Int. J. Sci. R. Tech., 2026 3(9)
A Multidisciplinary peer-reviewed Journal
www.ijsrtjournal.com [ISSN: 2394-7063]

Original Article

Lightweight And Privacy-Preserving Intrusion Detection For The Internet Of Things: A Systematic Review Of Datasets, Models, And Emerging Directions

Uttish Bavoriya*

GDC Samba

ABSTRACT

The rapid expansion of Internet of Things (IoT) deployments has been matched by a corresponding rise in device- and network-targeted attacks, while the limited computation, memory, and energy budgets of most IoT endpoints make conventional, resource-intensive intrusion detection systems (IDS) difficult to deploy directly on-device. Over the past five years, a substantial and fast-growing literature has proposed lightweight machine learning (ML) and deep learning (DL) approaches to this problem, spanning classical feature-reduction pipelines, pruned and quantized deep architectures, spiking neural networks, TinyML-optimized models, and federated and privacy-preserving training schemes. This paper presents a systematic review of this literature, organized around four themes: (1) the benchmark datasets that underpin current evaluation practice, (2) model-level techniques for achieving detection accuracy under strict resource budgets, (3) federated and privacy-preserving training architectures suited to distributed IoT deployments, and (4) the emerging role of explainability in IoT IDS. We synthesize reported results across more than twenty recent studies, present a comparative summary of representative approaches, and identify recurring gaps in the literature, including limited cross-dataset generalization evidence, inconsistent resource-cost reporting, and uneven treatment of class imbalance for rare but high-impact attack types. We conclude with a set of open research directions intended to guide future work toward more reproducible and deployment-realistic evaluation of lightweight IoT intrusion detection.

Keywords: Internet of Things (IoT), TinyML-optimized models, literature, intrusion detection systems.

INTRODUCTION

Internet of Things devices now span consumer, industrial, and critical-infrastructure settings, and their growing numbers have expanded the attack surface available to adversaries, including botnet recruitment, denial-of-service, reconnaissance, and protocol-level exploitation. Prior surveys have established that IoT security requires solutions distinct from conventional enterprise network defense, given the constrained processing power, memory, and energy supply typical of IoT endpoints, and given the heterogeneity of IoT device populations and communication protocols. Intrusion detection is one of the central defensive mechanisms proposed for this setting, and a large body of work has focused specifically on making IDS models lightweight enough for on-device or near-edge deployment.

This review focuses on that lightweight IoT-IDS literature specifically, rather than IoT security broadly. Our motivation is that the field has grown quickly enough, across enough disconnected sub-threads, that a consolidated view of what has been tried, what results have been reported, and what remains open is useful both to newcomers and to researchers seeking to identify genuinely novel contributions rather than incremental replications. We restrict scope to detection (rather than prevention or response), to IoT/IIoT network and device traffic (rather than general-purpose network intrusion detection), and to approaches that explicitly report a resource-efficiency dimension (model size, inference latency, energy, or communication cost) alongside detection accuracy.

The remainder of this paper is organized as follows. Section 2 describes the review methodology,

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

including search strategy and inclusion criteria. Section 3 surveys benchmark datasets. Section 4 reviews model-level lightweighting techniques. Section 5 covers federated and privacy-preserving approaches. Section 6 discusses explainability. Section 7 presents a comparative synthesis across the reviewed studies. Section 8 discusses open challenges and future directions, and Section 9 concludes.

2. REVIEW METHODOLOGY

This review follows a narrative-systematic approach informed by standard systematic-review search practice. Search terms combined IoT/IIoT with intrusion detection, lightweight, TinyML, edge, federated, and quantization/pruning, applied to major indexing sources (IEEE Xplore, ScienceDirect, SpringerLink, arXiv, and general web search) for work published primarily between 2019 and 2026, with emphasis on 2023–2026 publications to capture the current state of the field. Studies were included if they (a) targeted IoT or IIoT network/device traffic specifically, (b) proposed or evaluated a detection approach with an explicit resource-efficiency consideration (model size, latency, energy, communication cost, or an equivalent constrained-deployment framing), and (c) reported quantitative detection performance on at least one named benchmark dataset. Purely theoretical or position papers without an evaluated model, and papers addressing physical-layer or protocol-level security mechanisms unrelated to learned detection models, were excluded. Given the volume and rapid turnover of this literature, this review does not claim exhaustive coverage; it prioritizes recent (2023–2026), well-cited, and thematically representative work sufficient to characterize the current landscape and its gaps.

3. BENCHMARK DATASETS

Evaluation in this literature relies on a small number of recurring benchmark datasets. The Bot-IoT dataset provided an early realistic botnet-focused benchmark combining legitimate and simulated botnet traffic for network forensic analytics. The ToN_IoT dataset family, developed at UNSW Canberra Cyber, extended this by integrating IoT sensor telemetry, Linux and Windows OS audit logs, and network flow records from a multi-layer edge–fog–cloud testbed, covering nine attack categories with pronounced class imbalance (for example, man-in-the-middle traffic representing roughly 1,000 of over 200,000 total records in commonly used subsets). CICIoT2023, released by the Canadian Institute for Cybersecurity, was generated from a topology of 105 real IoT devices and captures 33 attack types across seven categories (DDoS, DoS, reconnaissance, web-based, brute force, spoofing, and Mirai-related botnet activity), and its authors explicitly note dataset-transferability analysis as an open direction. Additional datasets recurring in the reviewed literature include Edge-IIoTset and RT-IoT2022, both used to evaluate industrial and real-time IoT scenarios, and N-BaIoT, a botnet-focused dataset collected from commercial IoT devices infected with Mirai and BASHLITE. A recurring observation across dataset-overview studies is the lack of standardization in feature sets and attack taxonomies across these benchmarks, which complicates direct cross-study comparison and motivates the harmonization concerns discussed in Section 8.

4. MODEL-LEVEL LIGHTWEIGHTING TECHNIQUES

4.1 Feature Reduction and Classical Classifiers

A recurring strategy pairs dimensionality reduction with classical classifiers. Reported work combining principal component analysis with metaheuristic feature selection (particle swarm optimization) ahead of a support vector machine classifier reduced a 41-feature representation to 18 features on NSL-KDD while reporting 98.5% accuracy, illustrating that substantial feature-set compression can be achieved with limited accuracy cost. Comparative studies of decision trees, gradient-boosted ensembles (e.g., LightGBM), and compact neural networks on ToN_IoT have found tree-based classifiers competitive on raw F1-score (reported as high as 0.987 for a decision tree) but substantially larger in serialized size (hundreds of kilobytes) than compressed neural alternatives, illustrating that accuracy alone is an incomplete basis for comparison in this setting.

4.2 Pruning, Quantization, and Compact Deep Architectures

A second thread applies model compression — pruning, post-training or dynamic quantization, and

architecture-level compaction — directly to deep models. Reported hybrid DNN-BiLSTM architectures combined with dynamic quantization and incremental PCA have achieved model sizes near 25–31 kilobytes while retaining accuracy above 93–99% depending on the benchmark (CIC-IDS2017 and CICIoT2023 respectively), and hybrid CNN-BiLSTM architectures explicitly designed for on-board IoT deployment have reported 97.28% binary and 96.91% multiclass accuracy on UNSW-NB15. Separately, pipelines combining pruning and quantization with generative oversampling (CTGAN) to address class imbalance have reported comparable accuracy to unpruned baselines while reducing training time by up to 94% and memory consumption by up to 90%. TinyML-optimized feedforward networks evaluated on ToN_IoT have reported F1-scores near 0.976 at 31 kilobyte model size with inference throughput near 120,000 packets per second, presented as a favorable balance point relative to larger classical baselines.

4.3 Spiking Neural Networks and Energy-Centric Approaches

A smaller thread targets energy consumption directly rather than treating it as a secondary consequence of model size. Spiking neural network formulations evaluated against conventional CNN baselines on NSL-KDD have reported up to 4.8× lower energy consumption at a moderate accuracy cost (approximately 75% accuracy and 72% F1 versus higher-accuracy CNN baselines), framing an explicit accuracy-versus-energy trade-off rather than optimizing accuracy alone.

4.4 TinyML and On-Device Deployment

TinyML-specific work extends compression techniques with deployment validation on constrained hardware and multi-dataset evaluation. Recent work applying feature-importance-based selection, quantization, pruning, and sparsification across tree-based and neural models has reported maintained accuracy across four benchmark datasets (CICIoT2023, Edge-IIoTset, CIC-IDS2017, and RT-IoT2022) after optimization, and a 2026 framework combining TinyML with graph attention networks, federated training, and explainability reported up to 95% accuracy across DDoS, advanced-persistent-threat, and zero-day attack scenarios evaluated on Edge-IIoTset, CICIoT2023, and RT-IoT2022, alongside a 20% reduction in federated communication overhead in a ten-node simulation.

5. Federated and Privacy-Preserving Approaches

A distinct and rapidly growing thread applies federated learning (FL) to IoT intrusion detection, motivated by the observation that centralizing raw traffic data from distributed IoT deployments raises privacy and bandwidth concerns. Systematic reviews of FL-based IDS have catalogued a broad range of architectures, and IoT-specific surveys have specifically examined privacy-preserving FL for intrusion detection. Reported IoT-focused FL-IDS work spans lightweight neural architectures combined with differential privacy and homomorphic encryption for real-time detection, mist- and edge-assisted FL architectures for heterogeneous IoT networks, and dynamic feature-fusion FL approaches targeted at vehicular network security. A consistent theme across this sub-literature is a distinction between two threat layers: external network intrusions, which FL-based IDS aim to detect, and internal model-layer attacks such as poisoning, backdoor insertion, and gradient leakage, which target the federated training process itself and are treated as a separate, compounding security concern. Non-IID data distribution across clients and class imbalance for rare attack types are repeatedly identified as open challenges specific to the federated setting, with proposed mitigations including cross-client aggregation adjustments and synthetic data generation for underrepresented classes.

6. Explainability in IoT Intrusion Detection

As lightweight and federated IDS models have matured, a parallel thread has emphasized explainability, motivated by the operational need for analysts to trust and act on model outputs rather than treat them as opaque alerts. Reported approaches include post-hoc explanation methods (e.g., LIME and SHAP) applied to multi-layer perceptron IDS models, attention-based interpretability built into graph neural network architectures for edge intrusion detection, and logic-based explainable frameworks proposed for adjacent network domains such as 5G. Explainability is increasingly reported as a module integrated alongside lightweight and federated components rather than as a standalone contribution, reflecting a trend toward multi-objective IDS designs

that jointly target accuracy, efficiency, privacy, and interpretability.

7. Comparative Synthesis

Table 1 summarizes representative reported results across the studies reviewed in Sections 4–6, organized by technique, evaluation dataset, headline detection result, and the resource-cost dimension emphasized by each study. Figures are as reported by the original authors under their own experimental conditions; because evaluation datasets, preprocessing, and hardware platforms differ across studies, these figures should be read as illustrative of technique-level trends rather than as a directly comparable leaderboard.

Approach	Technique	Dataset	Reported Result	Model Size / Cost
PCA+PSO+SVM	Feature selection (41→18 features)	NSL-KDD	98.5% accuracy	Reduced feature set; low compute
SNN vs. CNN	Spiking neural network	NSL-KDD	75% acc / 72% F1	4.8× less energy than CNN
BCIDS-IoT (ANN)	Binary classification ensemble	UNSW-NB15	95% accuracy	Low false-positive rate
LIDSuFNN / LIDSuCNN	Pruning + quantization + CTGAN	Multiple benchmarks	Comparable accuracy to baseline	−94% training time, −90% memory
DNN-BiLSTM (quantized)	Dynamic quantization + IPCA	CIC-IDS2017	99.73% accuracy	25.6 KB model
DNN-BiLSTM (quantized)	Dynamic quantization + IPCA	CICIoT2023	93.95% accuracy	31.3 KB model
CNN-BiLSTM	Hybrid lightweight deep model	UNSW-NB15	97.28% (binary) / 96.91% (multiclass)	Designed for on-board IoT deployment
Decision Tree	Classical baseline	ToN_IoT	F1 = 0.987	593 KB
TinyML-FNN	Compressed feedforward NN	ToN_IoT	F1 = 0.976	31 KB; ≈120,000 pkt/s throughput
Decision Tree	Classical baseline	CICIoT2023	99.56% acc / 99.62% F1	—
XGBoost	Feature-selected ensemble	Real-world traffic	89.09% accuracy	Optimized for constrained deployment
TinyML + GAT + FL + XAI	Federated TinyML w/ explainability	Edge-IIoTset, CICIoT2023, RT-IoT2022	Up to 95% accuracy	20% lower FL communication overhead

Table 1. Representative reported results across lightweight IoT-IDS techniques (figures as reported in cited original studies; not independently reproduced).

Several patterns emerge from this synthesis. First, high headline accuracy (often above 95%) is now common across nearly all technique families, suggesting that detection accuracy in isolation is no longer a strongly differentiating metric for this literature. Second, resource-cost reporting is inconsistent: some studies report model size only, others report inference throughput or energy directly, and few report more than two of {size, latency, energy, communication cost} simultaneously, which limits direct comparison across studies. Third, almost all reviewed studies evaluate on a single dataset; cross-dataset evaluation, where a model trained on one benchmark is tested on an independently collected benchmark without retraining, remains rare despite explicit acknowledgement in dataset-originating papers (e.g., CICIoT2023, ToN_IoT heterogeneity analyses) that this is an important open question.

8. Open Challenges and Future Directions

- Cross-dataset generalization: Systematic evaluation of whether models trained on one IoT benchmark retain performance on independently collected benchmarks remains uncommon and is a priority gap given documented heterogeneity across datasets.
- Standardized resource-cost reporting: The field would benefit from a common reporting template covering model size, inference latency, energy per inference, and (for federated approaches) communication cost, to enable fair cross-study comparison.
- Rare-class performance: Aggregate accuracy and even macro-F1 can mask poor recall on rare but high-impact attack types (e.g., MITM, ransomware); per-class reporting should be standard practice rather than the exception.
- Real-hardware validation: A substantial share of reviewed work validates resource efficiency analytically or in simulation rather than on physical constrained hardware; broader adoption of on-device benchmarking (e.g., on microcontroller- or single-board-computer-class platforms) would strengthen deployment claims.
- Security of the learning process itself: As federated and distributed training approaches grow, model-layer threats (poisoning, backdoors, gradient leakage) require treatment as a first-class concern alongside network-layer intrusion detection, not an afterthought.
- Explainability at the edge: Most explainability techniques (SHAP, LIME, attention mechanisms) were not designed with the same resource constraints as the detection models they explain; lightweight, on-device-compatible explanation methods remain underexplored.

CONCLUSION

Lightweight intrusion detection for IoT has matured rapidly, with feature-reduction, pruning and quantization, spiking neural networks, TinyML deployment, federated training, and explainability each contributing an active and productive sub-literature. Reported detection accuracy across these techniques is now consistently high, which shifts the practically important open questions toward generalization across deployment environments, consistent and comprehensive resource-cost reporting, robustness for rare attack classes, and the security of the training process itself in federated settings. This review has synthesized representative results across these threads and identified cross-dataset generalization and standardized efficiency reporting as the most immediately actionable gaps for future empirical work, including the companion experimental methodology the authors intend to pursue in follow-on work using ToN_IoT and CICIoT2023 as a harmonized generalization testbed.

REFERENCES

1. N. Koroniotis, N. Moustafa, E. Sitnikova, and B. Turnbull, "Towards the development of realistic botnet dataset in the Internet of Things for network forensic analytics: Bot-IoT dataset," Future Generation Computer Systems, vol. 100, pp. 779–796, 2019.
2. A. Alsaedi, N. Moustafa, Z. Tari, A. Mahmood, and A. Anwar, "TON_IoT telemetry dataset: A new generation dataset of IoT and IIoT for data-

driven intrusion detection systems," IEEE Access, vol. 8, pp. 165130–165150, 2020.
3. N. Moustafa, M. Keshk, E. Debie, and H. Janicke, "Federated TON_IoT Windows datasets for evaluating AI-based security applications," in Proc. IEEE TrustCom, 2020, pp. 848–855.
4. E. C. P. Neto, S. Dadkhah, R. Ferreira, A. Zohourian, R. Lu, and A. A. Ghorbani, "CICIoT2023: A real-time dataset and benchmark for large-scale attacks in IoT environment," Sensors, 2023.
5. M. Jouhari and M. Guizani, "Lightweight CNN-BiLSTM based intrusion detection systems for resource-constrained IoT devices," arXiv:2406.02768, 2024.
6. Anonymous, "Lightweight intrusion detection system for IoT with improved feature engineering and advanced dynamic quantization," Discover Internet of Things (Springer Nature), 2025.
7. Anonymous, "A lightweight intrusion detection method for IoT based on deep learning and dynamic quantization," PMC/NIH, 2023.
8. Anonymous, "Lightweight machine learning models for intrusion detection on IoT devices," NTNU NIKT proceedings, 2025.
9. Anonymous, "Smart IoT security: Lightweight machine learning techniques for multi-class attack detection in IoT networks," arXiv:2502.04057, 2025.
10. Anonymous, "Intrusion detection on resource-constrained IoT devices: Hardware-aware model optimization," arXiv:2512.02272, 2025.
11. Anonymous, "Lightweight intrusion detection system using a hybrid CNN and ConvNeXt-Tiny model for Internet of Things networks," arXiv:2509.06202, 2025 (surveying Raghunath et al. 2025 [PCA+PSO+SVM], Shafique et al. 2025 [spiking neural networks], Agarwal et al. 2025 [BCIDS-IoT], and Das et al. 2025 [LIDSuFNN/LIDSuCNN with CTGAN]).
12. L. Nassef, M. I. Alghamdi, S. Ben Chaabane, Q. Abbas, W. M. Alawad, O. H. Albalawi, O. I. Alqaisi, and B. Fakieh, "Lightweight and energy-aware intrusion detection for industrial IoT using TinyML and edge AI," Scientific Reports, vol. 16, art. 18524, 2026.
13. S. El Haddouti and W. Lazraq, "TinyML strategies for privacy-preserving and cyber threat multi-classification in edge-IoT networks," Computing, vol. 107, art. 173, 2025.
14. J. L. Hernández Ramos, R. M. Parizi, A. Dehghantanha, S. Pouriyeh, V. Mothukuri, and K.-K. R. Choo et al., "Intrusion detection based on federated learning: A systematic review," ACM Computing Surveys, vol. 57, no. 12, pp. 1–65, 2025.
15. Anonymous, "A comprehensive survey of federated learning for intrusion detection systems in Internet of Things environments," ScienceDirect (in press), 2026.
16. A. Vyas et al., "Privacy-preserving federated learning for intrusion detection in IoT environments: A survey," IEEE Access, 2024.
17. A. Puviarasu and V. K. Sudha, "Enhanced IoT security: Privacy-preserving federated learning model for accurate, real-time intrusion detection across devices," Ain Shams Engineering Journal, vol. 17, no. 1, art. 103866, 2026.
18. A. Khraisat, A. Alazab, and M. Alazab et al., "Federated learning for intrusion detection in IoT environments: A privacy-preserving strategy," Discover Internet of Things, vol. 5, art. 72, 2025.
19. J. Li, Y. Ma, J. Bai, C. Chen, T. Xu, and C. Ding, "A lightweight intrusion detection system with dynamic feature fusion federated learning for vehicular network security," Sensors, vol. 25, no. 15, art. 4622, 2025.
20. Anonymous survey (MDPI Sensors), "Intrusion detection in the Internet of Things: A comprehensive review of techniques, architectures, datasets, and emerging trends," Sensors, vol. 26, no. 11, art. 3405, 2026.
21. D. Gaspar, P. Silva, and C. Silva, "Explainable AI for intrusion detection systems: LIME and SHAP applicability on multi-layer perceptron," IEEE Access, 2024.
22. Anonymous, "ExAI5G: A logic-based explainable AI framework for intrusion detection in 5G networks," arXiv:2604.18052, 2026.

HOW TO CITE: Uttish Bavoriya*, Lightweight And Privacy-Preserving Intrusion Detection For The Internet Of Things: A Systematic Review Of Datasets, Models, And Emerging Directions, Int. J. Sci. R. Tech., 2026, 3 (9), 484-489. https://doi.org/10.5281/zenodo.22916752