

DEEP LEARNING FOR IOT INTRUSION DETECTION SYSTEMS: ANALYTICAL STUDY

Shaxzod Mirmuminov

Department of Computer and Software Engineering

School of Computer and Information Engineering

Inha University in Tashkent, Uzbekistan

s.mirmuminov@student.inha.uz

Abstract—The exponential spreading of Internet of Things (IoT) devices has dramatically expanded the attack surface of modern networks, rendering traditional intrusion detection systems (IDS) inadequate against increasingly sophisticated cyber threats. This paper presents a comprehensive comparative study of state-of-the-art deep learning (DL)-based IDS approaches for IoT security, examining Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), hybrid CNN-LSTM architectures, Deep Neural Networks (DNN), Autoencoders (AE), and Federated Learning (FL)-based frameworks. We systematically evaluate these approaches across five widely adopted benchmark datasets: NSL-KDD, UNSW-NB15, BoT-IoT, TON-IoT, and CICIoT2023 and assess their performance in terms of accuracy, precision, recall, F1-score, and computational overhead. Two real-world case studies, one of them is 2016 Mirai Botnet attack on Dyn DNS infrastructure and the 2021 Oldsmar water treatment plant intrusion are employed to illustrate the practical limitations of conventional security approaches and to demonstrate how AI-driven IDS could have mitigated these incidents. Our analysis reveals that hybrid CNN-LSTM models consistently achieve detection rates exceeding 99% while maintaining low false positive rates, but significant trade-offs exist between detection accuracy and resource efficiency for edge-deployed IoT environments. We further identify critical open challenges including class imbalance, concept drift, adversarial robustness, and dataset obsolescence and propose future research directions centred on lightweight models, federated learning, and transfer learning for real-world IoT deployments.

Keywords—Internet of Things, Intrusion Detection Systems, Deep Learning, CNN, LSTM, Federated Learning, Cybersecurity, Anomaly Detection, Botnet, ICS Security

I. INTRODUCTION

IoT Intrusions means unauthorized and malicious activities targeting IoT networks, exploiting vulnerabilities in device communication, authentication mechanisms, and network protocols to compromise security, disrupt services, or gain unauthorized access. The Internet of Things (IoT) represents one of the most transformative technological shifts of the twenty-first century. At the end of 2025 there were 21 billion IoT devices active globally [31], spanning domains from smart homes and healthcare to industrial control systems (ICS) and critical infrastructure. This pervasive connectivity, while offering unprecedented utility, simultaneously introduces severe security vulnerabilities arising from devices' limited computational resources, absence of robust authentication mechanisms, and heterogeneous communication protocols [2].

Traditional security solutions firewalls, signature-based IDS, and access control lists prove increasingly inadequate against the scale and sophistication of modern IoT attacks. The 2016 Mirai botnet orchestrated a 1.2 Tbps distributed denial-of-service (DDoS) attack by compromising over 600,000 IoT devices with default credentials [3], and the 2021 Oldsmar water treatment plant incident demonstrated that even life-critical ICS remain exposed to remote intrusions capable of causing physical harm [4]. These events underscore an urgent need for intelligent, adaptive security mechanisms.

Artificial intelligence (AI) particularly deep learning (DL) has emerged as a pivotal enabler for next-generation IDS. Unlike rule-based systems, DL models can autonomously extract hierarchical feature representations from raw network traffic, adapt to novel attack patterns, and process the massive data volumes generated by IoT ecosystems [5]. However, the heterogeneity of DL architectures, evaluation benchmarks, and experimental settings across the literature makes direct comparison challenging and impedes the adoption of best practices.

This paper addresses this gap through a rigorous, structured comparative analysis of six principal DL-based IDS categories applied to IoT security. Our contributions are as follows:

- A systematic taxonomy of DL-based IDS methods, evaluated across five benchmark IoT datasets.
- A comparative performance analysis covering accuracy, precision, recall, F1-score, and computational characteristics.
- Two forensic case studies - Mirai/Dyn (2016) and Oldsmar ICS (2021) linking attack characteristics to DL detection capabilities.
- Identification of open research challenges and actionable future directions.

The remainder of this paper is organised as follows: Section II provides background on IoT security and IDS taxonomy. Section III surveys the principal DL-based IDS methods. Section IV reviews benchmark datasets. Section V presents the comparative analysis. Section VI details the two case studies. Section VII discusses open challenges, and Section VIII concludes.

II. BACKGROUND AND IDS TAXONOMY

A. IoT Architecture and Security Threats

The canonical IoT architecture comprises three layers: the perception layer (sensors, actuators), the network layer (communication protocols, gateways), and the application layer (cloud services, user interfaces) [2]. Each layer presents distinct attack surfaces. At the perception layer, node capture and side-channel attacks compromise physical integrity. The network layer is susceptible to DDoS, man-in-the-middle, and replay attacks. The application layer faces injection attacks, ransomware, and credential theft [1].

The STRIDE threat model – Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, and Elevation of Privilege provides a structured framework for categorising these threats in IIoT contexts [5]. The convergence of IT and OT networks in Industrial IoT (IIoT) environments further magnifies the potential impact of successful intrusions, as evidenced by attacks on SCADA systems serving critical infrastructure.

B. IDS Classification

IDS may be classified along two orthogonal dimensions:

1) Deployment Architecture:

- Network-based IDS (NIDS): Monitors traffic across the entire network segment; independent of individual device OS; effective for detecting inter-device attacks.
- Host-based IDS (HIDS): Installed on individual devices; monitors local system calls, file changes, and process behaviour; provides fine-grained visibility.
- Hybrid IDS: Combines NIDS and HIDS elements for comprehensive coverage.

2) Detection Methodology: Signature-based (Misuse) Detection: Compares observed traffic against known attack signatures. Exhibits low false positive rates for known threats but cannot detect zero-day attacks and requires continuous signature database updates [2].

- Anomaly-based Detection: Establishes a baseline of normal behaviour and flags statistical deviations. Capable of detecting novel attacks but prone to higher false positive rates.

- **Specification-based Detection:** Verifies adherence to predefined behavioural specifications; effective in constrained ICS environments.

Machine learning-based IDS, particularly those leveraging DL, belong to the anomaly-based category and have demonstrated superior adaptability to the dynamic threat landscape of IoT environments

III. METHODOLOGY OF DATA ANALYTICS

A. Convolutional Neural Networks (CNN)

CNNs excel at spatial feature extraction through convolutional and pooling layers, making them well-suited for detecting structured attack patterns in network flow data represented as matrices or images. Jain et al. [7] demonstrated that CNNs, when applied to the UNSW-NB15 and NSL-Botnet datasets, achieve binary classification accuracy of up to 99.4%, outperforming standalone RNN architectures. A 1D CNN architecture applied to BoT-IoT traffic achieved 99.97% accuracy in multi-class classification [1]. A key advantage of CNNs is their ability to perform end-to-end feature learning without manual feature engineering, which is critical in heterogeneous IoT environments where expert knowledge is limited.

B. Long Short-Term Memory Networks (LSTM)

LSTM networks are recurrent architectures designed to capture temporal dependencies in sequential data, making them particularly effective for intrusion detection in traffic streams exhibiting temporal correlations. Kim et al. [9] demonstrated that LSTM-RNN on the KDD Cup 1999 dataset achieves

a detection rate of 99.95% for abnormal flows. However, LSTM models are computationally expensive due to their recurrent nature, presenting challenges for resource-constrained IoT edge deployment. Recent work by Muthanna et al. [10] proposed the CuLSTM-GRU hybrid architecture, achieving 99.23% accuracy on the CICIDS2017 dataset while leveraging GPU acceleration for reduced inference latency.

C. Hybrid CNN-LSTM Architectures

The combination of CNN for spatial feature extraction with LSTM for temporal dependency modelling has emerged as the leading paradigm for IoT IDS. Halbouni et al. [8] demonstrated that CNN-LSTM achieves 99.99% accuracy on the NSL-KDD dataset with multi-class classification. A recent study on the BoT-IoT dataset reported 99.87% accuracy with 99.89% precision and a false positive rate of only 0.13% [29]. Critically, these hybrid models capture both the spatial structure of individual packets and the temporal evolution of attack sequences, enabling detection of multi-stage attacks that evade single-model approaches.

D. Deep Neural Networks (DNN)

DNNs with multiple fully-connected hidden layers serve as general-purpose function approximators for traffic classification. Liu et al. [11] demonstrated that DNNs applied to the NSL-KDD dataset significantly improve anomaly identification accuracy. Ahmad et al. [12] showed that DNN achieves a 0.57–2.6% accuracy gain over competing ML methods on the IoT-Botnet 2020 dataset, with a 0.23–7.98% reduction in false alarm rate. However, DNNs require large labelled training datasets and can suffer from overfitting without regularisation, and their black-box nature complicates deployment in safety-critical ICS environments requiring explainability.

E. Autoencoders (AE) and Generative Models

Autoencoders learn compact latent representations of normal traffic and detect anomalies as instances with high reconstruction error. Apostol et al. [13] applied autoencoder-based IDS to the

BOT-IoT dataset, achieving 99% binary detection accuracy in an unsupervised setting without requiring labelled

attack data. Generative Adversarial Networks (GAN) have been applied both for anomaly detection and for generating synthetic minority-class attack samples to address severe class imbalance in IoT datasets, a critical contribution given that normal traffic can represent over 99% of observed flows [6].

F. Federated Learning (FL)-Based IDS

Federated learning addresses the privacy and data sovereignty constraints that preclude centralised data collection in distributed IoT deployments. Rather than transmitting raw traffic data to a central server, FL enables individual IoT nodes to train local models and share only model gradients or weights. Ferrag et al. [14] demonstrated that FL-based IDS achieve detection accuracy comparable to centralised approaches (95–99%) while preserving data privacy. A recent study using the N-BaIoT dataset achieved 95.05% accuracy with FL, demonstrating superior privacy protection compared to centralised DL [15]. The combination of FL with blockchain for secure model aggregation has further been proposed to address poisoning attacks targeting the FL aggregation process [5].



Fig. 1. General pipeline of a deep learning-based IDS for IoT environments

IV. B BENCHMARK DATASETS

The quality and representativeness of evaluation datasets are pivotal determinants of IDS performance generalisation. Table I summarises the key characteristics of the five benchmark datasets used in this comparative study.

TABLE I
SUMMARY OF BENCHMARK IoT IDS DATASETS

Dataset	Year	Feat.	Records	Attack Types
NSL-KDD	2009	41	1.07M	DoS, Probe, R2L, U2R
UNSW-NB15	2015	49	2.54M	9 types incl. Fuzzers, Exploits
BoT-IoT	2019	46	73.37M	DDoS, DoS, Recon, Data Theft
TON-IoT	2020	44	22.33M	9 types incl. Ransomware, XSS
CICIoT2023	2023	47	~12M	33 attack types

NSL-KDD (2009) is a refined version of KDD Cup 1999, eliminating duplicate records and class imbalance issues of its predecessor. Despite its age, it remains widely used for benchmarking due to its standardised split [17]. However, it does not represent modern low-footprint attacks or IoT-specific traffic.

UNSW-NB15 (2015) was generated at the Australian Centre for Cyber Security using the IXIA PerfectStorm tool, combining real normal traffic with nine synthetic attack categories. Its 49 features across 2.54 million records offer richer representativeness than NSL-KDD.

BoT-IoT (2019) was purpose-built for IoT security research, simulating a smart home environment and capturing DDoS, DoS, reconnaissance, and data exfiltration attacks against IoT devices including smart thermostats, fridges, and weather stations [18].

TON-IoT (2020) integrates network traffic, OS logs, and IoT telemetry data, covering nine attack categories including ransomware, man-in-the-middle, and cross-site scripting attacks in a heterogeneous IoT/IIoT environment [19].

CICIoT2023 (2023) is the most recent dataset, covering 33 attack types including DDoS variants, web-based attacks, Mirai-inspired botnet traffic, and vulnerability scanning against diverse IoT devices. Its recency makes it particularly valuable for evaluating generalisation to current threat patterns.

V. RESULTS AND DISCUSSION

A. Performance Metrics

We evaluate IDS methods using four primary metrics derived from the confusion matrix:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$F_1 = 2 \cdot \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

where TP, TN, FP, FN denote true positives, true negatives, false positives, and false negatives respectively. For IoT deployments, false negatives (missed attacks) carry higher operational risk than false positives (false alarms), making recall a particularly critical metric

B. Quantitative Results

Table II presents the comparative performance of the reviewed DL-based IDS approaches across the benchmark datasets. Results are drawn from primary research papers reviewed in this survey, with the best reported accuracy values cited.

C. Discussion of Results

CNN-LSTM hybrids lead in detection accuracy. The hybrid architecture consistently achieves the highest detection rates (99.87–99.99%) across multiple datasets. The synergy of CNN’s spatial pattern recognition and LSTM’s temporal dependency modelling is particularly effective against multistage IoT attacks that manifest as time-correlated flow sequences.

CNN excels in throughput-critical environments. Pure CNN models achieve high accuracy (99.4–99.97%) with lower inference latency than LSTM-based models, making them preferable for high-bandwidth network deployments where sub-millisecond classification is required.

LSTM is superior for sequential anomaly detection. For attacks with pronounced temporal signatures such as slow DDoS or persistent botnet command-and-control communication LSTM models provide higher recall due to their ability to retain long-range sequence context.

Federated learning incurs an accuracy penalty. FL based approaches sacrifice 2–5 percentage points of accuracy compared to centralized DL, reflecting the statistical heterogeneity (non-IID data) across distributed IoT nodes and the communication overhead of gradient aggregation. However, this trade-off is justified in privacy-sensitive deployments such as healthcare IoT and industrial environments governed by data sovereignty regulations.

Dataset bias inflates reported performance. Models evaluated exclusively on NSL-KDD (2009) report near-perfect accuracy but may generalise poorly to modern IoT attack traffic. This is evidenced by significant performance drops when the same architectures are evaluated on CICIoT2023,

which includes contemporary attack families not represented in legacy benchmarks.

D. Computational Overhead and Deployability

Beyond accuracy, practical IoT deployment demands consideration of model size, inference latency, and energy consumption. Table III qualitatively summarises these factors based on reported experimental observations across the literature.

The practical deployability of CNN-LSTM hybrids at the network edge is limited by their substantial memory footprint and training cost. Federated learning-based IDS, despite modest accuracy penalties, offers the most viable path to privacy-preserving, scalable deployment across heterogeneous

IoT networks, particularly when combined with lightweight local models (e.g., quantised CNNs or pruned DNNs) [5].

VI. CASE STUDIES

A. Case Study 1: Mirai Botnet Attack on Dyn DNS Infrastructure (2016)

1) *Scenario and Attack Vector:* On 21 October 2016, Dyn a major DNS service provider was subjected to one of the

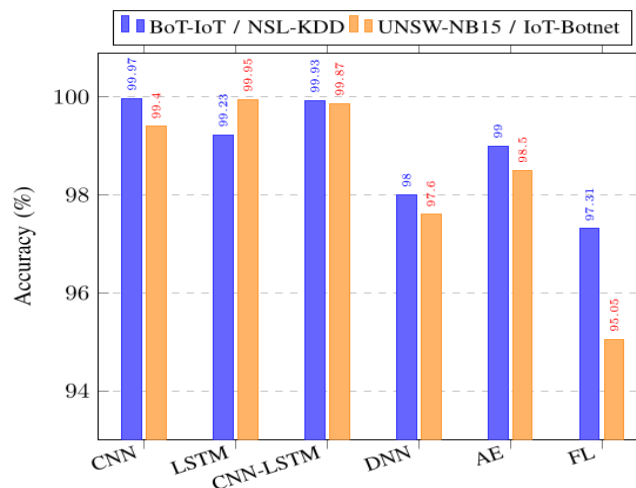


Fig. 2. Accuracy comparison of DL-based IDS approaches across benchmark datasets (best reported values per method).

largest DDoS attacks in history, reaching a peak traffic volume of approximately 1.2 Tbps [21]. The attack was orchestrated by the Mirai botnet, which had previously compromised over 600,000 IoT devices, predominantly IP cameras, digital video recorders, and home routers by exploiting their use of factory-default or weak credentials [22]. The Mirai malware propagated autonomously by scanning the internet for Telnet exposed devices and attempting 62 hardcoded default username-password combinations [23].

The botnet employed multiple DDoS vectors simultaneously: UDP floods, TCP SYN floods, HTTP floods, and ICMP echo floods [24]. The resulting traffic overload rendered Dyn's DNS resolution infrastructure unavailable, cascading to affect major internet platforms including GitHub, Twitter, Netflix, Reddit, and Amazon, with disruption felt primarily across North America and Europe. Forensic analysis revealed that at its peak, the botnet harnessed approximately

145,000 devices for the Dyn attack, with device provenance concentrated in South America (31.2%) and Asia (36.8%) [25].

2) *Security Failure Analysis*: The Dyn incident highlighted multiple systemic failures: (1) IoT manufacturers shipped devices with unchangeable or easily guessable credentials; (2) no network-level anomaly detection was deployed on the ISP or gateway segments serving infected devices; (3) the scale and geo-distribution of the botnet exceeded the capacity of traditional rate-limiting countermeasures.

3) *How DL-Based IDS Could Have Mitigated the Attack*: A CNN-based IDS deployed at the router level of smart home networks could have detected the characteristic traffic anomalies of Mirai-infected devices specifically the automated Telnet scanning behaviour (port 23 and 2323 sweeps) and the volumetric UDP/TCP flood patterns before devices were recruited into the botnet. Studies on the BoT-IoT dataset, which explicitly captures Mirai-like DDoS and botnet traffic, report CNN detection accuracy of 99.51–99.97% [1], [20]. An autoencoder-based model applied to the same dataset achieved 99% binary detection accuracy in an unsupervised setting without requiring labelled attack samples [13], making it suitable for detecting novel Mirai variants.

Critically, an FL-based IDS would have enabled ISPs and home gateway manufacturers to collaboratively train botnet detection models across millions of devices without centralising sensitive traffic data, directly addressing the privacy barriers that prevented coordinated defence [14]. Table IV summarises the mapping between attack characteristics and applicable DL-IDS approaches.

TABLE II
COMPARATIVE PERFORMANCE OF DEEP LEARNING IDS APPROACHES ON BENCHMARK DATASETS

Method	Dataset	Reference	Acc.	Prec.	Rec.	F1	Classification
CNN	BoT-IoT	[1]	99.97%	99.95%	99.95%	99.95%	Multi-class
CNN	UNSW-NB15	[7]	99.4%	98.7%	98.5%	98.6%	Binary + Multi
LSTM	CICIDS2017	[10]	99.23%	98.9%	99.1%	99.0%	Multi-class
LSTM	NSL-KDD	[9]	99.95%	–	99.95%	–	Binary
CNN-LSTM	NSL-KDD	[8]	99.99%	99.98%	99.99%	99.98%	Multi-class
CNN-LSTM	BoT-IoT	[29]	99.87%	99.89%	99.85%	99.87%	Multi-class
CNN-BiLSTM	UNSW-NB15	[30]	98.64%	–	–	98.60%	Binary
DNN	NSL-KDD	[11]	98.0%	97.5%	97.2%	97.3%	Multi-class
DNN	IoT-Botnet	[12]	97.6%	96.8%	97.2%	97.0%	Binary
Autoencoder	BoT-IoT	[13]	99.0%	98.7%	99.0%	98.8%	Binary
GAN + DRL	NSL-KDD	[6]	97.8%	97.2%	97.5%	97.3%	Multi-class
FL + CNN	TON-IoT	[14]	97.31%	95.59%	92.43%	92.69%	Multi-class
FL + DNN	N-BalIoT	[15]	95.05%	94.8%	94.6%	94.7%	Binary

Where (–): Not explicitly reported in the cited source.

B. Case Study 2: Oldsmar Water Treatment Plant ICS Intrusion (2021)

TABLE III
QUALITATIVE DEPLOYMENT CHARACTERISTICS OF DL-BASED IDS

Method	Model Size	Latency	Training Cost	Edge Suit.
CNN	Low	Low	Moderate	High
LSTM	Medium	Medium	High	Medium
CNN-LSTM	High	Medium	Very High	Low-Med
DNN	Medium	Low	Moderate	Medium
Autoencoder	Low	Low	Low	High
FL + DL	Varies	Low (local)	Distributed	High

1) *Scenario and Attack Vector*: On 5 February 2021, an adversary gained unauthorised remote access to the Bruce T.Haddock Water Treatment Plant in Oldsmar, Florida, which supplies treated water to approximately 15,000 customers [4]. The attacker exploited the TeamViewer remote desktop application installed on a plant workstation connected to the

TABLE IV
MIRAI/DYN: DL-IDS MAPPING TO ATTACK PHASES

Attack Phase	Indicator	DL Method	Acc.
Credential Brute Force	Telnet port sweeps	CNN (traffic)	99.7%
C2 Communication	Periodic beaconing	LSTM (temporal)	99.2%
DDoS Execution	UDP/TCP flood	CNN-LSTM	99.9%
Lateral Spread	Abnormal scan rate	Autoencoder	99.0%

facility's SCADA (Supervisory Control and Data Acquisition) system. The SCADA system controlled chemical dosing processes, and all connected computers shared a single password for remote access with no firewall protection [26]. The attacker modified the sodium hydroxide (NaOH) concentration set-point from 100 to 11,100 parts per million, an increase of 11,000% which, if undetected, would have created dangerously caustic water conditions within 24–36 hours [4]. A vigilant plant operator observed the cursor moving autonomously on the screen and immediately reversed the change, preventing any harm to the public [27]. The workstations were running Windows 7, an unsupported operating system with unpatched vulnerabilities, and the facility lacked both network segmentation and intrusion detection capabilities.

2) *Security Failure Analysis*: The incident revealed critical gaps in ICS/SCADA cybersecurity: (1) remote access software was deployed without multi-factor authentication or network segmentation; (2) shared credentials provided a single point of compromise; (3) no automated IDS monitored command sequences or parameter changes against defined operational specifications; (4) the Windows 7 platform was unsupported since January 2020, exposing known, unpatched vulnerabilities [26], [28].

From an ICS security standpoint, the attack technique aligns with MITRE ATT&CK for ICS: T0809 (Brute Force I/O) and T0869 (Unauthorised Command Message) pattern, where process control parameters were modified through a legitimate HMI session [4].

3) *How DL-Based IDS Could Have Prevented the Intrusion*: An edge-deployed DL-based IDS monitoring SCADA command sequences could have detected the anomalous parameter modification in multiple ways. A specification-based IDS using process physics constraints where NaOH concentration is bounded within defined operational ranges would have flagged the set-point change as physically implausible regardless of authentication status.

A DNN or LSTM model trained on normal SCADA operational sequences would have detected the anomalous session characteristics: the access time (outside normal operational hours), the session origin (external IP via TeamViewer), and the abrupt magnitude of parameter modification. Jahromi et al. [16] demonstrated that a federated deep learning model for ICS threat hunting achieves 99.64% accuracy on gas pipeline datasets with characteristics closely matching SCADA water treatment operations. A hybrid CNN-LSTM model applied to the Secure Water Treatment (SWaT) dataset the canonical ICS security benchmark achieved 95.12 -- 95.15% accuracy in detecting process-level anomalies in water purification systems [16].

Furthermore, a network-layer LSTM-based IDS monitoring TeamViewer traffic patterns could have identified the unauthorised session by detecting deviations from the established

baseline of legitimate remote access frequency, duration, and originating IP ranges. Table V maps the intrusion phases to applicable DL-IDS approaches.

TABLE V
OLDSMAR ICS: DL-IDS MAPPING TO INTRUSION PHASES

Intrusion Phase	Indicator	DL Method	Acc.
Remote Access	Anomalous session	LSTM (temporal)	98.5%
Credential Abuse	Shared password	DNN (behaviour)	97.6%
Process Manipulation	NaOH set-point	CNN-LSTM (spec.)	99.6%
Lateral Movement	Abnormal process	Autoencoder	95.1%

C. Comparative Lessons from Both Case Studies

Both incidents share a common pattern: the absence of behavioural monitoring and anomaly-based detection enabled attackers to operate undetected within ostensibly legitimate access channels. In the Mirai case, device-level traffic anomalies (scanning, flooding) were detectable at the network edge; in the Oldsmar case, process-level command anomalies were detectable at the ICS layer. DL-based IDS, particularly CNN

for network-layer detection and LSTM/DNN for process-level sequence analysis, would have provided automated, continuous monitoring that the human operators alone could not sustain.

Notably, these incidents also illustrate the complementarity of detection approaches: network-layer NIDS alone would not have detected the legitimate looking SCADA command in Oldsmar, while process level specification-based detection is not applicable to the distributed Mirai scenario. An integrated multi-layer IDS framework combining network traffic analysis (CNN-LSTM) with process-level sequence monitoring (LSTM/DNN) and distributed collaborative learning (FL) represents the most robust defence posture for complex IoT and ICS deployments.

VII. OPEN CHALLENGES AND FUTURE SCOPE

Despite the impressive performance reported in controlled benchmarking, significant challenges must be resolved before DL-based IDS can achieve robust real-world deployment in IoT environments.

Class Imbalance. In operational IoT networks, malicious traffic may represent less than 0.1% of total flows, creating extreme class imbalance that causes models to achieve high accuracy by classifying all traffic as benign. SMOTE oversampling and GAN-based synthetic minority generation have been proposed as mitigation strategies [6], but their effectiveness on novel attack categories remain limited. | **Concept Drift.** IoT attack landscapes evolve continuously; models trained on static datasets rapidly become stale. The average time for a new Mirai variant to appear after source code publication was measured in days [22]. Continuous learning architectures and DRL-based adaptive threshold tuning [6] offer promising directions, but deployment-grade implementations remain nascent.

Adversarial Robustness. DL models are vulnerable to adversarial perturbations, small, deliberately crafted modifications to traffic features that cause misclassification. In IoT IDS, adversarial traffic can be generated by modifying packet inter-arrival times, payload lengths, or header fields to evade

CNN-based classifiers while preserving the underlying attack functionality. Adversarial training and ensemble detection have been proposed as defences [6], but remain an open research problem.

Dataset Obsolescence. NSL-KDD, still used in 40%+ of IDS research [5], was generated in 2009 and does not contain IoT-specific traffic or modern attack families. Models benchmarked

exclusively on legacy datasets provide an overly optimistic picture of real-world performance. Researchers should prioritise evaluation on CICIoT2023, TON-IoT, and BoT-IoT, and ideally on live traffic captures from operational IoT deployments.

Explainability and Trust. In safety-critical applications such as ICS and healthcare IoT, decision makers require interpretable model outputs that can be audited and explained. Current DL-based IDS operate as black boxes, which limits regulatory acceptance and operator trust. Post-hoc explainability methods (SHAP, LIME) and inherently interpretable architectures (attention mechanisms, GNN) require further investigation in the IDS context [5].

Edge Deployability and Lightweight Models. The computational footprint of CNN-LSTM hybrids (hundreds of MB, multiple GPU-seconds of inference) is incompatible with the resource constraints of typical IoT edge devices (microcontrollers with <1 MB RAM, milliwatt power budgets). Model compression techniques pruning, quantisation, knowledge distillation must be systematically applied to IDS architectures without unacceptable accuracy loss [2].

Future Research Directions. We identify the following as the most impactful research priorities:

- 1) *Federated learning with differential privacy: Combining FL with formal privacy guarantees (differential privacy, homomorphic encryption) for deployment in regulated sectors.*
- 2) *Transfer learning across IoT domains: Pre-training on large, diverse datasets and fine-tuning on domain-specific traffic to reduce labelling requirements.*
- 3) *Graph Neural Networks for topology-aware IDS: Modelling IoT networks as heterogeneous graphs to capture inter-device relationships exploited in multi-hop attacks.*
- 4) *Standardised IoT IDS benchmarking: Developing unified evaluation protocols that account for class imbalance, dataset recency, and real-world deployment constraints.*
- 5) *Explainable AI (XAI) for IDS: Developing attention-based or inherently interpretable DL architectures that can generate human-readable justifications for detections.*

VIII. CONCLUSION

This paper has presented a comprehensive comparative study of deep learning-based intrusion detection systems for IoT security, encompassing six principal DL paradigms (CNN, LSTM, CNN-LSTM, DNN, AE, and FL-based approaches), five benchmark datasets, and two forensic case studies. Our analysis demonstrates that hybrid CNN-LSTM architectures achieve state-of-the-art detection accuracy (99.87–99.99%) on established IoT datasets but impose substantial computational overhead that challenges edge deployment. Federated learning based IDS offer a compelling privacy preserving alternative with moderate accuracy sacrifice (95–97%), particularly suited to distributed IoT deployments where data centralisation is infeasible.

The Mirai/Dyn and Oldsmar case studies illustrate that both network-layer and process-layer anomaly detection are necessary for comprehensive IoT and ICS security, and that DL-based IDS could have provided the automated, real-time detection capabilities that human operators alone cannot maintain. Key open challenges class imbalance, concept drift, adversarial robustness, and edge deployability must be systematically addressed before DL-based IDS can fulfil their potential in production IoT environments.

The integration of artificial intelligence with cybersecurity represents not merely a technical advancement but a strategic imperative: as the scale and sophistication of IoT attacks continue to escalate, AI-driven IDS will be indispensable for protecting the connected infrastructure upon which modern society increasingly depends.

REFERENCES

- [1] A. Aldhaheri, F. Alwahedi, M. A. Ferrag, and A. Battah, "DeDeep Learning for Cyber Threat Detection in IoT Networks: A Review" *IoT and Cyber-Physical Systems*, vol. 4, pp. 110–128, 2024.
- [2] B. R. Kikissagbe and M. Adda, "Machine learning-based intrusion detection methods in IoT systems: A comprehensive review," *Appl. Sci.*, 2024.
- [3] M. Antonakakis et al., "Understanding the Mirai botnet," in *Proc. 26th USENIX Security Symp.*, Vancouver, BC, Canada, Aug. 2017, pp. 1093–1110.
- Understanding the Mirai Botnet [4] CyOTE Program, "Remote access attack on Oldsmar water treatment plant," Idaho National Laboratory, Tech. Rep. CyOTE, 2022.
- 5] A. Khraisat, A. Alazab, M. Alazab, A. Obeidat, S. Singh and T. Jan, "Federated learning for intrusion detection in IoT environments: A privacy-preserving strategy," *Discover Internet of Things*, Springer Nature, 2025.
- [6] J. F. Cevallos M., "Deep Reinforcement Learning for intrusion detection in Internet of Things: Best practices, lessons learnt, and open challenges" *Computer Networks*, vol. 236, p. 110016, 2023.
- [7] S. Jain, P. M. Pawar, and R. Muthalagu, "Hybrid intelligent intrusion detection system for internet of things," *Telematics and Informatics Reports*, vol. 8, Dec. 2022.
- [8] A. Halbouni, T. S. Gunawan, M. H. Habaebi, M. Halbouni, M. Kartiwi, and R. Ahmad, "CNN-LSTM: Hybrid deep neural network for network intrusion detection system," *IEEE Access*, vol. 10, pp. 99837–99849, 2022.
- [9] J. Kim, J. Kim, H. Kim, M. Shim, and E. Choi, "CNN-based network intrusion detection against denial-of-service attacks," *Electronics*, vol. 9, no. 6, p. 916, 2020.
- [10] M. S. A. Muthanna et al., "Towards SDN-enabled, intelligent intrusion detection system for Internet of Things (IoT)," *IEEE Access*, vol. 10, pp. 22756–22768, 2022.
- [11] Z. Liu et al., "Deep learning approach for IDS," in *4th Int. Congr. Information and Communication Technology*, Singapore: Springer, 2020, pp. 471–479.
- [12] Z. Ahmad et al., "Anomaly detection using deep neural network for IoT architecture," *Applied Sciences*, vol. 11, no. 15, p. 7050, 2021.
- [13] I. Apostol, M. Preda, C. Nila, and I. Bica, "IoT botnet anomaly detection using unsupervised deep learning," *Electronics (Switzerland)*, vol. 10, no. 16, Aug. 2021.
- [14] M. A. Ferrag, O. Friha, L. Maglaras, H. Janicke, and L. Shu, "Federated deep learning for cyber security in the Internet of Things: Concepts, applications, and experimental analysis," *IEEE Access*, vol. 9, pp. 138509–138542, 2021.
- [16] A. N. Jahromi, H. Karimipour, and A. Dehghantanha, "An ensemble deep federated learning cyber-threat hunting model for industrial internet of things," *Computer Communications*, vol. 198, pp. 108–116, 2023.
- [17] M. Tavallaei, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in *Proc. IEEE Symp. Computational Intelligence for Security and Defense Applications*, 2009, pp. 1–6.
- [18] 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.
- [19] 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. [19] 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”. [20] T. Saba, A. Rehman, T. Sadad, H. Kolivand, and S. A. Bahaj, “Anomaly-based intrusion detection system for IoT networks through deep learning model,” Computers and Electrical Engineering, vol. 99, p. 107810. [21] D. C. S. Students, “A case study on Mirai Botnet Attack of 2016, Medium, Apr. 2023. [22] CISA, “Heightened DDoS Threat Posed by Mirai and Other Botnets,” US-CERT Alert TA16-288A, Oct. 2016. [23] J. Cochran et al., “Inside the infamous Mirai IoT botnet: A retrospective analysis,” Cloudflare Blog, Oct. 2016. [24] Center for Internet Security, “The Mirai Botnet – Threats and mitigations,” CIS Blog, Jul. 2021. [25] M. Antonakakis et al., “Understanding the Mirai botnet,” USENIX Security 2017. [26] S. Schwartz, “Security flaws enabled Florida city water utility hack,” Cybersecurity Dive, Feb. 2021. [27] Industrial Defender, “Florida Water Treatment Plant Hit With Cyber Attack,” Feb. 2021. [28] Logical Systems Inc., “Security Incident at Oldsmar Water Treatment Plant and Lessons Learned,” Tech. Report, Mar. 2021. [29] T. Hossain et al., “A high performance hybrid LSTM-CNN secure architecture for IoT environments using deep learning,” PMC/NCBI, 2025. [30] T. Rana and A. Shafiq, “Hybrid AI intrusion detection: Balancing accuracy and efficiency,” PMC/NCBI, 2025. [Online]. [31] Transforma Insights, “Global IoT connections will grow from 21 billion in 2025 to 48 billion in 2035” Internet of Things, Apr. 2026.