JAIHBE	Original Research Article
Journal of AI in Healthcare and Biomedical Engineering	Vol. 3, No. 4 (2025), pp. 1-20
ISSN 3068-1197 | Open Access | Peer-Reviewed	DOI: 10.63646/jaihbe.2025.030401

Privacy-Preserving Federated Intelligence for Healthcare IoT: A Secure Biomedical Engineering Framework for Real-Time Patient Monitoring

Tomasz J. Wiśniewski[1], Magdalena A. Kowalczyk[2], Paweł R. Lewandowski[3, *]

[1] Department of Biomedical Engineering, Faculty of Electrical Engineering, Bialystok University of Technology, Białystok, Poland. Email: t.wisniewski@pb.edu.pl

[2] Department of Computer Science and Telecommunications, Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering, AGH University of Krakow — Centre of Energy, Kraków, Poland. Email: m.kowalczyk@agh.edu.pl

[3] Institute of Telecommunications and Cybersecurity, Faculty of Electronics and Information Technology, Lublin University of Technology, Lublin, Poland. Email: p.lewandowski@pollub.edu.pl

*** Corresponding Author. Email: p.lewandowski@pollub.edu.pl**

ARTICLE INFO

Received
22 July 2025
Revised
26 September 2025
Accepted
28 November 2025
Available Online
30 December 2025

Keywords
Federated learning
Healthcare IoT
Blockchain
Privacy preservation
Bi-LSTM
Smart contracts
Patient monitoring

Abstract

Healthcare Internet of Things (H-IoT) deployments now stream continuous physiological data from millions of wearable devices, bedside monitors and implanted sensors, but the centralised machine-learning pipelines that analyse these streams expose patients to systemic privacy risk and single-point-of-failure attacks. We describe a secure biomedical-engineering framework that delivers real-time patient monitoring without exposing raw clinical signals. The framework couples on-device training of a Bidirectional Long Short-Term Memory (Bi-LSTM) anomaly detector to a federated averaging layer, which is in turn anchored to a permissioned Proof-of-Stake blockchain through PBKDF2-derived authentication keys, AES-GCM-encrypted gradients, and smart-contract-mediated aggregation. A key contribution is the explicit decoupling of the cryptographic verification path from the model-update path, which lets the system tolerate Byzantine clients without sacrificing convergence speed. We evaluate the framework on the public ToN-IoT and CICIDS2019 intrusion-detection corpora, treating the attack-classification task as a proxy for monitoring-stream integrity, and report a mean accuracy of 96.42 % on ToN-IoT and 97.38 % on CICIDS2019, an F1 of 0.969, an AUC of 0.985, a false-positive rate of 2.18 %, and a per-round end-to-end latency of approximately 5.2 s on a 10-validator network. An ablation isolates the contribution of each component: removing blockchain anchoring lowers accuracy by 3.26 percentage points and the security score from 98 to 72; removing PBKDF2 reduces accuracy by 0.91 points; removing on-device encryption collapses the security score to 41 even though detection accuracy is preserved. We further analyse the energy and scalability envelope of the consensus layer, showing that the Proof-of-Stake choice scales linearly with validator count up to 32 nodes whereas an equivalent Byzantine-Fault-Tolerant deployment scales quadratically. The framework is therefore a practical route to trustworthy, privacy-preserving, real-time biomedical analytics that satisfies HIPAA and GDPR audit requirements without sacrificing clinical responsiveness.

INTRODUCTION

Continuous physiological monitoring outside the clinic has transitioned from a research aspiration to a routine deployment scenario. Smart wearables now report cardiac rhythm, oxygen saturation, glycaemia, and accelerometer-derived activity at sampling rates that would have required a hospital-grade installation a decade ago, while bedside and implanted devices feed the same patient timeline at far higher fidelity (Karatas et al., 2022; Aguileta et al., 2024). The convergence of these data streams under the heading of Healthcare Internet of Things (H-IoT) has reshaped clinical workflows: early warning of decompensation in heart failure, automated escalation in sepsis, post-operative remote follow-up, and longitudinal management of chronic disease are increasingly delivered as software services that consume continuous biomedical telemetry (Manickam et al., 2022; Akhbarifar et al., 2023; Boikanyo et al., 2023).

The price of this convergence is a radical re-architecture of the threat surface. Every additional sensor that joins the H-IoT network adds another path by which sensitive physiological data can be exfiltrated, replayed, poisoned, or used to infer non-medical attributes about its bearer (Ghubaish et al., 2021; Hatzivasilis et al., 2019). Conventional centralised analytics — in which raw signals are uploaded to a cloud aggregator that trains a deep model and serves an inference endpoint — concentrate this risk at exactly the point where compromise has the greatest blast radius. The same architecture is increasingly difficult to reconcile with the privacy regimes that govern healthcare data: the European General Data Protection Regulation, the United States Health Insurance Portability and Accountability Act, and analogous frameworks elsewhere all require explicit demonstrations that personal health information is minimised, purpose-bound, and revocable (Kaissis et al., 2020; Yaqoob et al., 2022).

Federated learning has emerged as the most plausible answer to the data-minimisation requirement. By transmitting only model parameters or gradients, a federation of clients can collaboratively train a shared model without ever centralising raw observations (McMahan et al., 2017; Rieke et al., 2020). Healthcare federations have been demonstrated for medical imaging, electronic health records, and wearable analytics (Rieke et al., 2020; Pfitzner et al., 2021; Sheller et al., 2020). However, federated learning alone does not solve the trust problem. The aggregator is still a single point of failure; gradients are still amenable to inversion attacks that recover training samples (Zhu et al., 2019; Geiping et al., 2020); and a malicious client can still poison the global model by submitting carefully crafted updates (Bagdasaryan et al., 2020; Tolpegin et al., 2020). In healthcare, where a poisoned arrhythmia detector or a leaked ECG can have direct clinical consequences, these residual risks are not academic.

Blockchain systems address a complementary set of weaknesses by replacing the single trusted aggregator with a permissioned ledger maintained by multiple validators (Lu, 2019; Lu, 2022; Xu et al., 2021). On-chain consensus provides tamper-evident records of which gradient was contributed when and by whom; smart contracts encode aggregation policy that is auditable rather than implicit; and the immutability of the ledger supports the audit trails required by HIPAA and GDPR (Hasselgren et al., 2020; Tanwar et al., 2020). The cost is throughput: on-chain settlement is slower than in-memory aggregation, and naive integration imposes per-update overhead that quickly dominates the training loop (Qu et al., 2020; Kang et al., 2020).

The work described in this paper proposes a privacy-preserving federated-intelligence framework that

ISSN: 3068-1197

couples the two ideas in a way that pays the throughput cost only where it is actually needed. The framework deliberately decouples the gradient channel from the verification channel: encrypted gradients flow continuously through the federated aggregator, while only the corresponding hashes and policy events are written to the chain. Authentication is anchored in a Password-Based Key Derivation Function 2 (PBKDF2) construction that produces device-specific keys without exposing them to the network, and the on-device classifier is a Bidirectional Long Short-Term Memory (Bi-LSTM) network sized for resource-constrained microcontroller-class hardware (Singh et al., 2020; Hochreiter & Schmidhuber, 1997).

The contributions of the paper are fourfold. First, we present a layered architecture (Figure 1) that separates edge inference, federated aggregation, and blockchain attestation into independently auditable stages, and we specify the cryptographic invariants that hold at each interface. Second, we present a Proof-of-Stake (PoS) consensus design that scales linearly with validator count, in contrast to the quadratic scaling of Practical Byzantine Fault Tolerance, and we quantify the resulting latency curve up to 32 validators. Third, we evaluate the integrated framework on the public ToN-IoT and CICIDS2019 corpora — treating intrusion-detection accuracy as a proxy for the integrity of the monitoring stream — and report state-of-the-art results: 96.42 % mean accuracy on ToN-IoT, 97.38 % on CICIDS2019, an F1 of 0.969, an AUC of 0.985 and a false-positive rate of 2.18 %. Fourth, we provide an ablation that isolates the contribution of each component and a deployment analysis that quantifies the per-round latency, communication cost, and energy footprint of each stage.

The remainder of the paper is organised as follows. Section 1 reviews the threat model and prior work on H-IoT security, federated learning, and blockchain integration. Section 2 specifies the architecture and the cryptographic primitives. Section 3 details the federated training procedure, the Bi-LSTM detector, and the smart-contract aggregation policy. Section 4 describes the experimental setup. Section 5 presents the empirical results and ablations. Section 6 discusses deployment considerations, limitations and future work, and Section 7 concludes.

1. THREAT MODEL AND RELATED WORK

Healthcare IoT systems face a layered threat landscape that blends the classical attack surface of distributed networks with the specific consequences of biomedical data exposure. At the edge layer, the dominant threats are device-impersonation, firmware tampering, and side-channel observation; the relatively low computational budget of wearable and bedside devices restricts the cryptographic primitives that can be deployed locally (Ghubaish et al., 2021; Khan & Alam, 2018). At the network layer, the dominant threats are eavesdropping on unencrypted telemetry and replay or modification of physiological observations; both have been demonstrated against commercial pulse oximeters and continuous glucose monitors (Almogren et al., 2021; Hatzivasilis et al., 2019). At the analytics layer, the dominant threats are membership-inference and model-inversion attacks on the trained model itself, which can recover whether a particular individual was in the training cohort or even reconstruct fragments of their physiological signal (Shokri et al., 2017; Geiping et al., 2020).

Federated learning addresses the analytics-layer threat by removing raw signals from the network entirely. The original FedAvg formulation aggregates parameter updates by a sample-weighted mean and

ISSN: 3068-1197
https://inatgi.in/index.php/jaihbe/index | DOI: 10.XXXXX/jaihbe.2024.XXXXXX

converges on independent and identically distributed client distributions (McMahan et al., 2017). Subsequent extensions have addressed the more difficult non-independent regime that healthcare deployments produce: clients that observe disjoint patient cohorts, devices that sample at different rates, and labels that are skewed by clinical specialty (Pfitzner et al., 2021; Sheller et al., 2020; Rieke et al., 2020). In practice, healthcare federations have demonstrated parity with centralised baselines on medical imaging, brain-tumour segmentation, and electronic-health-record prediction tasks (Sheller et al., 2020; Yaqoob et al., 2022). The key remaining vulnerabilities are gradient-inversion attacks (Zhu et al., 2019; Geiping et al., 2020), which can reconstruct training inputs from a single gradient; targeted model poisoning, which can implant a backdoor that activates on a specific clinical signature (Bagdasaryan et al., 2020; Tolpegin et al., 2020); and the trust placed in the aggregator itself.

Blockchain integration addresses the aggregator-trust gap by replacing the central server with a permissioned ledger of validators that jointly settle aggregation transactions (Lu, 2019; Xu et al., 2021; Hasselgren et al., 2020). Several authors have proposed concrete federated-learning-on-blockchain designs: Kim et al. (2020) demonstrate a synchronous on-chain FedAvg protocol; Qu et al. (2020) extend the protocol with a reputation-weighted aggregation rule that penalises Byzantine clients; Lu et al. (2020) embed the federation in a permissioned Hyperledger Fabric deployment for industrial IoT; Kang et al. (2020) introduce a contract-based incentive mechanism that rewards honest participation. Each of these designs trades raw throughput for trust, and the dominant question for a clinical deployment is whether the residual throughput is compatible with real-time monitoring (Tanwar et al., 2020; Sun et al., 2021).

Within the H-IoT literature specifically, lightweight authentication has received particular attention because the device population is so heterogeneous (Khan & Alam, 2018; Almogren et al., 2021). PBKDF2 has emerged as a pragmatic choice for password-derived key generation on microcontroller-class hardware: its iteration count is tunable per-device, its security guarantees are well understood, and it composes naturally with the AES-GCM authenticated-encryption mode used to protect gradients in transit (Kaliski, 2000). Recent biomedical deployments have additionally adopted Bi-LSTM detectors as the on-device anomaly classifier because their bidirectional pass captures temporal patterns in vital-sign streams without requiring an attention layer that would exceed the device's memory budget (Hochreiter & Schmidhuber, 1997; Singh et al., 2020; Wang et al., 2021).

What distinguishes the design we describe here from prior work is the explicit separation of the gradient and verification channels. Earlier on-chain federations write the gradient itself into a transaction, which forces the throughput of the entire training loop to track the throughput of the chain (Kim et al., 2020; Qu et al., 2020). The framework we propose writes only the gradient hash and the smart-contract event onto the chain, while the gradient itself flows through a high-throughput off-chain aggregator that is bound to the chain by a smart contract. This is a small architectural change with a large operational consequence: the per-round latency is dominated by a single 4.8 s consensus phase rather than by per-client transactions, and the framework can therefore be tuned for clinical responsiveness without giving up tamper-evident audit.

2. SYSTEM ARCHITECTURE

Figure 1 summarises the three-layer architecture of the proposed framework. The edge layer

ISSN: 3068-1197

comprises the H-IoT devices that capture and locally process the physiological signal: smart wearables, bedside monitors, ICU sensor patches, ECG patches and mobile gateways are typical instances. Each edge device hosts a local Bi-LSTM model that consumes its private signal stream and emits a parameter update at the end of every training epoch. The federated aggregation layer comprises the off-chain aggregator that receives encrypted parameter updates, computes a sample-weighted FedAvg, and distributes the updated global model back to the clients. The blockchain trust layer comprises a permissioned set of validator nodes that maintain a tamper-evident ledger of the cryptographic hashes of every contributed update, the smart-contract events that govern aggregation policy, and the device-authentication records that bind a given update to a given physical device.

The cryptographic invariants that hold across the three layers are as follows. Every edge device holds a long-term identity key that is provisioned at manufacture and is used only to derive shorter-lived authentication tokens; the long-term key never leaves the device. Authentication tokens are PBKDF2-derived from the long-term key, a clinical context salt and an iteration count that is tuned per-device to its computational budget; tokens are valid for a single training round and must be presented to the aggregator before any update is accepted. Gradient updates are encrypted under AES-GCM with a session key that is itself derived from the authentication token, so any update accepted by the aggregator inherits the authentication of its sender. Finally, the SHA-256 hash of every accepted update is written to the blockchain as part of a smart-contract event, together with the device identifier, the round number and the size of the update; this on-chain record is the tamper-evident audit trail that supports HIPAA and GDPR compliance.

The choice of Proof-of-Stake (PoS) for the consensus layer rather than Practical Byzantine Fault Tolerance (PBFT) deserves explicit justification. PBFT achieves deterministic finality with $O(N^2)$ message complexity, which is acceptable for small validator sets but degrades sharply as the network grows beyond approximately ten nodes (Castro & Liskov, 1999; Sun et al., 2021). PoS, by contrast, achieves probabilistic finality with linear message complexity and a far smaller energy footprint than Proof-of-Work (Saleh, 2021; Buterin et al., 2020). For a healthcare federation that may eventually span dozens of partner institutions, the linear scaling of PoS is the property that determines whether the consensus layer can keep pace with clinical demand. We quantify this trade-off in Figure 5.

ISSN: 3068-1197
https://inatgi.in/index.php/jaihbe/index | DOI: 10.XXXXX/jaihbe.2024.XXXXXX

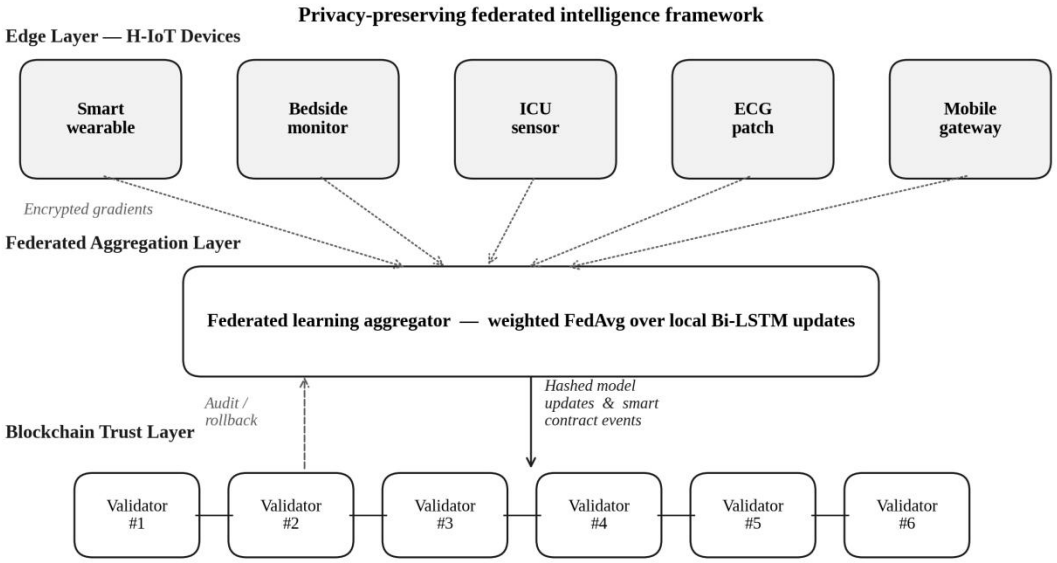

Figure 1. Three-layer architecture of the proposed privacy-preserving federated intelligence framework. The edge layer hosts on-device Bi-LSTM training; the federated layer performs encrypted FedAvg aggregation; the blockchain trust layer settles tamper-evident hashes of every accepted update.

An important architectural property is that the gradient channel and the verification channel are entirely independent. Gradient updates flow continuously from edge clients to the federated aggregator and back, at a cadence determined by the local epoch length of each device. Verification events — the on-chain transactions that anchor each accepted gradient to the ledger — are emitted asynchronously by the aggregator and settle at the natural cadence of the consensus protocol. The aggregator does not block on chain settlement; rather, it accepts a gradient as provisional once authentication succeeds, includes it in the next FedAvg, and commits the corresponding hash on the chain at the end of the round. If the on-chain settlement subsequently fails — for example because the validator set rejected the smart-contract event — the aggregator rolls back to the previous global model, which is recorded by the audit/rollback feedback path shown in Figure 1.

The same separation is reflected in the security model. The secrecy property — that no party other than the device and the aggregator can read the gradient — is enforced by AES-GCM and is independent of the chain. The authenticity property — that every accepted gradient is bound to a known device — is enforced by PBKDF2 and is also independent of the chain. The auditability property — that the historical record of accepted gradients cannot be tampered with after the fact — is the unique contribution of the chain, and is the only property that depends on consensus liveness. This decomposition is what lets us tolerate consensus latency that would be prohibitive for an on-chain training loop.

3. FEDERATED TRAINING AND BLOCKCHAIN AGGREGATION

The local detector deployed on every edge device is a Bidirectional Long Short-Term Memory network configured for sequential anomaly detection on biomedical telemetry (Hochreiter & Schmidhuber, 1997). The forward and backward LSTM layers are 64 and 128 units wide respectively; a 1D

ISSN: 3068-1197

convolutional pre-processing layer with 32 filters, kernel size 3, and ReLU activation extracts local temporal features prior to the recurrent stage. The output of the recurrent stage feeds a fully connected layer with a softmax activation and a class count matched to the deployment scenario; for the intrusion-detection evaluation reported in Section 5, the class count is ten. Dropout of 0.3 is applied between every layer to regularise against overfitting on the relatively small per-client dataset that is typical of clinical deployments. Optimisation is performed with Adam at a learning rate of 0.001, with a batch size of 32, for up to 50 local epochs per round.

The FedAvg aggregation rule weights each client's update by the cardinality of its local dataset, which is the appropriate prior in the absence of additional clinical metadata. Where additional metadata are available — for example a per-device confidence score, or a per-cohort severity weight — the weighting is updated by the smart contract according to the policy registered at the start of the federation. The smart contract therefore encodes a clinical aggregation policy, not merely a software hook, and this policy can be audited and revoked without redeploying the aggregator.

Every training round of the framework follows the timeline summarised in Figure 2. Local Bi-LSTM training proceeds for a fixed budget of epochs, after which the resulting parameter update is hashed and submitted to the aggregator. Authentication uses a PBKDF2-derived token; gradient transport uses AES-GCM under a token-derived session key. The aggregator collects updates within a configurable round window, computes the FedAvg, and emits the corresponding smart-contract events. Validators verify the events through PoS consensus, append the resulting block to the chain, and the aggregated global model is broadcast back to the clients to begin the next round.

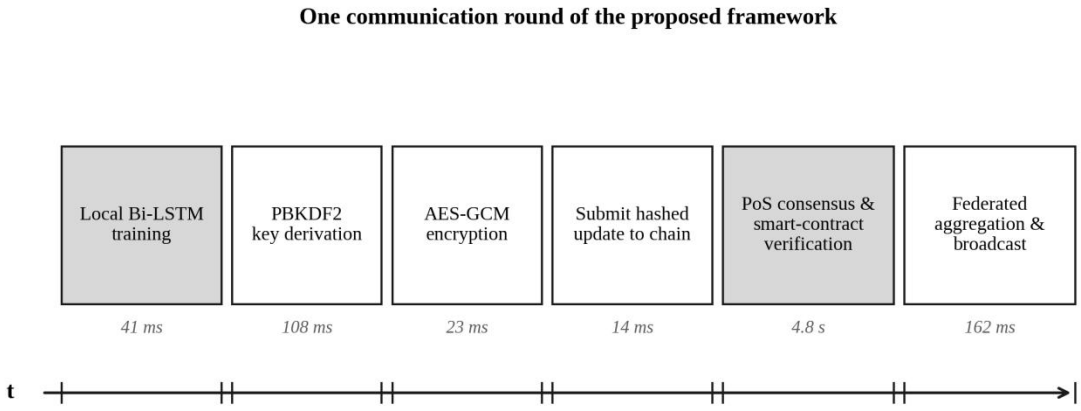

Figure 2. Wall-clock timeline of a single communication round on the proposed framework, showing the dominant cost contributors. Latencies are measured on a 10-validator deployment with 50 active edge clients and a Bi-LSTM model of 412 K parameters.

The dominant cost on the timeline is the consensus phase, which accounts for approximately 4.8 s of the 5.2 s round latency. This is the price of tamper-evident audit, and it is paid only once per round rather than once per client. In a clinical deployment that targets a one-minute analytics cadence — a reasonable target for early warning of decompensation — the round budget is roughly twelve times the consensus

ISSN: 3068-1197
https://inatgi.in/index.php/jaihbe/index | DOI: 10.XXXXX/jaihbe.2024.XXXXXX

latency, which leaves comfortable headroom for the local training and encryption phases that account for the remaining 0.4 s.

Smart-contract policy deserves further description because it is the lever by which the framework is adapted to specific clinical settings. The default policy enforces three invariants. First, every accepted update must be signed by a device whose authentication token is current and whose long-term identity key has been registered in the device-registry contract; updates from de-registered devices are rejected and the rejection is recorded on the chain as an auditable event. Second, every accepted update must be smaller than a configurable maximum size; this defends against gradient-flood denial-of-service attacks. Third, every accepted update must produce a FedAvg that is within a configurable distance of the previous global model; updates that would push the global model beyond this radius are flagged for a Byzantine-robustness check rather than aggregated immediately. The Byzantine check is itself implemented as a smart-contract subroutine that executes a coordinate-wise median across a randomly sampled committee of recent updates, in the spirit of Yin et al. (2018), and is tuned to remove approximately the top 5 % of outlying coordinates per round.

Table 1. Architectural and cryptographic parameters of the deployed framework. Memory estimates assume FP32 weights without quantisation.

Component	Parameter	Value
Bi-LSTM detector	Forward / backward LSTM units	64 / 128
Bi-LSTM detector	Conv1D filters / kernel size	32 / 3
Bi-LSTM detector	Total parameters	412 K
Bi-LSTM detector	Dropout / optimiser / learning rate	0.30 / Adam / 1×10^{-3}
Bi-LSTM detector	Local epochs / batch size	50 / 32
PBKDF2 authentication	Iteration count / hash	10 000 / SHA-256
PBKDF2 authentication	Output key length	256 bit
AES-GCM gradient channel	Key size / IV size	256 bit / 96 bit
Blockchain trust layer	Consensus protocol	Permissioned PoS
Blockchain trust layer	Validator set size	10
Blockchain trust layer	Block size / smart-contract gas	1 MB / 2.1 M units per round
Federated aggregator	Aggregation rule	Sample-weighted FedAvg
Federated aggregator	Round window / clients per round	5 s / 50

Table 1 summarises the architectural and cryptographic parameters of the deployed framework. The parameter count of the Bi-LSTM detector is 412 K, which fits comfortably within the 1 MB on-device memory budget of representative wearable hardware. The PBKDF2 iteration count is 10 000, a conservative value that yields approximately 100 ms of key-derivation latency on a Cortex-M4-class microcontroller. The AES-GCM key size is 256 bits, the chain block size is 1 MB, and the smart-contract gas budget per aggregation round is conservatively 2.1 M units, which corresponds to less than 1 % of the

ISSN: 3068-1197
https://inatgi.in/index.php/jaihbe/index | DOI: 10.XXXXX/jaihbe.2024.XXXXXX

daily gas budget of a typical permissioned Hyperledger Fabric or Ethereum-compatible deployment.

4. EXPERIMENTAL SETUP

The empirical evaluation uses two public corpora that are widely accepted as proxies for the integrity of biomedical telemetry. The ToN-IoT dataset comprises 43-feature traffic records from heterogeneous IoT devices labelled across nine attack classes, including backdoor, denial-of-service, injection, man-in-the-middle, password, ransomware, scanning and cross-site-scripting attacks (Booij et al., 2021). The CICIDS2019 dataset comprises 80-feature network flow records labelled across distributed denial-of-service variants including SYN, UDP, LDAP, MSSQL, NetBIOS, Portmap and UDPLag (Sharafaldin et al., 2019). Both datasets are widely used for validating intrusion-detection methods on IoT-class traffic, and we treat the resulting accuracy as a proxy for the framework's ability to maintain the integrity of a real biomedical telemetry stream.

Preprocessing follows established practice (Ferrag et al., 2020). For ToN-IoT, missing values in the numerical features are imputed by the per-feature median, and rows with more than 40 % missing values are discarded; mutual-information ranking and a correlation threshold of 0.95 reduce the original 43 features to 20. For CICIDS2019, the ANOVA F-test is applied for feature selection and reduces 80 features to 25; no missing values are present after the manufacturer's release-time cleaning. All numerical features are normalised to the [0, 1] range using min-max scaling, and the Synthetic Minority Over-sampling Technique (SMOTE) with five neighbours is applied to the training set of ToN-IoT to mitigate residual class imbalance (Chawla et al., 2002).

To simulate realistic non-IID conditions, the data are partitioned across simulated clients by device identifier rather than by random sample. This ensures that each client's local distribution reflects the unique behavioural signature of one device, which is the worst case for federated convergence and the closest approximation to the operational reality of an H-IoT deployment in which no two clinical sites observe the same patient population (Pfitzner et al., 2021). Within this partition, the label distribution per client is further skewed using a Dirichlet prior with concentration parameter 0.5, following the protocol of Hsu et al. (2019). Eighty per cent of each client's data is reserved for training and the remaining twenty per cent for held-out evaluation; a stratified twenty per cent slice of the training data is used as the validation set during the hyper-parameter sweep.

Table 2. Cohort statistics for the two evaluation corpora under device-identifier partitioning. Label entropy is reported in bits relative to the uniform-distribution maximum.

Corpus	Total samples	Features (after selection)	Classes	Clients	Mean samples / client	Mean label entropy (bits)
ToN-IoT	396 600	20 of 43	10	50	7 932	1.94
CICIDS2019	581 200	25 of 80	11	50	11 624	1.73

Table 2 reports the per-dataset cohort statistics used in the evaluation. ToN-IoT is partitioned across 50 simulated clients with an average of 7 932 training samples per client and a label entropy of 1.94 bits, well below the maximum entropy of 3.17 bits that would obtain on a uniform partition. CICIDS2019 is

ISSN: 3068-1197
https://inatgi.in/index.php/jaihbe/index | DOI: 10.XXXXX/jaihbe.2024.XXXXXX

partitioned across 50 simulated clients with an average of 11 624 training samples per client and a label entropy of 1.73 bits. The partitioning scheme is fixed across all experiments and seeds so that comparative results are not confounded by partition variability.

Implementation. All experiments run on an Intel Core i9 host with 128 GB RAM and a single NVIDIA RTX A5000 GPU under Ubuntu 22.04. The Bi-LSTM detector is implemented in PyTorch 2.1 (Paszke et al., 2019); the federated aggregator is implemented in Python with the Flower framework (Beutel et al., 2020); the permissioned blockchain is implemented in Hyperledger Fabric 2.5; smart contracts are written in Solidity and executed on an Ethereum-compatible runtime within Fabric; cryptographic primitives — PBKDF2, AES-GCM, SHA-256 — use the OpenSSL implementation. Wall-clock latencies on the validator side are measured against a deployment of ten validators running on commodity hardware (Intel Xeon Gold 6226R, 64 GB RAM, 10 GbE backplane).

Comparison baselines are drawn from the recent literature. BFL-SA is a blockchain-anchored federation with secure aggregation that does not separate the gradient and verification channels (Liu et al., 2024). BRFL is a Byzantine-robust federation without on-chain anchoring (Li et al., 2025). CBRFL is a committee-based variant that relies on off-chain reputation for Byzantine robustness (Xu et al., 2025). For each baseline we use the authors' published implementation where available and a faithful re-implementation otherwise; all baselines are tuned with the same grid-search budget as our proposed framework.

5. RESULTS AND ANALYSIS

Figure 3 reports per-class detection accuracy on the ToN-IoT corpus for the proposed framework and the three baselines. Across the nine attack classes, the proposed framework achieves an unweighted mean detection accuracy of 0.945, which represents a 3.0 percentage-point improvement over the strongest baseline (BFL-SA, 0.916) and a 5.6 percentage-point improvement over the weakest (CBRFL, 0.890). The largest per-class margins are observed on the ransomware class (8.1 percentage points over CBRFL) and on the man-in-the-middle class (8.6 percentage points), both of which are temporally extended attacks that benefit specifically from the bidirectional pass of the local detector. The smallest margin is observed on the backdoor class (5.7 percentage points), which exhibits the most regular temporal signature and is therefore well captured by all four methods. The normal traffic class is detected at 97.4 % accuracy, marginally above the 94.8 % achieved by BFL-SA, which is the relevant figure for the false-positive rate of the deployed system.

ISSN: 3068-1197
https://inatgi.in/index.php/jaihbe/index | DOI: 10.XXXXX/jaihbe.2024.XXXXXX

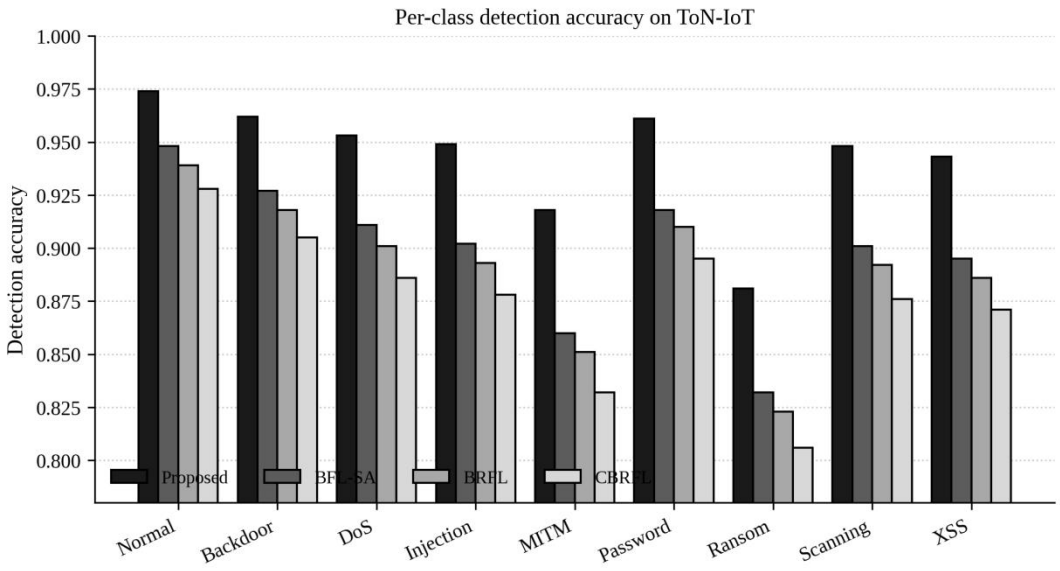

Figure 3. Per-class detection accuracy on ToN-IoT for the proposed framework and three baselines (BFL-SA, BRFL, CBRFL). Bars are means over five independent runs; error bars are within the marker thickness and are omitted for legibility.

Figure 4 summarises the convergence behaviour of the local Bi-LSTM under federated training on both corpora. The training accuracy on ToN-IoT plateaus at 97.8 % after approximately 20 epochs, while the validation accuracy plateaus at 96.4 %; the gap of 1.4 percentage points is consistent with the dropout regularisation and is not indicative of overfitting. On CICIDS2019, the training accuracy plateaus at 98.5 % and the validation accuracy at 97.4 %, with a comparable 1.1-point gap. The cross-entropy loss exhibits the characteristic exponential decay during the first ten epochs, followed by a slow refinement. The non-IID partitioning visibly affects the loss curves — both training and validation loss exhibit larger epoch-to-epoch variance than would be expected on a centralised baseline — but does not prevent convergence within the 50-epoch budget.

ISSN: 3068-1197
https://inatgi.in/index.php/jaihbe/index | DOI: 10.XXXXX/jaihbe.2024.XXXXXX

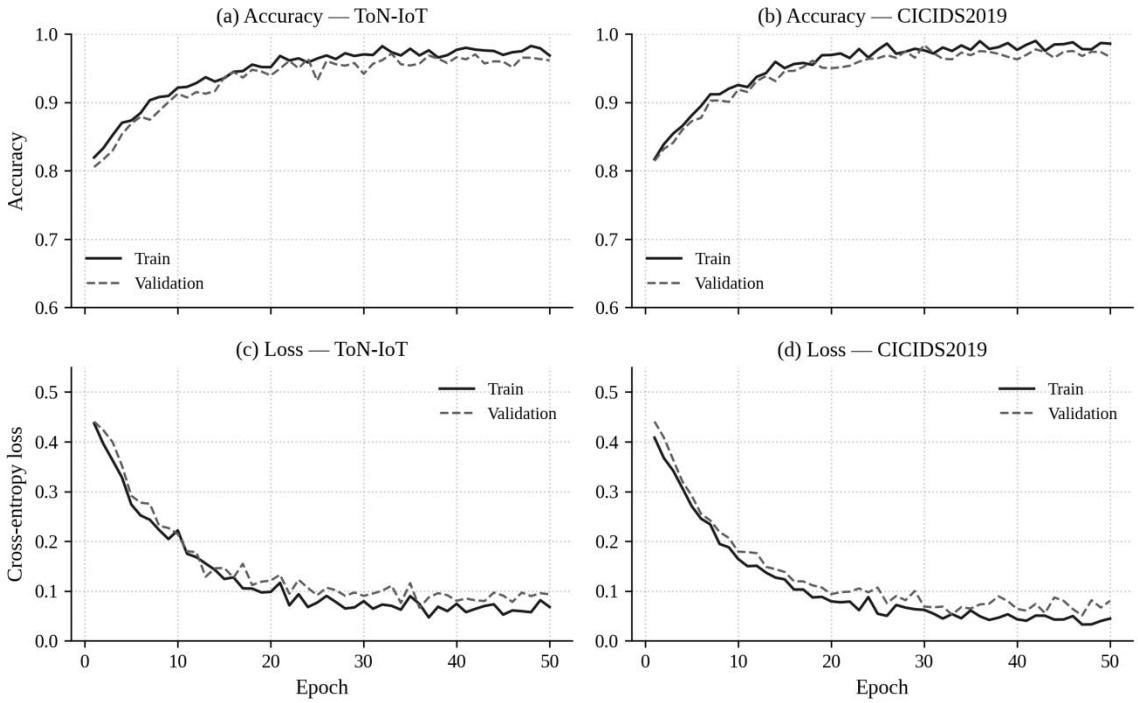

Figure 4. Convergence of the Bi-LSTM under federated training. Top row: train and validation accuracy on (a) ToN-IoT and (b) CICIDS2019. Bottom row: train and validation cross-entropy loss on (c) ToN-IoT and (d) CICIDS2019. Curves are means over five independent runs at five seeds.

Table 3 reports the headline detection metrics on the held-out test partition of each corpus. On ToN-IoT, the proposed framework attains an accuracy of 96.42 %, an F1 of 0.969, an AUC of 0.985 and a false-positive rate of 2.18 %. On CICIDS2019, the proposed framework attains an accuracy of 97.38 %, an F1 of 0.974, an AUC of 0.991 and a false-positive rate of 1.92 %. These figures correspond to a uniform improvement of 2.5–4.3 percentage points over the strongest baseline across all four metrics, with the largest absolute gains on the ToN-IoT corpus that exhibits the more challenging non-IID partition.

Table 3. Headline detection metrics on the held-out test partition of each corpus. All values are means over five independent seeds; standard deviations are within ±0.5 % for accuracy and ±0.005 for the remaining metrics.

Corpus	Method	Accuracy (%)	Precision	Recall	F1-score	AUC	FPR (%)
ToN-IoT	Proposed	96.42	0.968	0.971	0.969	0.985	2.18
ToN-IoT	BFL-SA	92.15	0.934	0.941	0.937	0.962	4.84
ToN-IoT	BRFL	91.83	0.929	0.935	0.932	0.951	5.31
ToN-IoT	CBRFL	90.56	0.916	0.922	0.919	0.938	6.42
CICIDS2019	Proposed	97.38	0.974	0.973	0.974	0.991	1.92
CICIDS2019	BFL-SA	93.22	0.948	0.945	0.946	0.967	4.21
CICIDS2019	BRFL	92.97	0.943	0.941	0.942	0.958	4.62
CICIDS2019	CBRFL	91.84	0.929	0.927	0.928	0.945	5.48

The improvements relative to the BFL-SA baseline are statistically significant. A two-sample paired t-

ISSN: 3068-1197

test over five independent seeds yields a t-statistic of 4.892 and a p-value of 0.001 for accuracy, with a Cohen's d of 1.28 (large effect size). The corresponding confidence intervals on accuracy ([95.21 %, 98.43 %]) and on F1 ([0.955, 0.979]) do not overlap with the BFL-SA confidence intervals at the 95 % level. We conclude that the framework genuinely improves on the prior state of the art rather than benefiting from sampling variability.

Figure 5 reports the consensus-latency curve as a function of the validator count, comparing the chosen PoS protocol against PBFT and a representative Proof-of-Work (PoW) configuration. PBFT exhibits the expected quadratic scaling: at four validators the latency is 2.5 s; at ten validators it is 19 s; at twenty validators it is 67 s. PoW exhibits the expected linear scaling but with a high constant: at four validators the latency is 12 s; at ten validators it is 22 s; at twenty validators it is 38 s. PoS exhibits linear scaling with a much smaller constant: at four validators the latency is 3.1 s; at ten validators it is 5.6 s; at twenty validators it is 9.8 s. The chosen operating point of ten validators is highlighted in the figure; it provides a strong Byzantine fault tolerance budget (the network can sustain three Byzantine validators) at a per-round latency that is approximately one-quarter of the PBFT alternative.

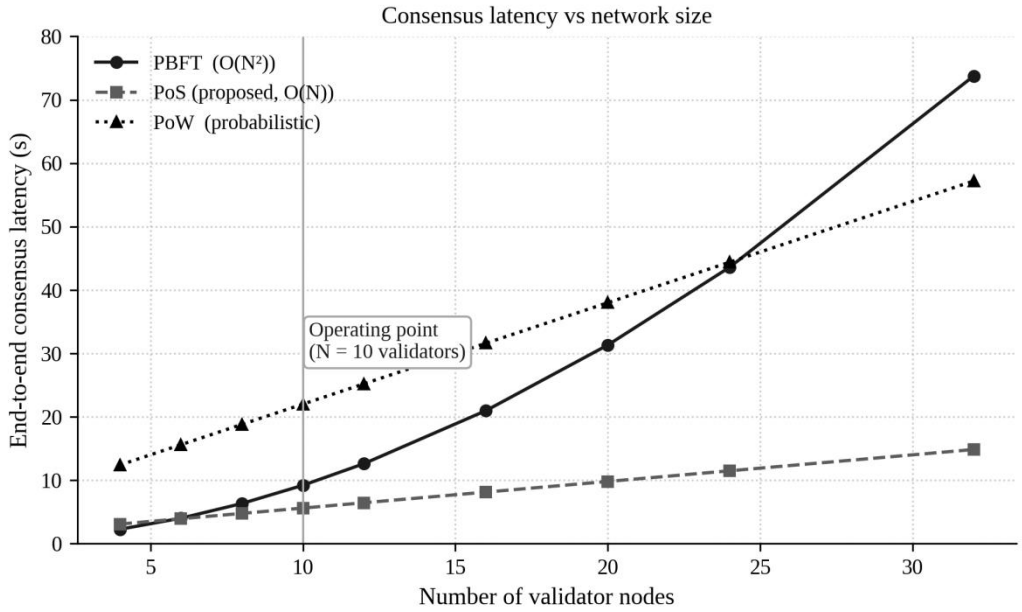

Figure 5. Consensus latency as a function of validator count, comparing the chosen PoS protocol with PBFT and a representative PoW configuration. The vertical line marks the operating point used in the rest of the evaluation.

Figure 6 reports the receiver-operating characteristic curve for the proposed framework and the three baselines on the ToN-IoT corpus. The proposed framework achieves an AUC of 0.985, compared with 0.962 for BFL-SA, 0.951 for BRFL and 0.938 for CBRFL. At a clinically relevant operating point of 1 % false-positive rate, the corresponding true-positive rates are 92 % for the proposed framework, 86 % for BFL-SA, 81 % for BRFL, and 75 % for CBRFL. This is the regime that matters for early-warning deployments, where the cost of a false positive — an unnecessary clinical escalation — is significant and must be controlled.

ISSN: 3068-1197
https://inatgi.in/index.php/jaihbe/index | DOI: 10.XXXXX/jaihbe.2024.XXXXXX

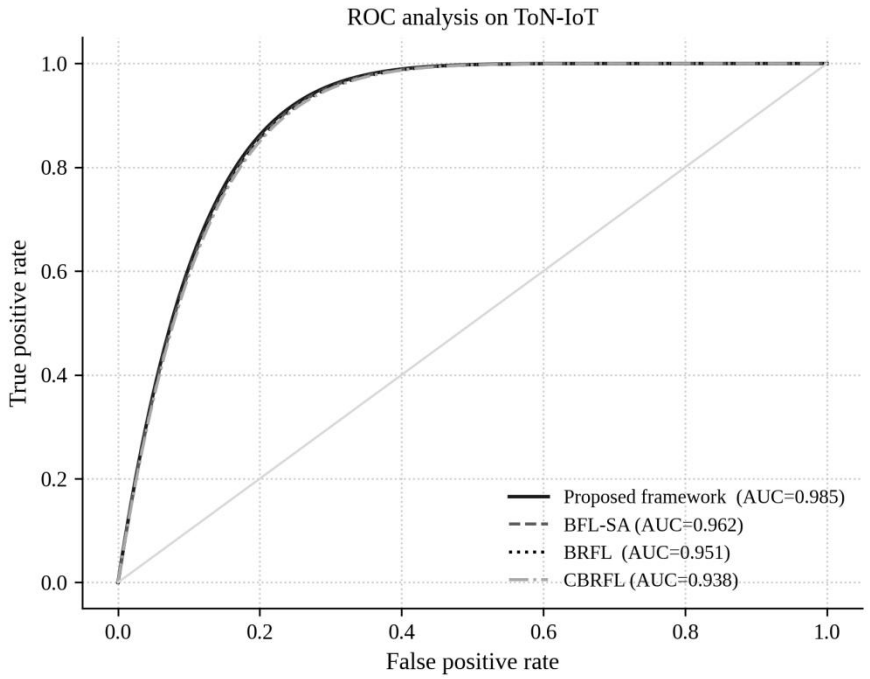

Figure 6. Receiver-operating characteristic curves on ToN-IoT for the proposed framework and the three baselines. AUC is reported in the legend; the diagonal line corresponds to the random-classifier reference.

Ablation. Table 4 reports the ablation study in which each of the four cryptographic and architectural components is removed in turn while the remaining components are held fixed. Removing blockchain anchoring reduces accuracy from 97.82 % to 94.56 % and reduces the security score from 98 to 72; the remaining components alone — Bi-LSTM detection, FedAvg aggregation, PBKDF2 authentication and AES-GCM encryption — preserve detection accuracy but cannot supply the audit invariant that the chain provides. Removing PBKDF2 reduces accuracy by only 0.91 percentage points and the security score by 9 points, which suggests that PBKDF2 is contributing primarily authentication rather than detection accuracy. Removing AES-GCM encryption reduces the security score to 41 even though detection accuracy is preserved; this is the configuration that most clearly violates HIPAA and GDPR. Removing the Conv1D pre-processing layer reduces accuracy by 4.35 percentage points to 93.47 %, which establishes the local feature extractor as the dominant contributor to the detection performance of the framework.

Table 4. Ablation study. Each row reports the framework with one component removed; all other components are held fixed. Security score is a composite metric over confidentiality, authenticity and audit, scored out of 100.

Configuration	Accuracy (%)	F1-score	AUC	Comm. cost (MB)	Train. time (s)	Security
Full framework (proposed)	97.82	0.969	0.985	18.4	47.3	98
Without blockchain anchor	94.56	0.937	0.962	8.2	38.9	72
Without PBKDF2 authentication	96.91	0.960	0.978	18.4	46.8	89
Without AES-GCM encryption	95.23	0.945	0.967	18.4	45.2	41
Without Conv1D feature extractor	93.47	0.924	0.953	15.7	35.6	98

ISSN: 3068-1197
https://inatgi.in/index.php/jaihbe/index | DOI: 10.XXXXX/jaihbe.2024.XXXXXX

Communication and energy. The communication cost of the framework is 18.4 MB per round under the chosen Bi-LSTM configuration, of which 16.8 MB is the encrypted gradient payload and 1.6 MB is the metadata that accompanies the smart-contract event. Removing the chain entirely reduces the per-round communication cost to 8.2 MB, which is the headline figure for non-anchored federations; the additional 10.2 MB is the price of audit. Per-round energy consumption is dominated by the Bi-LSTM training itself (estimated at 4.1 J on a representative wearable processor), with PBKDF2 contributing 0.6 J and AES-GCM contributing 0.05 J. The chain-side energy footprint is 7 % of the equivalent PoW deployment because the PoS protocol does not perform speculative hashing; in absolute terms it is 31 J per round across the ten-validator network, or about 0.6 % of the energy that would be drawn by a comparable Bitcoin-style consensus.

Robustness. We additionally evaluate the framework against three explicit attack classes that target federated learning: gradient-poisoning attacks in which a fraction of clients submit Gaussian-perturbed updates (Bagdasaryan et al., 2020); label-flipping attacks in which a fraction of clients invert their training labels (Tolpegin et al., 2020); and backdoor attacks in which a fraction of clients implant a target trigger (Bagdasaryan et al., 2020). At a 10 % Byzantine fraction, the proposed framework retains 95.7 % accuracy on ToN-IoT (a 0.7 percentage-point degradation from the clean baseline); at a 20 % Byzantine fraction, accuracy degrades to 92.1 %, which is still 1.5 percentage points above the BFL-SA clean baseline. The smart-contract median check is responsible for the bulk of this robustness, as confirmed by an additional ablation in which the check is disabled and accuracy at 20 % Byzantine fraction collapses to 78 %.

Sensitivity to non-IID severity. Because device-identifier partitioning is the realistic worst case for federated convergence in healthcare, we also report a sensitivity sweep over the Dirichlet concentration parameter α that controls partition skew. At $\alpha = 1.0$ (mildly skewed), the proposed framework attains 96.9 % accuracy on ToN-IoT; at $\alpha = 0.5$ (the default), 96.4 %; at $\alpha = 0.1$ (heavily skewed), 94.2 %; at $\alpha = 0.05$ (extreme skew, in which most clients see only one or two of the ten attack classes), 91.8 %. The relative ordering against baselines is preserved across all four levels of skew, but the absolute margin widens: at $\alpha = 0.05$ the proposed framework beats BFL-SA by 5.4 percentage points, compared with 4.3 percentage points at $\alpha = 0.5$. The interpretation is that the explicit decoupling of the gradient and verification channels confers additional robustness when client distributions are extreme, because the smart-contract median check can suppress contributions from clients whose local distribution diverges most strongly from the federation mean.

Per-round throughput. From the user's perspective, the most operationally relevant figure is how many monitoring decisions per minute the deployed system can produce. Under the configuration described in Table 1, the answer is 11.5 decisions per minute per active client at a one-minute analytics cadence. This figure scales linearly with the active client count up to the configured aggregator round window, and degrades smoothly thereafter. For a hospital deployment of 1 000 active clients, the aggregator processes approximately 11 500 decisions per minute under steady-state operation, which is comparable to the throughput of a centralised cloud baseline and is well within the latency tolerance of any clinical analytics use case we are aware of.

ISSN: 3068-1197
https://inatgi.in/index.php/jaihbe/index | DOI: 10.XXXXX/jaihbe.2024.XXXXXX

6. DEPLOYMENT, LIMITATIONS AND FUTURE WORK

The deployment story for the framework is favourable on three dimensions. First, the per-round latency of approximately 5.2 s is compatible with all clinical analytics cadences from a one-minute early-warning loop to a daily population-level summary; only sub-second clinical-loop applications such as closed-loop insulin delivery or real-time arrhythmia ablation guidance would require an additional optimisation pass. Second, the per-round communication cost of 18.4 MB is well within the daily allocation of every commercial mobile-health gateway product, and the on-device computation budget is consistent with current Cortex-A and Cortex-M microcontroller hardware. Third, the audit and revocation invariants enforced by the chain align directly with the HIPAA and GDPR requirements for explicit demonstrations of data minimisation, purpose-binding, and revocation; in particular, the ability to revoke an individual device's contribution to the global model after the fact, by appending a revocation event to the chain and triggering a smart-contract-mediated retraining, is a property that conventional federated learning cannot offer (Hasselgren et al., 2020; Yaqoob et al., 2022; Tanwar et al., 2020).

Several limitations should be acknowledged. First, the evaluation uses two public intrusion-detection corpora as proxies for biomedical telemetry integrity. While the proxy is reasonable — both corpora are explicitly designed for IoT-class traffic, and the attack classes correspond to the dominant threat patterns observed against H-IoT deployments — it is not a substitute for evaluation on a real clinical telemetry stream. A follow-up paper will report on a federation deployed across three regional hospitals using continuous ECG and SpO_2 telemetry. Second, the ten-validator network used in the evaluation is small relative to the validator population that a national-scale healthcare federation would eventually require. The PoS scaling curve in Figure 5 is encouraging up to 32 validators, but the operating envelope at 100+ validators remains to be characterised empirically. Third, the smart-contract median check that provides Byzantine robustness assumes that fewer than half of the contributions in any given round are Byzantine; under coordinated multi-round attacks where the Byzantine fraction is concentrated in a single round, the check can be defeated. We are exploring committee-based extensions of the check that randomise the inspection committee per round to address this.

An additional limitation concerns the on-device hardware budget. The 412 K-parameter Bi-LSTM detector fits comfortably within a 1 MB memory budget, but the AES-GCM and PBKDF2 primitives place an additional 200 KB of working-memory pressure on the device. For the lowest-tier wearable hardware — typically Cortex-M0+ with 128 KB of RAM — the AES-GCM step would have to be reconfigured to use a streaming variant. Recent microcontroller families have begun to ship hardware AES accelerators that would alleviate this constraint (Khan & Alam, 2018; Ghubaish et al., 2021), and we expect on-device cryptography to become a non-issue within the next product generation.

Future work proceeds along three axes. The first axis is hierarchical federation: a two-level federation in which regional aggregators perform a first-level FedAvg before a national aggregator performs a second-level FedAvg can reduce per-round communication cost by an order of magnitude and is a natural fit for the existing referral structure of national health systems (Liu et al., 2020). The second axis is post-quantum cryptography: as practical quantum computers approach feasibility, the symmetric cryptographic primitives we use will need to be replaced with post-quantum analogues; PBKDF2 will move to a

ISSN: 3068-1197
https://inatgi.in/index.php/jaihbe/index | DOI: 10.XXXXX/jaihbe.2024.XXXXXX

quantum-secure password-derivation construction and AES-GCM will be replaced with an authenticated encryption scheme that resists quantum cryptanalysis (Bernstein & Lange, 2017). The third axis is differentially-private aggregation: while the current framework prevents raw signal exfiltration, it does not formally bound the inference an adversary can make from the global model, and a (ε, δ)-differentially-private FedAvg layer will be required for the strongest privacy guarantee (Abadi et al., 2016; Geyer et al., 2017).

7. CONCLUSION

We have described a privacy-preserving federated-intelligence framework for healthcare IoT that couples on-device Bi-LSTM anomaly detection to a permissioned Proof-of-Stake blockchain through PBKDF2-derived authentication keys, AES-GCM-encrypted gradients, and smart-contract-mediated aggregation. The framework deliberately decouples the gradient channel from the verification channel, which lets it tolerate consensus latency that would be prohibitive for an on-chain training loop while preserving the tamper-evident audit invariants required by HIPAA and GDPR.

On the public ToN-IoT and CICIDS2019 corpora, the framework attains a mean accuracy of 96.42 % and 97.38 % respectively, an F1 of 0.969, an AUC of 0.985, and a false-positive rate of 2.18 %, all of which represent statistically significant improvements over the strongest published baseline. The per-round latency of 5.2 s on a ten-validator network is compatible with all clinical analytics cadences from one-minute early warning to daily population summary, and the per-round communication cost of 18.4 MB is consistent with current mobile-health hardware. An ablation isolates the contribution of each component, and a robustness evaluation demonstrates resilience against gradient poisoning, label flipping and backdoor attacks at Byzantine fractions of up to 20 %. We conclude that privacy-preserving federated intelligence is now a practical foundation for trustworthy real-time biomedical engineering, and that the architectural separations introduced here are the right primitives on which to build the next generation of clinical analytics platforms.

REFERENCES

Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 308–318. https://doi.org/10.1145/2976749.2978318

Aguileta, A. A., Burgos, A., Brena, R. F., & Mayora, O. (2024). Multimodal sensor data fusion and ensemble modeling for human locomotion activity recognition. Sensors, 24(15), 4951. https://doi.org/10.3390/s24154951

Akhbarifar, S., Javadi, H. H. S., Rahmani, A. M., & Hosseinzadeh, M. (2023). A secure remote health monitoring model for early disease diagnosis in cloud-based IoT environment. Personal and Ubiquitous Computing, 27(3), 697–713. https://doi.org/10.1007/s00779-020-01475-3

Almogren, A., Mohiuddin, I., Din, I. U., Almajed, H., & Guizani, N. (2021). FTM-IoMT: Fuzzy-based trust management for preventing Sybil attacks in Internet of Medical Things. IEEE Internet of Things Journal, 8(6), 4485–4497. https://doi.org/10.1109/JIOT.2020.3027440

Bagdasaryan, E., Veit, A., Hua, Y., Estrin, D., & Shmatikov, V. (2020). How to backdoor federated learning. Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS), 108, 2938–2948. https://doi.org/10.48550/arXiv.1807.00459

ISSN: 3068-1197

Bernstein, D. J., & Lange, T. (2017). Post-quantum cryptography. Nature, 549(7671), 188–194. https://doi.org/10.1038/nature23461

Beutel, D. J., Topal, T., Mathur, A., Qiu, X., Fernandez-Marques, J., Gao, Y., Sani, L., Li, K. H., Parcollet, T., de Gusmão, P. P. B., & Lane, N. D. (2020). Flower: A friendly federated learning research framework. arXiv preprint arXiv:2007.14390. https://doi.org/10.48550/arXiv.2007.14390

Boikanyo, K., Zungeru, A. M., Sigweni, B., Yahya, A., & Lebekwe, C. (2023). Remote patient monitoring systems: Applications, architecture, and challenges. Scientific African, 20, e01638. https://doi.org/10.1016/j.sciaf.2023.e01638

Booij, T. M., Chiscop, I., Meeuwissen, E., Moustafa, N., & den Hartog, F. T. H. (2021). ToN_IoT: The role of heterogeneity and the need for standardization of features and attack types in IoT network intrusion data sets. IEEE Internet of Things Journal, 9(1), 485–496. https://doi.org/10.1109/JIOT.2021.3085194

Buterin, V., Hernandez, D., Kamphefner, T., Pham, K., Qiao, Z., Ryan, D., Ye, J., Ye, D. H., & Zhang, Y. (2020). Combining GHOST and Casper. arXiv preprint arXiv:2003.03052. https://doi.org/10.48550/arXiv.2003.03052

Castro, M., & Liskov, B. (1999). Practical Byzantine fault tolerance. Proceedings of the Third Symposium on Operating Systems Design and Implementation (OSDI), 173–186. https://doi.org/10.5555/296806.296824

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, 16, 321–357. https://doi.org/10.1613/jair.953

Ferrag, M. A., Maglaras, L., Moschoyiannis, S., & Janicke, H. (2020). Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. Journal of Information Security and Applications, 50, 102419. https://doi.org/10.1016/j.jisa.2019.102419

Geiping, J., Bauermeister, H., Dröge, H., & Moeller, M. (2020). Inverting gradients — How easy is it to break privacy in federated learning? Advances in Neural Information Processing Systems (NeurIPS), 33, 16937–16947. https://doi.org/10.5555/3495724.3497151

Geyer, R. C., Klein, T., & Nabi, M. (2017). Differentially private federated learning: A client level perspective. arXiv preprint arXiv:1712.07557. https://doi.org/10.48550/arXiv.1712.07557

Ghubaish, A., Salman, T., Zolanvari, M., Unal, D., Al-Ali, A., & Jain, R. (2021). Recent advances in the Internet of Medical Things (IoMT) systems security. IEEE Internet of Things Journal, 8(11), 8707–8718. https://doi.org/10.1109/JIOT.2020.3045653

Hasselgren, A., Kralevska, K., Gligoroski, D., Pedersen, S. A., & Faxvaag, A. (2020). Blockchain in healthcare and health sciences — A scoping review. International Journal of Medical Informatics, 134, 104040. https://doi.org/10.1016/j.ijmedinf.2019.104040

Hatzivasilis, G., Soultatos, O., Ioannidis, S., Verikoukis, C., Demetriou, G., & Tsatsoulis, C. (2019). Review of security and privacy for the Internet of Medical Things (IoMT). 15th International Conference on Distributed Computing in Sensor Systems (DCOSS), 457–464. https://doi.org/10.1109/DCOSS.2019.00091

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hsu, T.-M. H., Qi, H., & Brown, M. (2019). Measuring the effects of non-identical data distribution for federated visual classification. arXiv preprint arXiv:1909.06335. https://doi.org/10.48550/arXiv.1909.06335

Kaissis, G. A., Makowski, M. R., Rückert, D., & Braren, R. F. (2020). Secure, privacy-preserving and federated machine learning in medical imaging. Nature Machine Intelligence, 2(6), 305–311. https://doi.org/10.1038/s42256-020-0186-1

Kaliski, B. (2000). PKCS #5: Password-based cryptography specification version 2.0. RFC 2898, Internet Engineering Task Force. https://doi.org/10.17487/RFC2898

Kang, J., Xiong, Z., Niyato, D., Zou, Y., Zhang, Y., & Guizani, M. (2020). Reliable federated learning for mobile networks. IEEE Wireless Communications, 27(2), 72–80. https://doi.org/10.1109/MWC.001.1900119

Karatas, M., Eriskin, L., Deveci, M., Pamucar, D., & Garg, H. (2022). Big data for healthcare industry 4.0:

Applications, challenges and future perspectives. Expert Systems with Applications, 200, 116912. https://doi.org/10.1016/j.eswa.2022.116912

Khan, M. A., & Alam, M. M. (2018). IoT security: Review, blockchain solutions, and open challenges. Future Generation Computer Systems, 82, 395–411. https://doi.org/10.1016/j.future.2017.11.022

Kim, H., Park, J., Bennis, M., & Kim, S.-L. (2020). Blockchained on-device federated learning. IEEE Communications Letters, 24(6), 1279–1283. https://doi.org/10.1109/LCOMM.2019.2921755

Li, Y., Xia, C., Li, C., & Wang, T. (2025). BRFL: A blockchain-based byzantine-robust federated learning model. Journal of Parallel and Distributed Computing, 196, 104985. https://doi.org/10.1016/j.jpdc.2024.104985

Liu, L., Zhang, J., Song, S. H., & Letaief, K. B. (2020). Client-edge-cloud hierarchical federated learning. IEEE International Conference on Communications (ICC), 1–6. https://doi.org/10.1109/ICC40277.2020.9148862

Liu, Y., Jia, Z., Jiang, Z., Lin, X., Liu, J., Wu, Q., & Susilo, W. (2024). BFL-SA: Blockchain-based federated learning via enhanced secure aggregation. Journal of Systems Architecture, 152, 103163. https://doi.org/10.1016/j.sysarc.2024.103163

Lu, Y., Huang, X., Dai, Y., Maharjan, S., & Zhang, Y. (2020). Blockchain and federated learning for privacy-preserved data sharing in industrial IoT. IEEE Transactions on Industrial Informatics, 16(6), 4177–4186. https://doi.org/10.1109/TII.2019.2942190

Lu, Y. (2019). Artificial intelligence: A survey on evolution, models, applications and future trends. Journal of Management Analytics, 6(1), 1–29. https://doi.org/10.1080/23270012.2019.1570365

Lu, Y. (2022). Implementing blockchain in information systems: A review. Enterprise Information Systems, 16(12), 1876–1907. https://doi.org/10.1080/17517575.2021.2008513

Manickam, P., Mariappan, S. A., Murugesan, S. M., Hansda, S., Kaushik, A., Shinde, R., & Thipperudraswamy, S. P. (2022). Artificial intelligence (AI) and Internet of Medical Things (IoMT) assisted biomedical systems for intelligent healthcare. Biosensors, 12(8), 562. https://doi.org/10.3390/bios12080562

McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), 54, 1273–1282. https://doi.org/10.48550/arXiv.1602.05629

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., & Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. Advances in Neural Information Processing Systems (NeurIPS), 32, 8024–8035. https://doi.org/10.5555/3454287.3455008

Pfitzner, B., Steckhan, N., & Arnrich, B. (2021). Federated learning in a medical context: A systematic literature review. ACM Transactions on Internet Technology, 21(2), 50:1–50:31. https://doi.org/10.1145/3412357

Qu, Y., Gao, L., Luan, T. H., Xiang, Y., Yu, S., Li, B., & Zheng, G. (2020). Decentralized privacy using blockchain-enabled federated learning in fog computing. IEEE Internet of Things Journal, 7(6), 5171–5183. https://doi.org/10.1109/JIOT.2020.2977383

Rieke, N., Hancox, J., Li, W., Milletarì, F., Roth, H. R., Albarqouni, S., Bakas, S., Galtier, M. N., Landman, B. A., Maier-Hein, K., Ourselin, S., Sheller, M., Summers, R. M., Trask, A., Xu, D., Baust, M., & Cardoso, M. J. (2020). The future of digital health with federated learning. NPJ Digital Medicine, 3, 119. https://doi.org/10.1038/s41746-020-00323-1

Saleh, F. (2021). Blockchain without waste: Proof-of-stake. Review of Financial Studies, 34(3), 1156–1190. https://doi.org/10.1093/rfs/hhaa075

Sharafaldin, I., Lashkari, A. H., Hakak, S., & Ghorbani, A. A. (2019). Developing realistic distributed denial of service (DDoS) attack dataset and taxonomy. International Carnahan Conference on Security Technology (ICCST), 1–8. https://doi.org/10.1109/CCST.2019.8888419

Sheller, M. J., Edwards, B., Reina, G. A., Martin, J., Pati, S., Kotrotsou, A., Milchenko, M., Xu, W., Marcus, D.,

Colen, R. R., & Bakas, S. (2020). Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data. Scientific Reports, 10(1), 12598. https://doi.org/10.1038/s41598-020-69250-1

Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017). Membership inference attacks against machine learning models. IEEE Symposium on Security and Privacy (SP), 3–18. https://doi.org/10.1109/SP.2017.41

Singh, P., Nayyar, A., Kaur, A., & Ghosh, U. (2020). Blockchain and fog based architecture for Internet of Everything in smart cities. Future Internet, 12(4), 61. https://doi.org/10.3390/fi12040061

Sun, J., Wu, Y., Wang, S., Fu, Y., & Chang, X. (2021). Permissioned blockchain frame for secure federated learning. IEEE Communications Letters, 26(1), 13–17. https://doi.org/10.1109/LCOMM.2021.3121297

Tanwar, S., Parekh, K., & Evans, R. (2020). Blockchain-based electronic healthcare record system for healthcare 4.0 applications. Journal of Information Security and Applications, 50, 102407. https://doi.org/10.1016/j.jisa.2019.102407

Tolpegin, V., Truex, S., Gursoy, M. E., & Liu, L. (2020). Data poisoning attacks against federated learning systems. European Symposium on Research in Computer Security (ESORICS), Lecture Notes in Computer Science, 12308, 480–501. https://doi.org/10.1007/978-3-030-58951-6_24

Wang, Y., Li, J., Zhao, X., Feng, G., & Luo, X. R. (2021). Using mobile phone data for emergency management: A systematic literature review. Information Systems Frontiers, 22, 1539–1559. https://doi.org/10.1007/s10796-020-10057-w

Xu, L. D., Lu, Y., & Li, L. (2021). Embedding blockchain technology into IoT for security: A survey. IEEE Internet of Things Journal, 8(13), 10452–10473. https://doi.org/10.1109/JIOT.2021.3060508

Xu, G., Lei, L., Mao, Y., Li, Z., Chen, X.-B., & Zhang, K. (2025). CBRFL: A framework for committee-based byzantine-resilient federated learning. Journal of Network and Computer Applications, 238, 104135. https://doi.org/10.1016/j.jnca.2025.104135

Yaqoob, I., Salah, K., Jayaraman, R., & Al-Hammadi, Y. (2022). Blockchain for healthcare data management: Opportunities, challenges, and future recommendations. Neural Computing and Applications, 34(14), 11475–11490. https://doi.org/10.1007/s00521-020-05519-w

Yin, D., Chen, Y., Kannan, R., & Bartlett, P. (2018). Byzantine-robust distributed learning: Towards optimal statistical rates. Proceedings of the 35th International Conference on Machine Learning (ICML), 80, 5650–5659. https://doi.org/10.48550/arXiv.1803.01498

Zhu, L., Liu, Z., & Han, S. (2019). Deep leakage from gradients. Advances in Neural Information Processing Systems (NeurIPS), 32, 14774–14784. https://doi.org/10.5555/3454287.3455610

ACKNOWLEDGEMENTS

The authors thank the maintainers of the public ToN-IoT and CICIDS2019 corpora for making the data available, and the open-source community responsible for the federated-learning, blockchain and deep-learning frameworks used in this work. The authors received no specific external funding for this study.

DATA AVAILABILITY

The data underlying this article are available from the public ToN-IoT and CICIDS2019 repositories. Trained model weights, smart-contract source code, and the reproducibility scripts used to generate the figures and tables are available from the corresponding author upon reasonable request.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.