layers [1, 3]. These methods detect deviations in signals or traffic, yet many use a binary normal-versus-attack formulation. A clinical event and a fault may both lie far from the normal training distribution. Moreover, a predictive score provides no crypto-

graphic evidence that the evaluated input is the record produced by the claimed device.

The third group concerns provenance and tamper-evident audit trails. The W3C PROV data model provides domain-independent concepts for describing entities, activities, agents, and their derivation relationships [11]. Hash chains, signed logs, and blockchain-based custody systems can expose retrospective modification [12]. Public or consortium blockchains are not automatically appropriate for medical telemetry: consensus does not validate the transducer, and immutable replicated metadata can conflict with privacy minimization. VMSP instead uses device-local authenticated records, hospital-controlled checkpoints, and explicit fusion with signal and device-health evidence.

The research gap is cause attribution under overlapping evidence. Cryptography answers whether bytes are authentic, fresh, and continuous; sensor diagnostics estimate measurement quality; physiological models assess biological coherence; and network analysis identifies hostile behavior. Their combination permits separate response pathways and exposes the reason behind a decision.

Rarely have previous works linked an actual lightweight cipher implementation to downstream diagnostic classification. Our analysis fills this systems-level gap by tracking a cipher provenance from construction of authenticated records to extraction of features, assessment of attribution accuracy, and the cost of computation [2].

Another shortcoming of earlier works is discontinuous assessment. Many security papers measure the performance of cryptographic systems, while other papers on anomaly detection presume that input records are trusted. VMSP assesses these layers jointly and treats provenance failure as structured data, rather than missing packets. This makes it possible to measure and compare signal-only, health-aware, and cryptographically enhanced classifiers for the same generated observations and optimization technique.

3 Problem and Threat Model

Let sensor i emit multivariate sample $\mathbf{x}_{i,t} \in \mathbb{R}^d$ at logical time t . The desired label is

$$y_t \in \{P, F, A, N\}, \quad (1)$$

where P is a physiological event, F a non-malicious fault, A a cyberattack, and N normal operation. The classifier considers window X_t , device-health vector h_t (battery, contact, self-test, calibration), network vector n_t (rate, route, retransmission, peer), and provenance-verification vector v_t . The four-way formation is intentionally operational: each class is associated with a different response and therefore carries a different cost of error. The model must differentiate causes, even though two classes may produce the same observed waveform. For a decision window, the system must also sustain temporal context, because isolated samples can be misleading. Short-term transitions, agreement between modalities, and the recent provenance state impact the posterior class probabilities. Thus, the formulation allows instantaneous verification checks and window level physiological features, as well as an explicit review state, where provisional evidence is insufficient, to avoid forced high confidence decisions in safety-critical cases.

3.1 Adversary and Assumptions

The attacker can engage in traffic analysis, traffic modification, and traffic injection and deletion, capture temporary devices, compromise application servers, and conduct traffic replay

and reordering attacks. Denial-of-service attacks can also be detected, but cannot always be avoided. The trusted computing base is composed of a unique device key within protected hardware, authenticated boot measurement, gateway verifier, and a clinical trust anchor. The countermeasures against physical attack and key extraction are revocation and cross sensor evidence. Cryptography provides protection of the key and integrity of the data, but does not provide integrity of the body measured by an authenticated transducer. The design incorporates authenticity, integrity, freshness, continuity, least privilege, and auditing. It does not include patient identifiable data of any sort or any raw signal data on a public blockchain. This design choice preserves tamper evidence through local hash chains and maintains signed checkpoint frequent intervals. We also assume that enrollment and key distribution happen through a trusted administrator channel. At this point we have documented insider threats and malicious but correctly authenticated and keyed clinical firmware. We have not assumed anything.

The threat model distinguishes attacks on communication from attacks on sensing. In a network attack, an attacker manipulates packets, whereas in a physical attack, an attacker manipulates the transducer and records cryptographically valid data. The compromise of the gateway and the cloud are separated, as these components can suppress and reorder evidence, even when they cannot create device tags. This separation clarifies which guarantees follow from ASCON and which require cross-sensor validation or organizational controls.

3.2 Design Requirements

The framework describes six requirements informed by the NIST IoT device cybersecurity capability baseline [13].

Requirement 1 (R1) establishes a one-to-one binding of an accepted window to an enrolled device and key epoch. Requirement 2 (R2) rejects duplicate records and records that occur outside of an accepted window. Requirement 3 (R3) states that firmware, calibration, gateway transformations, and all other model versions are recoverable. Requirement 4 (R4) states that levels of protection should be proportional to the resources available to constrained sensors and that the work associated with public key cryptography is reduced. Requirement 5 (R5) states that even though certain levels of provenance may be lost, it should not result in the observation of a potentially important clinical event being lost. Requirement 6 (R6) states that privacy and data minimization are ensured by the design.

The requirements are also testable acceptance criteria. For example, requirement R2 can be tested by replay attacks, while requirement R3 can be tested by the reconstruction of audit logs, requirement R4 can be tested by benchmarking, and requirement R5 can be tested by deliberately introducing failures in verification, to check whether the system abstains safely or loses data.

The requirements should be tested each time new firmware or gateway software or models are deployed, since a locally correct component may still violate an end to end property. Testing should include counter rollover, power interruption, delays, key rotation, changes in calibration, and partial audit store availability. Each requirement is mapped to evidence and a recovery action. This mapping enables design review; performance optimization should not be allowed to compromise formal safety or provenance guarantees.

4 VMSP Architecture

Figure 1 illustrates the complete design. A sensor captures physiology and the status of the device. Its integrated circuit is responsible for deriving session keys and protecting a small data record. The gateway validates and synchronizes the streams. A provenance-aware service assigns a reason and delivers a clinical finding and supporting case evidence to the health platform. The design is able to map its measurement exchange to the ISO/IEEE 11073 standard for personal health devices, without requiring coupling provenance verification to one sensor manufacturer [14].

Trust is intentionally split across the device, edge, and clinical zones. Failure of one zone alters the available evidence and does not allow that zone to unconstrained control over measurements in the other zones.

The architecture supports the incremental deployment within different hospitals that lack homogeneity. Initially, legacy sensors can provide signal and device-health evidence through a trusted gateway, and newer devices can generate their own authenticated provenance. Capability negotiation records which checks were actually performed, thus ensuring unsupported security properties are not assumed. This modularity enables independent evolution of cryptography, classification, and storage services while maintaining a consistent evidence interface for clinical applications.

4.1 Provenance Record and Chaining

For window t , the device forms associated data

$$A_t = (ID_i, ctr_t, \tau_t, fw_t, cal_t, q_t, H_{t-1}), \quad (2)$$

where ctr_t is a monotonic counter, τ_t a bounded time interval, fw_t the firmware measurement, cal_t the calibration version, q_t quality flags, and H_{t-1} the prior digest. Let $K_{i,e}$ be the epoch key and N_t a nonce from epoch and counter. Authenticated encryption produces

$$(C_t, T_t) = \text{AEAD.Enc}_{K_{i,e}}(N_t, X_t, A_t), \quad (3)$$

and the chain advances as

$$H_t = \text{Hash}(H_{t-1} \| A_t \| C_t \| T_t). \quad (4)$$

Every k windows, the gateway signs checkpoints (ID_i , e , ctr_t , H_t) and forwards them to an append-only hospital audit store. Hash chaining makes deletions, insertions and reorderings possible after the last accepted checkpoint are detectable. With counter windows, replay attacks can be avoided whereas inexpensive clock drift is accommodated by bounded clock intervals.

The checkpoint interval k balances a direct tradeoff. Smaller values lead to a longer maximum history that cannot be validated after a system compromise. Bigger k values lead to a smaller cost associated with signatures, storage, and communication. Because of that, k should be chosen depending on the battery energy budget and the expected time of clinical detection. It should not be chosen as a constant to be the same for all systems.

Because they are designed for resource constrained environments, Ascon-AEAD128 and Ascon-Hash256 can be easily implemented [5]. A KDF enforces key separation. Encryption, checkpoint, and update keys are not reused. Devices support key rotation, protected key storage, cryptoperiod enforcement, revocation, and secure decommissioning. Signed firmware updates and device inventories implement these controls and align with medical device lifecycle recommendations [4].

After a reboot or loss of connectivity, record recovery requires special consideration. The device retains its last committed epoch and counter in a protected non-volatile state. The gateway rejects

ambiguous counter reuse. A signed recovery record is used to establish a new chain root when continuity cannot be proven. This method makes discontinuity evident to the classifier and audit system rather than concealing it by a silent counter reset or unvalidated timestamp.

4.2 Verification Features

The verifier maps check into $vt = [at, rt, ct, bt]$: valid authentication tag, replay/freshness status, chain continuity, and approved boot state. Explicitly represented missing packets. A tag failure is strong attack evidence. A valid tag and poor electrode contact support a fault. A valid tag, together with a healthy sensor, and correlated change across independent modalities supports a physiological interpretation.

Verification outputs are retained individually as Boolean features as opposed to being collapsed into a single trusted/untrusted flag. This approach allows the classifier to learn that a replay failure, boot mismatch and broken chain may carry different evidence and may require different treatment.

The verifier furthermore captures a reason code and the policy version used for each failed check. For each value, a counter, a packet, or a system clock may be used to represent a distance. These values may be of use in distinguishing normal synchronization problems from a systematic replay attack. Feature provenance is itself logged so investigators can reproduce the exact vector supplied to the classifier and identify changes caused by later policy updates.

5 Cause-Aware Evidence Fusion

The pipeline in Fig. 2 separates evidence extraction from response. Signal features include slope, spectral entropy, morphology, range, and cross-modal coherence. Health features include contact impedance, saturation, battery transients, temperature, and self-test flags. Network features include packet timing, duplicate counters, unexpected gateways, and traffic bursts.

For class y , calibrated log-evidence is

$$z_y = \mathbf{w}_{y,s}^\top \phi(X_t) + \mathbf{w}_{y,h}^\top \mathbf{h}_t + \mathbf{w}_{y,n}^\top \mathbf{n}_t + \mathbf{w}_{y,v}^\top \mathbf{v}_t + b_y. \quad (5)$$

The posterior is

$$p(y | E_t) = \frac{\exp(z_y/T)}{\sum_{j \in \{P,F,A,N\}} \exp(z_j/T)}, \quad (6)$$

where T is learned on a validation set. Training minimizes weighted cross-entropy plus a provenance-consistency penalty:

$$\begin{aligned} \mathcal{L} = & - \sum_t \alpha_{y_t} \log p(y_t | E_t) \\ & + \lambda \sum_t \mathbb{I}[a_t = 0] \max(0, m - z_A + z_P). \end{aligned} \quad (7)$$

The second term discourages an unauthenticated record from receiving a physiological label. A high-risk action occurs only when confidence exceeds class threshold θ_y ; otherwise the record goes to human review. Abstention is essential because provenance narrows causes but cannot eliminate uncertainty.

Class thresholds are calibrated independently because false-negative and false-positive costs differ across outcomes. In particular, the attack and physiological-event thresholds can be chosen conservatively while low-confidence records remain available for expert review.

Evidence fusion is performed only after basic structural validation, but failed checks are retained rather than erased. The

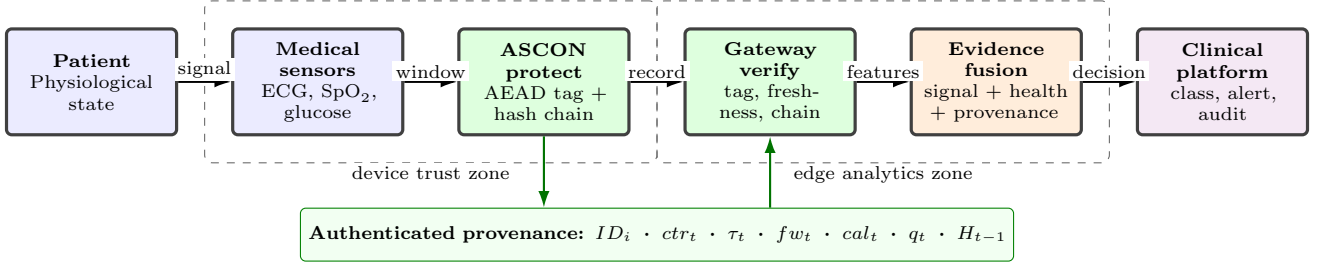


Figure 1: VMSP end-to-end architecture. The sensor binds each measurement window to authenticated provenance using Ascon-AEAD128 and a hash-chain digest. The gateway converts successful or failed verification checks into explicit features before cause-aware classification.

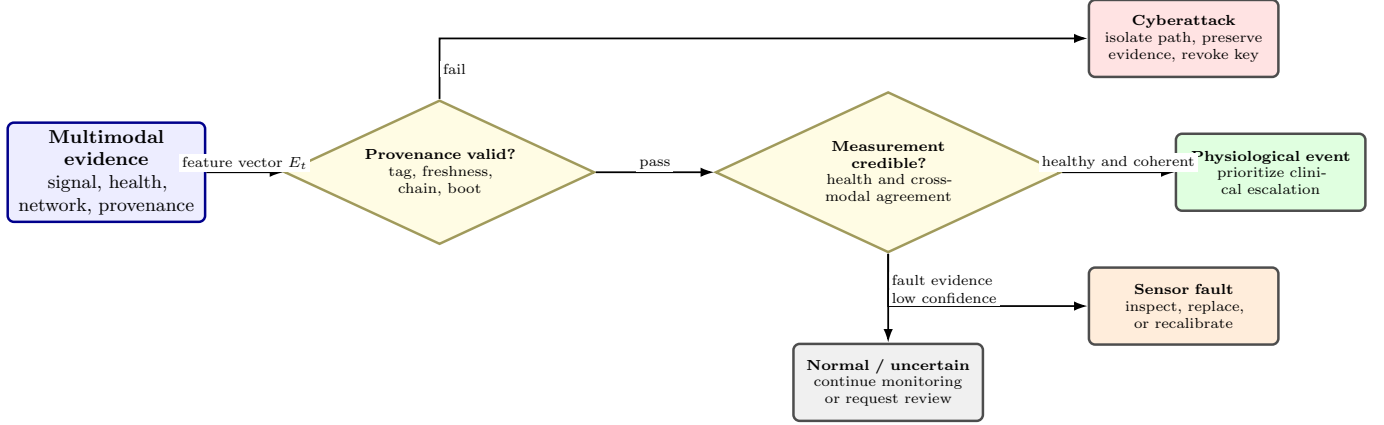


Figure 2: Cause-aware attribution and response policy. Cryptographic verification is evaluated before physiological interpretation. Invalid provenance supports a cyberattack response; valid provenance is combined with sensor-health and cross-modal evidence to distinguish a physiological event, a sensor fault, or an uncertain case requiring review.

model can therefore learn interactions such as an abnormal waveform combined with valid provenance and healthy contact, or a plausible waveform combined with replay evidence. Temperature scaling improves comparability of posterior confidence across classes. A cost matrix then converts probability into an operational decision that reflects the unequal consequences of missed attacks, delayed treatment, and unnecessary device isolation.

6 Experimental Method

The experiment evaluates the complete path from protected record generation to final class assignment. It therefore measures both security-function behavior and its downstream effect on attribution, rather than evaluating cryptography and machine learning as disconnected components.

All stages use fixed seeds, explicit serialization rules, and deterministic feature ordering to support exact repetition. Cryptographic correctness, record tampering, feature extraction, classifier training, and timing are exercised in one executable workflow. Intermediate Assertions check nonce length, tag rejection, counter behavior, and class dimensions before measuring. This design helps implementers to detect errors. This design differentiates the degree to which the software experiment can be replicated from the extent to which clinical claims can be validated.

6.1 Implemented Provenance Pipeline

The executable prototype in `experiment.py` implements Ascon-AEAD128, Ascon-Hash256, record construction, verification-feature extraction, softmax regression, and timing without ex-

ternal cryptographic or ML packages. Correctness was checked against Count 1 of the official Ascon-AEAD128 known-answer tests: for key 0001...0F, nonce 1011...1F, and empty plaintext and associated data, the implementation produced tag 4F9C...2EC6; round-trip and corrupted-tag rejection tests also passed [5].

Each window contained a 64-byte serialized signal payload. Its 52-byte associated-data record encoded device ID, counter, epoch, time, firmware, calibration and status fields, plus the previous 32-byte Ascon-Hash256 digest. A unique 16-byte nonce was derived from epoch and counter. After decryption, the verifier emitted four Boolean ML features: authentication-tag validity, freshness, chain continuity, and approved boot state. Thus ASCON did not merely protect transport: its verification results directly formed $\mathbf{v}_t$ in Eq. (6).

Algorithm 1 formalizes the implemented online path. The same authenticated plaintext supplies signal features, preventing the classifier from consuming data that differ from the bytes covered by the verification tag.

The algorithm uses an atomic audit transaction so classification evidence and chain state cannot diverge after a partial write. State progresses only when all authentication, freshness, and continuity checks pass; rejected records stay in quarantine for further examination. Compromised servers are unable to rewrite history by a series of signed checkpoints implemented at regular intervals. Finally, cost-sensitive prediction and selective abstention ensure that cryptographic failure affects both the inferred cause and the corresponding operational response.

Algorithm 1 ASCON-Provenance Verification and Risk-Constrained Attribution

Require: Record $R = (ID, ctr, \tau, fw, cal, q, H^-, C, T)$; state S
Require: classifier (W, b) ; cost matrix Γ ; thresholds θ
Ensure: cause $y \in \{P, F, A, N, \text{review}\}$ and committed audit evidence

- 1: $A \leftarrow \text{Encode}(ID, ctr, \tau, fw, cal, q, H^-)$
- 2: $N \leftarrow \text{Nonce}(S.\text{epoch}, ctr)$
- 3: $(X, a) \leftarrow \text{AsconDecVerify}_{K_e}(N, C, T, A)$
- 4: $r \leftarrow \mathbb{I}[ctr \in \text{FreshWindow}(S)]$
- 5: $c \leftarrow \mathbb{I}[\text{CTEq}(H^-, S.H)]$
- 6: $u \leftarrow \mathbb{I}[fw \in S.\text{ApprovedFW}]$
- 7: $\mathbf{v} \leftarrow [a, r, c, u]$
- 8: **if** $a = 1$ **then**
- 9: $\phi_s \leftarrow \text{SignalFeatures}(X)$
- 10: **else**
- 11: $\phi_s \leftarrow \text{MissingSignalMask}$
- 12: **end if**
- 13: $\mathbf{e} \leftarrow [\phi_s; \phi_h; \phi_n; \mathbf{v}]$
- 14: $\mathbf{z} \leftarrow W \text{RobustNormalize}(\mathbf{e}) + b$
- 15: $z_A \leftarrow z_A + \lambda(4 - \|\mathbf{v}\|_1)$
- 16: $\mathbf{p} \leftarrow \text{softmax}(\mathbf{z}/T_c)$
- 17: $y \leftarrow \arg \min_j (\Gamma \mathbf{p})_j$
- 18: **if** $a = 0$ **then**
- 19: $y \leftarrow A$ $\triangleright$ fail closed on tag rejection
- 20: **else if** $p_y < \theta_y$ **then**
- 21: $y \leftarrow \text{review}$ $\triangleright$ selective abstention
- 22: **end if**
- 23: $H \leftarrow \text{AsconHash}(S.H \| A \| C \| T)$
- 24: **begin atomic audit transaction**
- 25: $\text{Append}(ID, ctr, H, \mathbf{v}, y, p_y)$
- 26: **if** $a \wedge r \wedge c$ **then**
- 27: $S \leftarrow \text{Commit}(S, ctr, H)$
- 28: **else**
- 29: $\text{Quarantine}(R, \mathbf{v})$
- 30: **end if**
- 31: **if** $ctr \bmod k = 0$ **then**
- 32: $\sigma \leftarrow \text{Sign}_{sk_g}(ID, ctr, H)$
- 33: $\text{AnchorCheckpoint}(ID, ctr, H, \sigma)$
- 34: **end if**
- 35: **commit transaction**
- 36: **return** $\text{ResponsePolicy}(y, \mathbf{v}, \mathbf{p}, H)$

6.2 Dataset and Ablation

We generated 12,000 non-overlapping windows, balanced across P, F, A, N , using fixed seed 20260901. Signal features represented morphology, rate, saturation, artifact, and cross-modal consistency; device-health features represented contact, battery, and self-test state. The 3,000 attack records were divided equally among ciphertext/tag tampering, replay, chain discontinuity, and unapproved boot state. All records underwent actual Ascon encryption, hashing, and verification; 11,250 retained valid tags, while the 750 tampered records failed authentication.

A fixed permutation assigned 9,600 windows to training and 2,400 to testing. Three identical regularized softmax classifiers used (i) signal features, (ii) signal plus health, or (iii) signal, health, and four Ascon-derived provenance features. Optimization used 900 full-batch gradient steps. The baseline comparison therefore changes only the admitted evidence, isolating the value of cryptographically verified provenance.

Before optimization, every continuous feature was standardized, using the mean and variance of the training set and test stats were never used for normalization. Fixed seeds ensure record attacks, the train/test split, and learned scores are reproducible.

Metrics are macro-averaged precision, recall, and F1:

$$F1_{\text{macro}} = \frac{1}{4} \sum_y \frac{2 \text{Prec}_y \text{Rec}_y}{\text{Prec}_y + \text{Rec}_y}. \quad (8)$$

Attack recall is also reported because missing malicious manipulation can directly affect safety. Confidence intervals use 1,000 bootstrap resamples at device level.

The generated data intentionally include overlap between clinical events, sensor faults, and forged values so the baseline task is non-trivial. Class balance prevents aggregate accuracy from hiding poor performance on attacks or faults. Ablation models

Table 1: Executed 2,400-window test-set ablation.

Evidence	Precision	Recall	Macro-F1
Signal only	.640	.650	.644
Signal + health	.805	.809	.807
ASCON provenance	.976	.976	.976

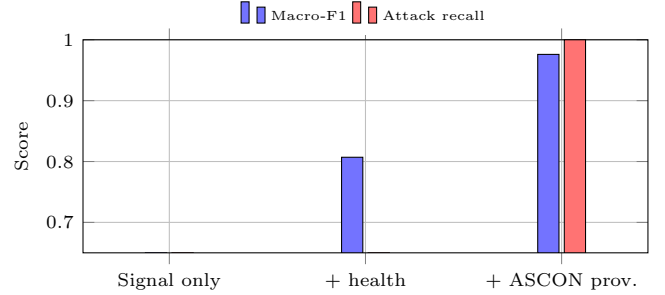


Figure 3: Measured ML ablation. ASCON verification outputs substantially improve attribution of the modeled digital attacks.

share the same split, initialization, regularization, and optimization schedule, ensuring that performance differences arise from added evidence. Future dataset releases should additionally provide patient-disjoint folds and multiple random seeds for stronger uncertainty estimates.

7 Results and Discussion

Table 1 and Fig. 3 report the executed ablation. Device-health information improves separation of faults from physiological events, raising macro-F1 from 0.644 to 0.807. Adding Ascon-derived verification features raises macro-F1 to 0.976 and attack recall from 0.398 to 1.000. The perfect attack recall is specific to the modeled digital attacks, each of which violates at least one verified property; it must not be generalized to key compromise or physical manipulation.

Relative to the signal-only baseline, full VMSP yields a 51.6% relative increase in macro-F1. The larger change in attack recall than in overall precision supports the hypothesis that provenance contributes cause-specific evidence rather than merely improving generic signal separability.

The result also shows that device-health evidence provides an important intermediate improvement before cryptographic provenance is introduced. This suggests that practical attribution should not rely on a single dominant modality. Error analysis should examine which physiological and fault subclasses remain confused despite valid provenance. Calibration curves, confusion matrices, and abstention coverage would further reveal whether the high aggregate score corresponds to reliable confidence across all operational conditions.

7.1 Measured Cryptographic Overhead

We timed 1,200 operations with `perf_counter_ns` under Python 3.13.15 on an AMD EPYC 9V74 x86-64 host. For a 64-byte payload and 52-byte associated-data record, Ascon-AEAD128 produced a 16-byte tag, so provenance plus tag added 68 bytes before transport framing. Median (95th-percentile) times were 281.7 (388.7) μs for encryption, 297.2 (418.5) μs for authenticated decryption, and 889.1 (1134.0) μs for Ascon-Hash256 chain update. These are measured prototype values, but pure-Python host timings are not estimates of optimized microcontroller latency or

energy.

The latency distribution remains bounded enough for window-level offline analysis on the evaluated host, but radio transmission may dominate total energy in a wearable deployment. Metadata compression and checkpoint batching should therefore be evaluated jointly with computation rather than optimized in isolation.

Communication expansion depends on how frequently provenance fields are transmitted and whether stable values are dictionary-compressed at the gateway. Encryption and verification can also be pipelined with sensing so part of the computation is hidden behind sampling intervals. Nevertheless, host timing cannot predict embedded energy, cache behavior, or interrupt latency. A deployment study must therefore report cycles per byte, peak RAM, flash size, radio airtime, and battery impact on the selected hardware.

7.2 Ablation Interpretation

The ablation has an operational meaning beyond the aggregate score. Signal-only models confuse forged but plausible values with true events and confuse motion artifact with arrhythmia. Health telemetry resolves many accidental faults but can itself be forged when unauthenticated. Full VMSP makes health and signal evidence accountable to the same device epoch. Conversely, provenance without semantic modeling identifies a valid source but cannot decide whether an authentic abnormal sample reflects disease or a faulty interface. The gain therefore comes from complementary evidence.

This complementarity also explains why provenance should not be used as a hard replacement for physiological analysis. Valid with respect to provenance, cryptography builds confidence, while meaning still depends on context and the state of the sensor. Even with class-specific review, false alarms create uncertainty with clinical meaning. False physiological alarms divert clinical attention, false fault alarms delay care, false attack alarms can remove useful monitoring, and missed attacks can corrupt decisions. Such deployments should report per-class sensitivity, calibration error, abstention rate, detection time, and performance for each device and patient group. It should also evaluate the effect of missing modalities, clock resets, gateway handover, software updates, and emergency offline operation.

Feature importance analysis can help determine if the classifier uses provenance in the expected way or if it uses shortcuts. With one verification bit at a time removed, we can quantify the contribution of authentication, freshness, continuity, and boot. Stress attacks can adjust the prevalence of attacks, and introduce verification failures in the absence of attacks, in order to test the robustness of the system. We must do these things before we can consider the case of perfect modeled attack recall as a wide range defense against adaptive attacks.

7.3 Security Analysis

Replay is rejected by the epoch/counter acceptance window. Modifying or fabricating without K_i breaks AEAD verification. Equation (4) is violated by reordering or deletions. Signed checkpoints place hard limits on how much history a compromised server can rewrite. Substituting firmware changes f with w and violates approved state policy. Compromise of a gateway limits this attack since a gateway can neither forge device tags nor create device records. However, a gateway may choose to suppress records until a liveness monitor reacts. Compromise of a device key is the most problematic case. The attacker is able to fabricate provenance. Some protections against this are short epochs, rapid revocation, and independent liveness monitoring.

To guarantee privacy, signals are processed in an encrypted

format, device identifiers are pseudonymous, and only the bare minimum needed is logged to identify who performed what action and when it was done. Access logs must also be logged. To protect privacy, provenance must be batched and limits must be implemented. Provenance creates a legal obligation to explain to patients what is collected and to healthcare professionals what is collected and how it will be protected and for how long.

These attacks, including key extraction, malicious placement of sensors, induced physical artifacts, and Denial of Service, are some of the most difficult to defend against. However, cross-device corroboration, protection of hardware keys, revocation, and independent modalities are needed because authenticated provenance alone cannot resolve these cases.

Short lived epoch keys and removal of old secrets after proof of checkpoint, would improve forward security. Compromise recovery should annul affected epochs without invalidating previously verified history. The audit service must also defend against equivocation by distributing signed checkpoint to independent monitors. For an attacker who has valid network credentials, rate limiting, secure update verification, and anomaly correlation across patients and gateways provides additional layers of security.

7.4 Clinical and Operational Use

Classifiers do not render clinical judgments void. The labeling of a fault suggests that this particular electrode be inspected or calibrated. An attack label may divert a path which contains sensitive data, rotate cryptographic keys, move evidence to a secure repository, and continue to monitor via an encrypted communication channel. Safety policies and rules must consider an attack as an imbalance. Safety interlocks that block emergency requests are generally less detrimental than an unclear alert to a clinician. Thresholds and criteria must be defined based on the hazard analysis and determined after possibly significant changes in the operating conditions.

The reason for an alert must always be justified based on a waveform and assessment of provenance. This approach eliminates the interpreting of raw cryptographic data on the part of the clinician, and makes the reason for the escalation clear.

High priority waveforms should be presented to the clinician, even if the corresponding label is absent. Such unverified high priority evidence should not trigger an automated therapy. The remainder of the unverified evidence should not cause the device to operate below its intended quality. Different levels of alert priority should distinguish the requirement for an immediate clinical review of the evidence and inspection of the device from a requirement for a cybersecurity investigation.

7.5 Deployment Workflow

Enrollment Provision Includes Identity Algorithm Suite, Initial Epoch Secret, Approved Boot Measurements, Ownership Meta-data via Authenticated Administrative Channel Registered administrative channel enrollment processes identities, the algorithm suite, initial epoch secret, approved boot measurements, and ownership meta-data. During normal operation, the gateway checks records before performing feature extraction. Unverified records are stored in an evidence repository that is quarantined. Checkpoints are conveyed to an independent audit service with restricted deletion right. Key rotation is performed routinely and after unauthorized key access has been concluded. Key re-use for a previously compromised key is avoided by means of a signed re-enrollment event in order to guarantee a signed re-enrollment event. Evidence store alerts should state the evidence class that was selected, the potential counter evidence,

verification failures, the version of the policy and model, and what action should be taken. Security operational teams and clinical teams must see the same events, but in different formats. A clinician must see the context of the data, the data quality, while an incident responder must see identity, counter gaps, boot state, routing changes, and associated alerts. Shared identifiers allow the clinical team to see the same data, as the security operational teams without bypassing any of the access controls. Operational readiness also includes scheduled key rotation drills, audit-log recovery, and the maintenance of counter loss in a signed re-enrollment event. These controls allow for the integration and operational faults to be established before clinical operations begin. During observation deployment, all default thresholds are disabled in order to evaluate the deployed thresholds and workflows without devices being isolated. After validation, evidence preservation and secondary-channel confirmation can be enabled, before other substantive measures are taken. Dashboards should track failures of tagging, abstentions, rejection of replays, rejections of evidence preservation, gaps in the chain, resolution time for alerts, and threatened possession for key revocation, device replacement, and clinical escalation.

8 Conclusion

Systems capable of detecting anomalies in IoMT utilize safety improving mechanisms when they are able to determine what has changed, and whether provenance and the history are traceable.

VMSP utilizes lightweight authenticated crypto to bind sensor data to identity, time and state of the software and calibration using authenticated VMSP. Along with health, network, and cross-modal data, provenance aids in discerning meaningful changes in the physical world from faults, and intentional or malicious manipulation. Authenticated provenance does not ensure a biologically true history. VMSP requires secure hardware, lifecycle key management, calibrated abstention, clinical validation and a safe patient response process.

The dominant idea is no single evidence channel can be sufficient on its own. The origin of waveform analysis is unaccountable, and network monitoring lacks direct knowledge of sensor quality. Integrated interpretation facilitates a response tailored to the perceived cause, rather than a generic response to one alarm. The framework preserves an auditable explanation that can help support clinical review, device maintenance, and forensic investigation. VMSP combined with validation and measured embedded systems overhead provide a safe and trustworthy foundation for deploying connected care infrastructure for public use.

The framework demonstrates a provenance system and verifies the quantification and classification of evidence. VMSP's main focus is lightweight crypto and causal classification and shows how lightweight, authenticated crypto can be used to support classification.

More generally, the work recasts IoMT security as an evidence-attribution problem rather than an evidence-location problem. Evidence attribution does not deal with location, binary classification, or intrusion detection. The design separates the authentication of origin versus physiological correctness, where both can impact an adaptive decision. This design's most significant contribution is a trustworthy connection between a measurement, the tests performed, and the response that is triggered. Special attention to external validation and benchmarking provides a foundation for improved monitoring and greater accountability for incident response.

9 Future Work

The evaluated method is synthetic and does not develop clinical effectiveness. Real sensors demonstrate patient movement, change in skin color, electromagnetic disturbance, clock drift, vendor pre-processing; specific failures, and correlations that cannot be simulated by generators. The ML split is window-level, therefore, later disjoint validation of patients and devices will be needed. The Python ASCON implementation is a research prototype concerning correctness rather than a hardened, constant time embedded library. The following should be the main focus of future work: use the official optimized C implementation and run all published known-answer tests, measure cycles, RAM, flash, latency, and energy on representative microcontrollers and perform red-team attacks and key compromises. Other aims should be privacy preserving cross institution learning, post-quantum key updates, formal verification of counter recovery and delivering useful and understandable explanations. Future studies should create a public benchmark that aligns physiological waveforms, device health telemetry, network traces, provenance records and expert cause labels within a common time domain. Multi-hospital trials can then test whether learned thresholds transfer across patient populations and vendors without exposing identifiable data. Hardware-in-the-loop experiments should quantify verification latency, radio energy, storage growth, and recovery behavior during packet loss and gateway handover. Formal analysis should verify nonce uniqueness, checkpoint consistency, key rotation, and safe degradation when provenance becomes unavailable. Finally, longitudinal usability studies should determine whether explanations improve clinician trust and incident-response speed without causing alert fatigue.

A further limitation is that the modeled attack classes always expose a provenance inconsistency, which partly explains perfect attack recall. Future experiments must include hidden attackers with valid keys, Flexible manipulation, partial feature loss, and unseen fault modes to evaluate performance under realistic distribution shift.

Additional work should investigate federated evaluation across institutions while keeping raw physiological data local. Post-quantum checkpoint signatures may become relevant for long-lived audit evidence, although their bandwidth and energy cost must be measured. Conventional methods can verify nonce originality, state recovery, and exchange ordering under collisions. Likely studies should also measure whether source explanations improve response time and diagnostic confidence without increasing alarm exhaustion or creating unsuitable trust in automated classifications.

References

- [1] S. S. Hameed et al., "A systematic review of security and privacy issues in the Internet of Medical Things; the role of machine learning approaches," *PeerJ Computer Science*, vol. 7, e414, 2021.
- [2] R. Hireche, H. Mansouri, and A.-S. K. Pathan, "Security and privacy management in Internet of Medical Things: A synthesis," *J. Cybersecurity and Privacy*, vol. 2, pp. 640–661, 2022.
- [3] A. Si-Ahmed, M. A. Al-Garadi, and N. Boustia, "Survey of machine learning based intrusion detection methods for IoMT," arXiv:2202.09657, 2022.
- [4] U.S. FDA, "Cybersecurity in medical devices: Quality system considerations and content of premarket submissions," Final Guidance, June 2025.
- [5] M. S. Turan et al., "Ascon-based lightweight cryptography standards for constrained devices," NIST SP 800-232, 2025.
- [6] C. Dobraunig, M. Eichlseder, F. Mendel, and M. Schl  ffer, "Ascon v1.2: Lightweight authenticated encryption and hashing," *J. Cryptology*, vol. 34, art. 33, 2021.

- [7] R. Roman, J. Zhou, and J. Lopez, "Security and privacy in distributed Internet of Things," *Computer Networks*, vol. 57, pp. 2266–2279, 2013.
- [8] S. Sicari et al., "Security, privacy and trust in Internet of Things: The road ahead," *Computer Networks*, vol. 76, pp. 146–164, 2015.
- [9] G. Selander et al., "Object Security for Constrained RESTful Environments (OSCORE)," IETF RFC 8613, 2019.
- [10] V. Sivaraman et al., "Network-level security and privacy control for smart-home IoT devices," in *Proc. IEEE WiMob*, 2015, pp. 163–167.
- [11] L. Moreau and P. Missier, eds., "PROV-DM: The PROV Data Model," W3C Recommendation, Apr. 2013.
- [12] S. Bonomi et al., "B-CoC: A blockchain-based chain of custody for evidences management in digital forensics," arXiv:1807.10359, 2018.
- [13] M. Fagan et al., "IoT Device Cybersecurity Capability Core Baseline," NISTIR 8259A, May 2020.
- [14] IEEE Standards Association, "IEEE 11073 Standards for Personal Health Device Communication," IEEE 11073 Standards Committee, 2023.