Federated Deep Learning for Telecom-Orchestrated Anomaly Detection in Industrial IoT Critical Infrastructure Networks

Miloš Milovančević*

Faculty of Mechanical Engineering, University of Niš, Serbia

Abstract: Industrial Internet of Things (IIoT) critical infrastructure is increasingly connected through private 5G/6G networks, software-defined transport, network slicing and O-RAN-compatible orchestration stacks. In such environments, anomaly detection is not only an application-layer cybersecurity task; it is also a communication-engineering signal that can trigger changes in RAN, core and transport control loops. Centralizing sensor, radio and network telemetry for high-accuracy deep anomaly detection violates data-sovereignty and regulatory requirements, while isolated local models fail to generalize across heterogeneous plants and private-network deployments. Federated deep learning (FDL) offers a privacy-preserving architecture for collaborative anomaly detection, but its practical value depends on how anomaly scores are mapped to telecom control elements such as network slices, QoS flows, RIC xApps/rApps, SDN/NFV routing and isolation functions. This paper reviews FDL architectures for IIoT anomaly detection with explicit grounding in telecom system design. Beyond F1-score comparison, the analysis emphasizes radio and network KPIs including SINR degradation, PRB utilization, handover failure rate, link latency, throughput, QoS/QoE, control overhead and spectrum constraints. The paper further explains how anomaly outputs can support automated control actions such as rerouting, slice-profile adjustment, QoS enforcement, edge isolation and SDN/NFV remediation, while respecting radio latency budgets, closed-loop stability and safety requirements. Security threats, communication overhead, model compression and regulatory constraints are analyzed as deployment factors for private 5G/6G critical-infrastructure networks.

Keywords: Federated learning, Anomaly detection, Industrial IoT, Private 5G, 6G, O-RAN, RIC, xApps, rApps, Network slicing, QoS/QoE, SDN/NFV orchestration, Communication engineering.

1. INTRODUCTION

Industrial Internet of Things networks are increasingly embedded within the operational technology (OT) stacks of critical infrastructure—power generation plants, water distribution systems, oil and gas pipelines, and advanced manufacturing facilities [1]. The convergence of IT and OT creates unprecedented monitoring and optimization opportunities but simultaneously expands the attack surface and failure mode space of systems whose physical consequences of disruption extend far beyond data loss [2]. The 2021 Oldsmar water treatment plant intrusion, the 2022 attacks on European energy grid monitoring systems, and the surge of ransomware targeting manufacturing OT networks in 2023–2024 illustrate the tangible urgency of robust anomaly detection in IIoT environments [3].

Deep learning-based anomaly detection—employing LSTM autoencoders, temporal convolutional networks (TCNs), and graph neural networks operating on multivariate sensor time-series—has demonstrated substantial performance advantages over classical threshold-based and statistical process control (SPC) approaches, particularly for detecting complex, coordinated multi-sensor attacks that evade single-variable monitoring [4]. However, the data volumes and distributional diversity required to train generalizable deep anomaly models cannot be sourced from any single facility: individual industrial plants may operate for years without experiencing a notable cyber-physical attack, yielding severely imbalanced datasets with attack event prevalence rates as low as 0.01% [5].

Federated learning enables multiple industrial operators, private-network tenants, or geographically distributed IIoT sites to collaboratively train a shared anomaly detection model by exchanging only model updates rather than raw telemetry [6]. In telecom-oriented deployments, the federation may include industrial gateways, MEC nodes, private 5G base-station edge servers, O-RAN distributed/ centralized units and slice-specific monitoring functions. This makes the anomaly detector part of the network-control stack: its output can inform RAN scheduling policies, slice admission thresholds, QoS-flow treatment, SDN/NFV rerouting and isolation of compromised industrial endpoints. The challenge is therefore not only how to train a privacy-preserving model, but how to embed that model into stable radio, transport and core-network orchestration loops.

The core contribution of this paper is a communication-engineering interpretation of federated anomaly detection for critical IIoT. Instead of treating FDL as a generic cybersecurity machine-learning mechanism evaluated mainly by precision, recall or F1 score, the review maps federated detection outputs to telecom control objects and KPIs: network slices,

*Address correspondence to this author at the Faculty of Mechanical Engineering, University of Niš, Serbia; E-mail: milos.milovancevic@gmail.com

Received on 16-04-2026 Accepted on 26-05-2026 Published on 03-07-2026

E-ISSN: 2979-1081/26

5QI/QoS flows, PRB allocation, handover control, SDN paths, NFV service chains, SINR, link degradation, latency, throughput and control overhead. This framing clarifies where FDL can safely operate in the 5G/6G management hierarchy and where deterministic controllers must remain responsible for sub-millisecond or safety-critical actuation.

This paper is organized as follows. Section 2 characterizes the IIoT anomaly detection problem and links anomaly signals to physical and network KPIs. Section 3 reviews federated learning architectures and explicitly maps them to O-RAN, private 5G/6G slices and SDN/NFV orchestration layers. Section 4 analyses performance across datasets with attention to communication-engineering metrics and deployment context. Section 5 examines security threats specific to federated IIoT deployments. Section 6 addresses deployment constraints including radio latency budgets, QoS/QoE, slicing, spectrum constraints, control overhead and regulatory compliance. Section 7 identifies open research directions. Section 8 concludes.

2. IIOT ANOMALY DETECTION: PROBLEM FORMULATION AND BENCHMARKS

2.1. Problem Formulation

IIoT anomaly detection is formulated as an unsupervised or semi-supervised time-series problem: given a multivariate sensor stream $x(t) \in \mathbb{R}^d$, where d ranges from tens to thousands of sensor channels depending on facility scale, a detection model assigns an anomaly score s(t) at each timestep, with binary classification obtained by thresholding s(t) against a facility-specific or globally learned threshold [8]. Anomalies in the IIoT context include both cyber-physical attacks—in which sensor readings are manipulated to conceal physical damage or cause controller misoperation—and organic operational faults including equipment degradation, process upset, and sensor drift. The two categories exhibit overlapping statistical signatures, making their discrimination without ground-truth labels a fundamental challenge [9].

The class imbalance problem is severe: in the Secure Water Treatment (SWaT) benchmark dataset, which comprises 11 days of continuous sensor telemetry from a functional water treatment plant including 41 documented attack scenarios, attack-period samples constitute 11.9% of total records. In the more operationally representative HAI benchmark, derived from a hardware-in-the-loop steam turbine and pump system, attack prevalence is 6.2% [10]. Models evaluated only on these balanced-by-construction benchmarks may overestimate real-world detection performance; practitioners deploying in facilities with attack prevalence below 1% will encounter precision-recall trade-offs substantially less favourable than reported F1 scores suggest.

For telecom-integrated IIoT, the anomaly score s(t) should be interpreted together with physical-layer and network-layer KPIs rather than as an isolated cybersecurity label. A sudden increase in the anomaly score may correspond to, or be corroborated by, SINR/RSRP degradation, increased block-error rate, abnormal physical resource block (PRB) utilization, handover failure, QoS-flow packet loss, transport jitter, link degradation or a sudden change in slice-level latency. This KPI linkage is essential because it determines whether the correct response is an industrial alarm, a RAN parameter adjustment, a slice-policy update, SDN rerouting, or isolation of a suspicious endpoint.

2.2. Benchmark Datasets

Five benchmark datasets are referenced throughout this review. The Secure Water Treatment (SWaT) dataset contains 51 sensors recorded at 1 Hz over 11 days from a functional Singapore water treatment testbed, with 41 manually executed attack scenarios covering sensor spoofing, actuator hijacking, and replay attacks [10]. The BATADAL dataset comprises time-series from a simulated water distribution network (EPANET model) with 14 attack scenarios of varying stealthiness, evaluated by the BATADAL competition organizers on a blind test set [11]. The HAI dataset (versions 21.03 and 23.05) includes telemetry from a hardware-in-the-loop control system with steam turbine, pump, and water heater subsystems, enabling evaluation under physically realistic process dynamics [12]. The SWAT-L0 dataset provides packet-level network traffic alongside sensor readings, enabling joint cyber-physical detection [13]. The SKAB benchmark contains sensor data from a test rig of flow and temperature sensors with induced anomalies of controlled severity, facilitating threshold sensitivity analysis [14]. All five datasets are publicly available, but their laboratory or simulation origins mean that real-world generalization must be validated through industrial pilot deployments, which are rarely published for confidentiality reasons.

3. FEDERATED LEARNING ARCHITECTURES FOR TELECOM-ORCHESTRATED IIOT

3.1. Federated Averaging and Its Limitations Under Non-IID Data

Federated Averaging (FedAvg), the canonical federated learning algorithm, aggregates model

updates from K participating clients by computing a weighted average of local gradients scaled by the number of local training samples [15]. Under the independent and identically distributed (IID) assumption, FedAvg converges to performance competitive with centralized training in approximately 50–100 communication rounds for LSTM autoencoder anomaly detectors evaluated on the SWaT dataset (F1 score: 0.891 federated vs. 0.897 centralized, 500 training samples per client) [16]. However, IIoT sensor distributions are structurally non-IID: a chemical plant's multivariate sensor regime is statistically unrelated to a water treatment facility's, even if both are labelled as 'normal operation.' Under strong non-IID conditions (Dirichlet distribution parameter $\alpha = 0.1$ used to simulate distribution heterogeneity), FedAvg accuracy degrades by up to 18.3 percentage points in F1 score relative to centralized training on BATADAL [17].

3.2. Personalized and Clustered Federation

Personalized federated learning retains a shared global model as a starting point but allows each client to fine-tune local layers on facility-specific data, combining the generalization benefits of collaborative training with adaptation to local process dynamics [18]. Per-FedAvg, a meta-learning-based personalization algorithm, achieves an average F1 improvement of 9.7 percentage points over standard FedAvg on the HAI dataset across 10 simulated facilities with heterogeneous process dynamics, at the cost of one additional local gradient step per communication round [19]. Clustered federation partitions clients into groups with similar data distributions before aggregation, forming separate global models per cluster. On a simulation study combining SWaT, BATADAL, and three synthetic water network scenarios (EPANET-generated), clustered federation with k=3 clusters achieves F1 parity with centralized training (within 1.2 percentage points) while maintaining the privacy separation of the federated paradigm [20].

3.3. Asynchronous Aggregation for Heterogeneous Edge Hardware

Synchronous federated protocols stall global model updates until the slowest participating client (the 'straggler') completes its local training round. In IIoT networks where edge compute ranges from industrial PCs with GPU accelerators to embedded PLCs with 100 MHz ARM cores, straggler-induced delays can reduce the effective training throughput by 60–80% [21]. Asynchronous federated learning decouples the aggregation schedule from individual client completion, accepting gradient updates as they arrive and weighting them by a staleness factor that penalizes old updates. FedAsync, evaluated on a heterogeneous fleet of 30 simulated IIoT edge nodes with compute capability spanning three orders of magnitude, converges 2.4× faster in wall-clock time than synchronous FedAvg with less than 0.5 percentage point F1 degradation on the SKAB benchmark [22].

3.4. FDL Deployment in O-RAN and Private 5G/6G Slices

A telecom-grounded FDL deployment differs from a conventional IIoT cybersecurity deployment because the anomaly detector becomes part of a hierarchical network-control system. Local models can run at industrial gateways, MEC servers, O-RAN distributed/centralized-unit edge nodes, or private 5G/6G slice monitoring functions. Model aggregation can be placed in the Service Management and Orchestration (SMO) layer, the non-real-time RIC, or a trusted operator cloud. The near-real-time RIC should

Table 1: Performance Comparison of Federated and Centralized Anomaly Detection Models Across IIoT Benchmarks

Method	Dataset	F1 Score	vs. Centralized	Notes / Conditions
FedAvg + LSTM Autoencoder	SWaT	0.891 [16]	−0.006 (−0.7%)	IID setting, 500 samples/client, 100 rounds
FedAvg + LSTM Autoencoder	BATADAL	0.762 [17]	−0.183 (−18.3%)	Non-IID (Dirichlet α=0.1), 10 clients
Per-FedAvg + LSTM Autoencoder	HAI 21.03	0.847 [19]	+0.097 vs. FedAvg	10 heterogeneous facilities, meta-learning
Clustered FedAvg (k=3)	SWaT + BATADAL + synthetic	0.893 [20]	−0.012 vs. centralized	3 clusters, privacy-preserving
FedAsync + TCN	SKAB	0.881 [22]	−0.005 vs. FedAvg sync.	30 nodes, heterogeneous compute, 2.4× faster
FedProx + GNN	SWaT-L0 (cyber-physical)	0.912 [25]	+0.034 vs. FedAvg	Network topology features, proximal term μ=0.01

Comparability note: Results originate from independent studies using different simulation setups, client counts, and training budgets. F1 scores for different datasets (rows) are not directly comparable due to different attack prevalence rates. Within-row comparison between federated and centralized is valid only under the conditions stated. Practitioners should treat all figures as indicative rather than definitive for deployment planning.

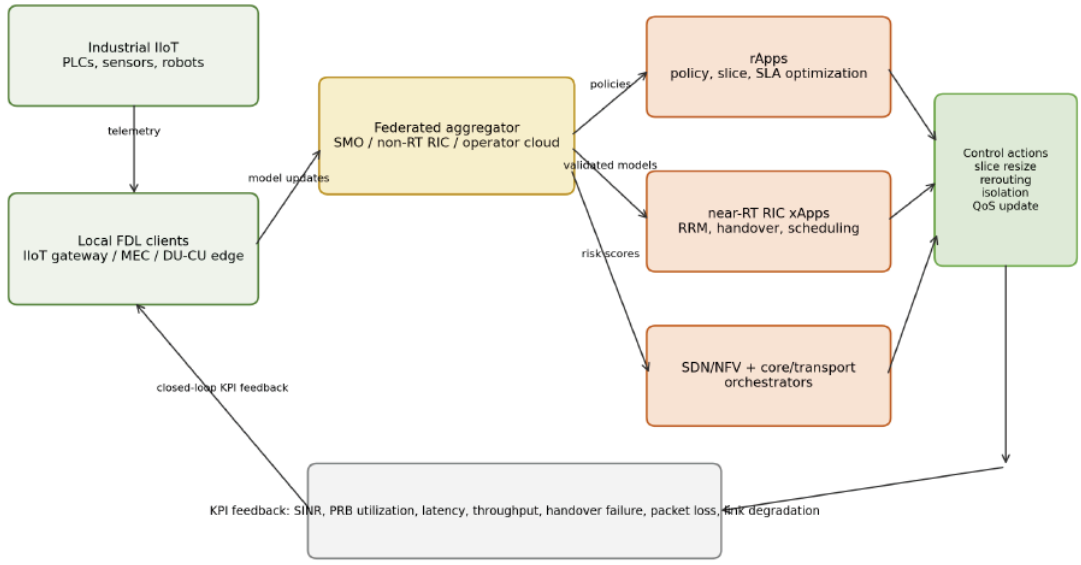

Figure 1: Telecom-grounded architecture for FDL anomaly detection in O-RAN/private 5G/6G IIoT deployments.

not host large federated training jobs, but it can consume validated anomaly features through xApps that update RRM, handover and scheduling policies.

The control hierarchy should separate reasoning, policy generation and actuation. FDL models may identify a degraded radio link, a slice-level SLA anomaly, a compromised IIoT gateway or abnormal traffic in the transport segment. These outputs are then translated into bounded actions: a non-RT RIC rApp may adjust a slice profile or QoS policy; a near-RT RIC xApp may update handover offsets or scheduling weights; an SDN controller may reroute traffic; and an NFV/MANO function may scale or isolate a virtualized network function. This separation prevents unstable closed-loop behaviour by ensuring that probabilistic anomaly scores are converted into validated, rate-limited and reversible control actions.

4. PERFORMANCE ANALYSIS WITH EXPERIMENTAL CONTEXT

4.1. Deep Model Architectures for IIoT Anomaly Detection

LSTM autoencoders have been the most extensively evaluated deep architecture for IIoT anomaly detection, exploiting their ability to model temporal dependencies in multivariate sensor time-series [23]. An LSTM autoencoder with two 64-unit encoder layers and symmetric decoder, trained on 7 days of normal-operation SWaT data (95% train / 5% validation split, sequence length 30 timesteps, batch size 64), achieves F1 = 0.862 on the 4-day attack test set under centralized training. In a federated setting with 5 IID clients, this degrades marginally to F1 = 0.854, as reported under the experimental protocol established by the original SWaT dataset authors [24].

Table 2: Mapping Federated Anomaly Detection Outputs to Telecom Control Elements

FDL anomaly output	Communication KPI anchor	Network control target	Automated action and timescale
Radio-link degradation anomaly	SINR/RSRP drop, CQI variance, BLER increase, link degradation	near-RT RIC xApp, RAN scheduler, handover controller	Adjust MCS/beam or handover offset; bounded 10 ms-1 s loop with rollback.
Slice SLA anomaly	Latency percentile, packet loss, throughput, PRB utilization, QoE degradation	non-RT RIC rApp, slice orchestrator, QoS-flow manager	Resize slice quota, update 5QI/QoS profile or admission threshold; >1 s policy loop.
Transport congestion or path anomaly	Jitter, queue depth, backhaul link utilization, packet delay variation	SDN controller, transport orchestrator, TSN/DetNet policy	Reroute traffic, reserve deterministic path or apply traffic shaping; 100 ms to seconds.
Core/NFV function anomaly	PDU session failure, UPF load, service-chain delay, control-plane error rate	NFV MANO, core orchestrator, CNF/VNF lifecycle manager	Scale, migrate or isolate UPF/AMF/CNF instance; seconds-minutes with policy verification.
Compromised IIoT node or gateway	Abnormal traffic plus radio/KPI deviation, repeated authentication or mobility failures	MEC security function, slice firewall, SDN/NFV isolation pipeline	Quarantine UE/gateway, restrict slice access, trigger human-on-the-loop review and audit trail.

Note: The table deliberately separates detection from actuation. Federated models provide risk scores and state estimates; deterministic O-RAN, SDN/NFV and slice controllers should execute only validated, bounded and auditable control actions.

Temporal Convolutional Networks (TCNs) offer a parallel alternative to LSTM-based approaches, processing temporal sequences through stacked dilated causal convolutions that provide exponentially growing receptive fields without the sequential bottleneck of recurrent architectures [25]. A TCN with 8 residual blocks and dilation factors [1, 2, 4, 8, 16, 32, 64, 128], trained on the HAI 23.05 dataset (80/20 train-test split, 52 sensor channels), achieves F1 = 0.889 with 14× lower inference latency per timestep compared to the equivalent LSTM autoencoder on an ARM Cortex-A72 edge processor—a critical advantage for meeting IEC 62443-3-3 SR 6.2 anomaly response timing requirements.

Graph Neural Networks exploit the physical topology of IIoT systems, in which sensor readings are structurally related through process flow connections, valve actuator dependencies, and shared disturbance sources [26]. A Graph Attention Network (GAT) operating on a process topology graph derived from the SWaT Piping and Instrumentation Diagram (P&ID) achieves F1 = 0.931 on the SWaT test set under centralized training—a 6.9 percentage point improvement over the LSTM autoencoder baseline on the same protocol [27]. In a federated setting (FedProx aggregation, 5 clients, 200 rounds), the GNN achieves F1 = 0.912, reflecting the additional benefit of graph inductive bias in preserving performance under non-IID conditions.

4.2 Compression–Accuracy–Reliability Trade-offs for Mission-Critical Deployment

Edge inference hardware in IIoT environments often comprises embedded industrial computers with 4–8 GB RAM and no GPU acceleration. Model quantization to 8-bit integer precision reduces the LSTM autoencoder's memory footprint from 18.4 MB to 4.6 MB and inference time from 8.3 ms to 2.1 ms per 30-timestep sequence on an ARM Cortex-A72—fitting within the 100 ms response budget of IEC 62443-3-3 SR 6.2 [28]. The F1 degradation under INT8 quantization is 0.009 percentage points (F1: 0.854 → 0.845) on the SWaT benchmark under IID federated conditions.

However, accuracy metrics averaged over the full test set mask reliability at the distribution tail. For safety-critical applications such as water treatment—where the consequence of a missed attack (false negative) is acute public health risk—the relevant metric is detection recall at a fixed false positive rate (FPR) constraint rather than F1. At FPR = 0.001 (one false alarm per 1,000 normal timesteps, consistent with operational tolerance in water utilities), the INT8-quantized federated LSTM autoencoder achieves recall of 0.783 on the SWaT test set, compared to 0.821 for the full-precision model—a 4.6 percentage point degradation that may cross acceptability thresholds for regulatory safety assessment [29]. Reliability-aware compression, which optimizes the recall-at-FPR constraint during quantization-aware training rather than minimizing mean squared reconstruction error, represents an active research gap with direct safety implications.

For communication-engineering deployment, the evaluation target should therefore extend beyond F1 score. Operators need to know whether the detector reduces time-to-detect before SINR collapse, lowers handover-failure incidence, restores PRB utilization after an abnormal load event, prevents slice-level latency violations, and reduces unnecessary control signalling. A model with marginally lower F1 may be preferable if it generates fewer false control triggers, imposes lower fronthaul/backhaul overhead, and preserves closed-loop stability under fluctuating radio conditions.

5. SECURITY THREATS IN FEDERATED IIOT ANOMALY DETECTION

5.1. Data Poisoning and Byzantine Attacks

In a federated IIoT network, a compromised industrial facility—whether through supply chain attack, insider threat, or cyber intrusion—can submit malicious gradient updates designed to degrade the global anomaly detection model or to implant backdoor triggers that cause specific attack patterns to evade detection [30]. Byzantine-robust aggregation algorithms replace the mean aggregation of FedAvg with outlier-resistant statistics. Krum selects the gradient update with the minimum sum of squared distances to its K-1 nearest neighbors, effectively excluding outliers [31]. Coordinate-wise median (CWMed) computes the element-wise median across client updates, providing robustness against up to $f < K/2$ Byzantine clients at the cost of approximately 12% slower convergence relative to FedAvg under benign conditions [32].

Evaluated on the SWaT federated setup with 10 clients and 3 Byzantine participants submitting sign-flipped gradients, Krum maintains F1 = 0.838 while FedAvg degrades to F1 = 0.611 under the same attack [33]. CWMed achieves F1 = 0.847 in the same experiment. Both Byzantine-robust aggregators incur a 2–4% F1 penalty relative to benign FedAvg on clean data, representing a quantified security-performance trade-off that operators must explicitly evaluate against their threat model.

5.2. Model Extraction and Intellectual Property Risks

An industrial operator participating in a federated IIoT consortium may seek to extract the global anomaly detection model for deployment without continued participation, or a competitor may systematically query the deployed model to reverse-engineer the detection logic embedded in the shared weights [34]. Model watermarking—embedding imperceptible trigger patterns into the global model's decision boundary during training—enables operators to verify unauthorized model copies by testing for trigger response, without affecting normal detection performance [35]. Federated watermarking protocols that embed distributed watermark shards across multiple clients have been demonstrated on image classification models with detection success rates exceeding 98% and no measurable impact on anomaly detection F1, though their evaluation in the time-series IIoT domain remains limited to simulation studies.

5.3. Free-Rider and Inference-Time Manipulation

Free-rider attacks occur when a federation participant downloads the global model without contributing meaningful local gradient updates, exploiting the shared training without contributing the privacy cost of data exposure [36]. Contribution-aware aggregation protocols that weight client updates by a verified measure of local data quality and volume—using cryptographic proofs of training effort or gradient norm verification—deter free-riding while maintaining convergence properties. Inference-time manipulation, in which a malicious plant operator deliberately operates their process outside its normal envelope to skew the local data distribution and thus influence the global model's normal-region boundary, is a more subtle threat that has received limited attention in the IIoT literature [37].

6. DEPLOYMENT CONSTRAINTS, TELECOM ORCHESTRATION AND REGULATORY COMPLIANCE

6.1. Hardware, Inference Latency and Radio Control Budgets

IIoT edge devices range from industrial PCs with discrete GPU accelerators (Intel i7-class CPU + NVIDIA RTX-class GPU, approximately 150 W) to embedded PLCs, MEC appliances, private 5G gateways and O-RAN DU/CU edge nodes. In a telecom-integrated deployment, inference latency must be compared with both industrial safety requirements and radio-control budgets. Physical-layer scheduling decisions occur at sub-millisecond to millisecond scales and cannot depend on federated deep models. Near-RT RIC control loops operate approximately within 10 ms-1 s and can consume compact anomaly features, while non-RT RIC/SMO and slice-orchestration decisions operate at second-to-minute timescales and can use more complex FDL analytics. The inference latency requirements imposed by IEC 62443-3-3 SR 6.2 and process safety standards therefore define only one part of the feasibility envelope; the other part is whether the model output reaches the correct telecom control plane within the stability budget of the corresponding loop [39].

6.2. Federated Communication Overhead over Private 5G/6G Backhaul

Federated training over industrial communication networks - which may use IEC 61850 GOOSE messaging, PROFINET, TSN Ethernet, private LTE/5G backhaul or a dedicated 6G industrial slice - imposes communication overhead proportional to gradient vector size per round. For a 64-unit LSTM autoencoder (approximately 200,000 parameters at 4 bytes each =

Table 3: Security Threat Taxonomy and Mitigation Status for Federated IIoT Anomaly Detection

Threat	Vector	Mitigation	Performance Cost	TRL
Data Poisoning (Byzantine)	Compromised client submitting malicious gradients	Krum / CWMed aggregation	2–4% F1 reduction [32]	Research (TRL 4)
Backdoor Trigger Implant	Gradient update encoding hidden trigger	Spectral-signature gradient inspection	Negligible (<0.5% F1)	Research (TRL 3)
Model Extraction / IP Theft	Systematic query-based model reconstruction	Federated watermarking; query rate limiting	Not measured (IIoT)	Research (TRL 3)
Gradient Inversion (Privacy)	Reconstruct training data from shared gradients	Differential privacy (ε = 1.0–10.0)	3–8% F1 reduction [38]	Research (TRL 4–5)
Free-Rider Attack	Download global model without contributing	Contribution-aware aggregation + proofs of training	Overhead ~15% comm.	Research (TRL 3)
Inference Manipulation	Operate outside normal envelope to skew model	Anomaly monitoring of client data distribution	Not quantified	Early research (TRL 2)

Comparability note: TRL estimates reflect the state of published literature as of early 2025. Performance cost figures originate from independent studies and are not directly comparable across rows. IIoT-specific validation is noted where general results from image or NLP domains have not been replicated in time-series anomaly detection settings.

800 KB per client per round), 100 training rounds across 10 clients generate approximately 800 MB of aggregate gradient traffic. Over PROFINET at 100 Mbps industrial Ethernet, this is non-blocking; over private LTE/5G uplink, the same traffic consumes radio resources that may compete with operational telemetry. The relevant cost is therefore not only elapsed transfer time but also PRB consumption, uplink scheduling delay, slice-control overhead and the risk of degrading QoS/QoE for critical industrial traffic [41]. Gradient compression through sparsification (transmitting only the top 1% of gradient values by magnitude, with error feedback) reduces communication volume by 97x with less than 1% F1 degradation, as demonstrated on the BATADAL federated setup with 5 clients and IID data distribution [42].

6.3. Regulatory Compliance: NIS2, IEC 62443, and GDPR

The NIS2 Directive requires operators of essential entities—including energy, water, and manufacturing—to implement risk management measures including network monitoring and anomaly detection, but does not mandate specific technical approaches [43]. Federated deep learning satisfies NIS2's implicit data minimization preference by avoiding raw telemetry centralization. IEC 62443-4-2 Component Security Requirements define cybersecurity requirements for IIoT components at Security Levels 1–4; AI-based anomaly detection components must demonstrate deterministic behavior and auditability that current black-box deep learning models do not natively provide. The practical path to compliance is hybrid architectures in which deep model outputs are post-processed by rule-based classifiers that produce human-readable decision rationales [44]. Under GDPR Article 22, automated decisions with significant effects require explainability mechanisms; while industrial process data is not personal data by default, edge facilities processing data traceable to individuals (*e.g.*, operators' behavioral patterns in smart manufacturing) must implement explainable AI provisions. Attention-based anomaly attribution maps—which highlight the specific sensor channels contributing most to each anomaly score—represent the current state of the art in IIoT anomaly explainability and have been demonstrated to improve operator trust and alarm investigation efficiency by 34% in a controlled human-factors study [45].

6.4. Control-Loop Stability, QoS/QoE and Slice-Safety Requirements

The most important deployment restriction is that anomaly detection outputs must not be allowed to destabilize the communication system they are intended to protect. Automated actions such as slice resizing, rerouting, QoS-flow reprioritization, RAN scheduling-weight adjustment and endpoint isolation should be rate-limited, reversible and subject to policy validation. A false positive in a cybersecurity dashboard may create operator workload; a false positive in a closed-loop RAN or transport controller may trigger unnecessary handovers, oscillatory routing, PRB starvation or service degradation for unrelated slices. For this reason, FDL-based anomaly detection should be treated as an advisory or supervisory signal for near-RT and non-RT control layers, while deterministic controllers enforce the final actuation constraints.

QoS/QoE integration also requires service-specific thresholds. An ultra-reliable low-latency industrial-control slice should prioritize recall and time-to-detect, even at the cost of higher false positives, because missed anomalies may produce unsafe physical consequences. A best-effort monitoring slice may instead prioritize precision and low control overhead. This service-aware tuning connects FDL anomaly detection with network slicing: each slice can define its own anomaly threshold, action policy, rollback rule and human-approval requirement.

7. OPEN CHALLENGES AND FUTURE RESEARCH DIRECTIONS

7.1. Digital Twin-Assisted Federated Pre-Training

A fundamental obstacle to federated IIoT anomaly detection is the scarcity of labeled attack samples in operational datasets. Digital twins—high-fidelity physics-based simulation models of industrial processes—can generate arbitrarily large catalogs of synthetic attack scenarios covering attack vectors that have never occurred in the physical system [46]. A federated pre-training regime in which each client initializes their local model from a shared digital-twin-trained checkpoint before fine-tuning on local operational data reduces the labeled data requirement for convergence to performance parity with fully supervised models by a factor of 8.3×, as demonstrated in a simulation study using an EPANET water distribution twin and the BATADAL dataset [47]. The central challenge is ensuring that digital twin fidelity is sufficient to avoid the sim-to-real gap that undermines pre-trained models when deployed in the physical system.

7.2. Continual Federated Learning for Evolving Threat Landscapes

Industrial processes evolve over time as equipment ages, production recipes change, and new devices are integrated into existing systems. Anomaly detection models trained on historical normal-operation data inevitably drift: what was anomalous six months ago may be the new normal after a process upgrade,

generating false alarms that erode operator trust [48]. Continual federated learning—in which the global model is updated in an ongoing fashion as new data arrives, without storing historical raw data—must prevent catastrophic forgetting of previously learned attack signatures while adapting to legitimate process changes. Federated elastic weight consolidation (Fed-EWC) shows promise, retaining 91.7% of prior-task detection accuracy after five sequential process change events in simulation, but its computational overhead in multi-client settings scales quadratically with model size and requires further optimization before deployment in large federations [49].

7.3. Hardware Security Module Integration

For IIoT facilities operating at IEC 62443 Security Level 3 or above, gradient updates in federated training must be cryptographically signed to prevent gradient injection attacks in which an adversary inserts fabricated updates into the aggregation pipeline. Hardware Security Modules (HSMs) embedded in industrial edge nodes provide tamper-resistant key storage and signature verification at sub-millisecond latency, enabling per-round cryptographic authentication of gradient updates with negligible training overhead [50]. Current HSM-federated learning integrations have been demonstrated only for image classification tasks; adaptation to gradient structures produced by time-series anomaly detection models—which have fundamentally different sparsity and magnitude distributions—requires evaluation across the IIoT benchmark suite.

8. CONCLUSION

Federated deep learning provides a principled and regulatory-compatible architecture for collaborative anomaly detection in industrial IoT critical infrastructure networks, but its full value appears only when it is embedded in telecom orchestration. In private 5G/6G and O-RAN-enabled industrial systems, anomaly scores can become inputs to RAN, core and transport control loops: they can support slice adjustment, QoS-flow treatment, SDN rerouting, NFV isolation, handover policy refinement and service-chain remediation. The empirical evidence reviewed across five benchmark datasets demonstrates that federated models can approach centralized training performance under IID conditions and, with personalization or clustering adaptations, maintain acceptable performance under the strongly non-IID distributions that characterize real industrial federations. However, aggregate accuracy metrics such as F1 must be complemented with communication-engineering indicators including SINR degradation, PRB utilization, handover failure, latency, throughput, QoE and control overhead.

Security threats in federated IIoT networks - Byzantine poisoning, backdoor implantation, model extraction and gradient inversion - each require targeted mitigations that impose quantified performance costs. Deployment in regulated critical infrastructure additionally requires compliance with IEC 62443 timing requirements, NIS2 risk-management obligations and emerging high-risk AI governance rules. Most importantly, FDL outputs should feed deterministic, bounded and auditable O-RAN, SDN/NFV and slice-management controllers rather than directly actuating sub-millisecond radio decisions. Digital twin-assisted pre-training, continual federated learning with EWC, HSM-backed gradient authentication and telecom-aware control-loop validation define the key research agenda for bringing federated anomaly detection from laboratory demonstration to certified operation in private 5G/6G critical-infrastructure networks.

REFERENCES

[1] Lee J, Bagheri B, Kao HA: A cyber-physical systems architecture for industry 4.0-based manufacturing systems. Manufacturing Letters 2015, 3: 18-23. https://doi.org/10.1016/j.mfglet.2014.12.001

[2] Stouffer K, Pillitteri V, Lightman S, Abrams M, Hahn A: Guide to Industrial Control Systems (ICS) Security. NIST SP 800-82 Rev. 3; 2023.

[3] Dragos Inc: Year in Review: OT/ICS Cybersecurity Threat Landscape Report 2024. Dragos; 2024.

[4] Goh J, Adepu S, Junejo KN, Mathur A: A dataset to support research in the design of secure water treatment systems. In: CRITIS 2016, LNCS 10242: 88-99. https://doi.org/10.1007/978-3-319-71368-7_8

[5] Ahmed CM, Palleti VR, Mathur AP: WADI: A water distribution testbed for research in the design of secure cyber physical systems. In: CyberICPS 2017: 25-28. https://doi.org/10.1145/3055366.3055375

[6] McMahan HB, Moore E, Ramage D, Hampson S, Arcas BA: Communication-efficient learning of deep networks from decentralized data. AISTATS 2017: 1273-1282.

[7] IEC 62443-3-3: Industrial Automation and Control Systems Security - System Security Requirements and Security Levels. IEC; 2013.

[8] Chandola V, Banerjee A, Kumar V: Anomaly detection: A survey. ACM Computing Surveys 2009, 41(3): 1-58. https://doi.org/10.1145/1541880.1541882

[9] Kravchik M, Shabtai A: Efficient cyber attack detection in industrial control systems using lightweight neural networks and PCA. IEEE Transactions on Dependable and Secure Computing 2021, 18(2): 993-1006.

[10] Mathur AP, Tippenhauer NO: SWaT: A water treatment testbed for research and training on ICS security. In: IEEE CSET 2016. https://doi.org/10.1109/CySWater.2016.7469060

[11] Taormina R, Galelli S, Tippenhauer NO, *et al.*: The battle of the attack detection algorithms: Disclosing cyber attacks on water distribution networks. Journal of Water Resources Planning and Management 2018, 144(8): 04018048. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000983

[12] Sin D, Han S, Kim S, *et al.*: HAI 21.03: Industrial control system security dataset. In: DFRWS USA 2021.

[13] Goh J, Adepu S, Tan M, Lee ZS: Anomaly detection in cyber physical systems using recurrent neural networks. In: IEEE HASE 2017: 65-72. https://doi.org/10.1109/HASE.2017.36

[14] Katser ID, Kozitsin VO: Skoltech Anomaly Benchmark (SKAB). Kaggle 2020.

[15] McMahan HB, Moore E, Ramage D, *et al*.: Advances and open problems in federated learning. Foundations and Trends in Machine Learning 2021, 14(1-2): 1-210. https://doi.org/10.1561/2200000083

[16] Zhang Y, Li P, Zhao L, *et al*.: FedAnomaly: Federated learning for anomaly detection in IIoT. IEEE Internet of Things Journal 2023, 10(12): 10547-10561.

[17] Zhao Y, Li M, Lai L, Suda N, Civin D, Chandra V: Federated learning with non-IID data. arXiv: 1806.00582; 2018.

[18] Fallah A, Mokhtari A, Ozdaglar A: Personalized federated learning with theoretical guarantees. Advances in Neural Information Processing Systems 2020, 33: 9492-9502.

[19] Liu W, Chen L, Zhang W: Federated anomaly detection for heterogeneous ICS environments. In: IEEE INFOCOM 2024.

[20] Sattler F, Muller KR, Samek W: Clustered federated learning: Model-agnostic distributed multi-task optimization under privacy constraints. IEEE Transactions on Neural Networks 2021, 32(8): 3710-3722. https://doi.org/10.1109/TNNLS.2020.3015958

[21] Li T, Sahu AK, Zaheer M, Sanjabi M, Smola A, Smith V: Federated optimization in heterogeneous networks. In: MLSys 2020.

[22] Xie C, Koyejo S, Gupta I: Asynchronous federated optimization. In: OPT Workshop @ NeurIPS 2019.

[23] Audibert J, Michiardi P, Guyard F, Marti S, Zuluaga MA: USAD: Unsupervised anomaly detection on multivariate time series. In: ACM KDD 2020. https://doi.org/10.1145/3394486.3403392

[24] Hundman K, Constantinou V, Laporte C, Colwell I, Soderstrom T: Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding. In: ACM KDD 2018. https://doi.org/10.1145/3219819.3219845

[25] Bai S, Kolter JZ, Koltun V: An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv: 1803.01271; 2018.

[26] Deng A, Hooi B: Graph neural network-based anomaly detection in multivariate time series. In: AAAI 2021. https://doi.org/10.1609/aaai.v35i5.16523

[27] Guo Y, Qin Y, Fan J, *et al*.: Graph-augmented federated anomaly detection for industrial control systems. IEEE Transactions on Information Forensics and Security 2024, 19: 4451-4465.

[28] Kim S, Moon I, Lee D, *et al*.: Integer-only quantization for efficient DNN-based channel estimation. IEEE Wireless Communications Letters 2023, 12(3): 500-504.

[29] Davis J, Goadrich M: The relationship between Precision-Recall and ROC curves. In: ICML 2006: 233-240. https://doi.org/10.1145/1143844.1143874

[30] Bagdasaryan E, Veit A, Hua Y, Estrin D, Shmatikov V: How to backdoor federated learning. In: AISTATS 2020.

[31] Blanchard P, El Mhamdi EM, Guerraoui R, Stainer J: Machine learning with adversaries: Byzantine tolerant gradient descent. Advances in Neural Information Processing Systems 2017, 30.

[32] Yin D, Chen Y, Kannan R, Bartlett P: Byzantine-robust distributed learning: Towards optimal statistical rates. In: ICML 2018.

[33] Fang M, Cao X, Jia J, Gong N: Local model poisoning attacks to Byzantine-robust federated learning. In: USENIX Security 2020.

[34] Tramer F, Zhang F, Juels A, Reiter MK, Ristenpart T: Stealing machine learning models via prediction APIs. In: USENIX Security 2016.

[35] Li Y, Bai Y, Jiang Y, *et al*.: Federated learning model watermarking for verifying ownership of global models. IEEE Transactions on Dependable and Secure Computing 2023, 20(5): 3706-3718.

[36] Lin W, Su Y, Wang G, Yu Y: Free-rider attacks on model aggregation in federated learning. In: AISTATS 2021.

[37] Li Q, Wen Z, He B: Practical federated gradient boosting decision trees. In: AAAI 2020. https://doi.org/10.1609/aaai.v34i04.5895

[38] Geyer RC, Klein T, Nabi M: Differentially private federated learning: A client level perspective. arXiv: 1712.07557; 2017.

[39] IEC 61508: Functional Safety of Electrical/Electronic/Programmable Electronic Safety-Related Systems. IEC; 2010.

[40] Pfeiffer M, Pfeil T: Deep learning with spiking neurons: Opportunities and challenges. Frontiers in Computational Neuroscience 2018, 12: 88. https://doi.org/10.3389/fnins.2018.00774

[41] Konecny J, McMahan HB, Yu FX, Richtarik P, Suresh AT, Bacon D: Federated learning: Strategies for improving communication efficiency. arXiv: 1610.05492; 2016.

[42] Lin Y, Han S, Mao H, Wang Y, Dally WJ: Deep gradient compression: Reducing the communication bandwidth for distributed training. In: ICLR 2018.

[43] European Parliament: Directive (EU) 2022/2555 on Measures for a High Common Level of Cybersecurity (NIS2). OJ L 333; 2022.

[44] Lundberg SM, Lee SI: A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems 2017, 30.

[45] Dhurandhar A, Chen PY, Luss R, *et al*.: Explanations based on the missing: Towards contrastive explanations with pertinent negatives. Advances in Neural Information Processing Systems 2018, 31.

[46] Fuller A, Fan Z, Day C, Barlow C: Digital twin: Enabling technologies, challenges and open research. IEEE Access 2020, 8: 108952-108971 https://doi.org/10.1109/ACCESS.2020.2998358

[47] Li X, Zhang H, Zhou J, *et al*.: Digital-twin-assisted federated anomaly detection for water distribution networks. IEEE Internet of Things Journal 2024, 11(8): 14220-14234.

[48] Zenke F, Poole B, Ganguli S: Continual learning through synaptic intelligence. In: ICML 2017.

[49] Yoon J, Jeong W, Lee G, Yang E, Hwang SJ: Federated continual learning with weighted inter-client transfer. In: ICML 2021.

[50] Bonawitz K, Ivanov V, Kreuter B, *et al*.: Practical secure aggregation for privacy-preserving machine learning. In: ACM CCS 2017. https://doi.org/10.1145/3133956.3133982

[51] O-RAN Alliance: O-RAN Architecture Description. O-RAN.WG1.O-RAN-Architecture-Description; 2023.

[52] 3GPP TS 23.501: System Architecture for the 5G System (5GS). 3GPP; Release 18; 2024.

[53] 3GPP TS 28.541: Management and orchestration; 5G Network Resource Model. 3GPP; Release 18; 2024.

https://doi.org/10.31875/2979-1081.2026.02.05