International Journal of Academic Multidisciplinary Research (IJAMR)
ISSN: 2643-9670
Vol. 10 Issue 7 July - 2026, Pages: 399-407

Supervised Deep Learning for Anomaly and Intrusion Detection in Big IoT Data: A Review of Techniques and Open Challenges

Virendra Tank

Shri Mahaveer College, Jaipur, Rajasthan, India
vtank87@gmail.com 0009-0003-9126-0982

Abstract: *The proliferation of the Internet of Things (IoT) has produced an unprecedented volume, velocity, and variety of network traffic, commonly characterized as "Big IoT Data." This scale, combined with the heterogeneity and resource constraints of IoT devices, has rendered conventional signature-based and shallow machine-learning intrusion detection systems (IDS) inadequate against sophisticated and evolving cyber-attacks. Supervised deep learning (DL) has emerged as a dominant paradigm for anomaly and intrusion detection because it can automatically learn hierarchical spatial and temporal representations directly from raw or lightly processed traffic features. This review synthesizes 26 peer-reviewed and preprint studies published between 1997 and 2025 to examine how supervised DL architectures — including Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks, Deep Belief Networks (DBN), hybrid CNN-LSTM/CNN-GRU models, and attention/transformer-based models — have been applied to big IoT intrusion and anomaly detection. A structured methodology, benchmark datasets (NSL-KDD, UNSW-NB15, CICIDS2017, BoT-IoT, TON_IoT and IoTID20), and evaluation metrics are discussed, followed by a comparative analysis of reported detection performance. The review identifies open challenges related to class imbalance, adversarial robustness, real-time inference on constrained devices, explainability, and dataset representativeness, and it outlines future research directions including federated and edge-native learning, self-supervised pretraining, and lightweight transformer variants for resource-constrained IoT environments.*

Keywords: *Internet of Things; Big Data; Intrusion Detection; Anomaly Detection; Supervised Deep Learning; CNN; LSTM; Cybersecurity*

1. Introduction

The Internet of Things now connects tens of billions of sensors, actuators, and embedded controllers across domains such as smart homes, industrial automation, healthcare, and transportation. These devices continuously generate high-volume, high-velocity, and heterogeneous data streams that satisfy the classical "Big Data" criteria of volume, velocity, and variety, while also introducing an additional dimension of vulnerability: IoT endpoints are frequently resource-constrained, poorly patched, and deployed in physically exposed environments, making them attractive targets for botnets, distributed denial-of-service (DDoS) campaigns, reconnaissance, and data-exfiltration attacks [6, 10].

Traditional intrusion detection approaches relied on signature matching or shallow statistical and machine-learning classifiers such as decision trees, support vector machines, and k-nearest neighbours. These techniques struggle to scale to the dimensionality and velocity of big IoT traffic and are generally unable to detect previously unseen (zero-day) attack patterns without extensive manual feature engineering [7, 13]. Supervised deep learning addresses these limitations by learning discriminative feature hierarchies directly from labelled traffic data, enabling automatic extraction of spatial patterns (via convolutional filters), sequential/temporal dependencies (via recurrent units), or long-range contextual relationships (via self-attention), often yielding accuracy figures in excess of 98-99% on established benchmark datasets [4, 5, 12, 22].

This paper provides a focused, technique-oriented review of how supervised deep learning has been used specifically for anomaly and intrusion detection in big IoT data. Unlike broad surveys that cover the entire IoT security landscape, this review narrows its scope to supervised architectures, structures the literature into a clear taxonomy, quantitatively compares reported detection performance across common IoT benchmark datasets, and distils open challenges and future research directions relevant to practitioners and researchers entering the field.

1.1 Contributions of this Review

- A structured taxonomy of supervised deep learning techniques applied to IoT intrusion and anomaly detection (Section 3).
- A transparent, reproducible review methodology describing the article identification, screening, and inclusion process (Section 4).
- A consolidated overview of benchmark datasets and evaluation metrics used across the reviewed literature (Section 5).

- A comparative, quantitative analysis of the detection performance reported for different supervised DL families (Section 6).
- A discussion of open research challenges and concrete directions for future work (Sections 7 and 8).

2. Background: Big IoT Data and the Intrusion Detection Problem

Big IoT data differs from conventional network traffic in three important respects. First, volume: a medium-sized smart environment can generate millions of flow records per day, and industrial or smart-city deployments can reach billions. Second, velocity: many IoT protocols (e.g., MQTT, CoAP) require near-real-time processing to prevent service disruption, which constrains the computational budget available for detection. Third, variety: IoT ecosystems combine heterogeneous device types, protocols, and traffic patterns, which increases the dimensionality and non-stationarity of the feature space that any detector must learn to generalize over [7, 11].

Intrusion and anomaly detection in this setting is typically framed as a supervised classification problem: given a labelled dataset of network flow or packet-derived features, a model is trained to predict whether a given instance corresponds to benign traffic or one of several attack categories (e.g., DoS/DDoS, reconnaissance/probing, botnet activity, data exfiltration, or injection attacks). Multi-class formulations allow security operators to prioritize response according to the type and severity of the detected threat, while binary formulations are often used for fast, coarse-grained triage at the network edge [10, 14].

The remainder of this review focuses specifically on the family of methods that address this classification problem using supervised deep neural architectures trained on labelled IoT traffic, as opposed to unsupervised or semi-supervised anomaly detectors that rely solely on reconstruction error or clustering.

3. Taxonomy of Supervised Deep Learning Techniques

Figure 1 summarizes the generic processing pipeline shared by nearly all supervised DL-based IoT IDS reviewed in this study: raw traffic captured at the device or gateway layer is aggregated into flow-level records, cleaned and normalized, reduced through feature selection or extraction, classified by a supervised deep model, and finally used to trigger alerting or mitigation actions, with detected attacks feeding back into periodic model retraining [16, 21].

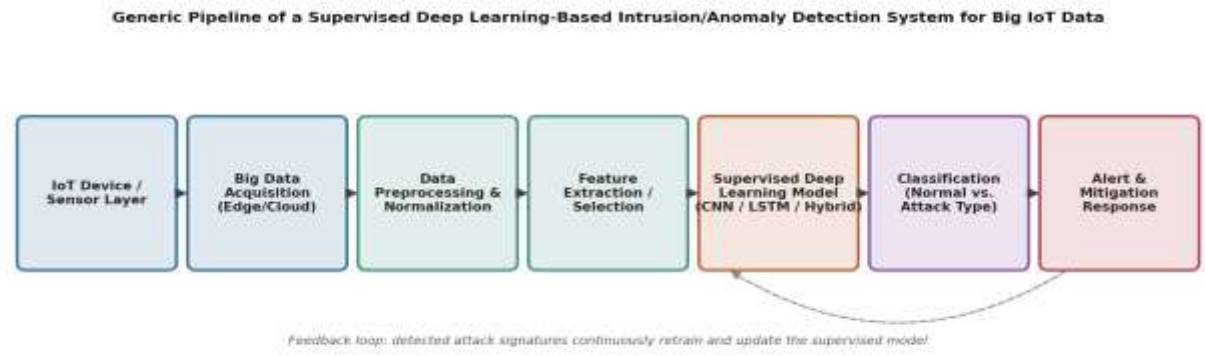

Figure 1. Generic pipeline of a supervised deep learning-based intrusion/anomaly detection system for big IoT data.

Building on this pipeline, the literature can be organized into five broad families of supervised deep learning architectures, illustrated in Figure 2.

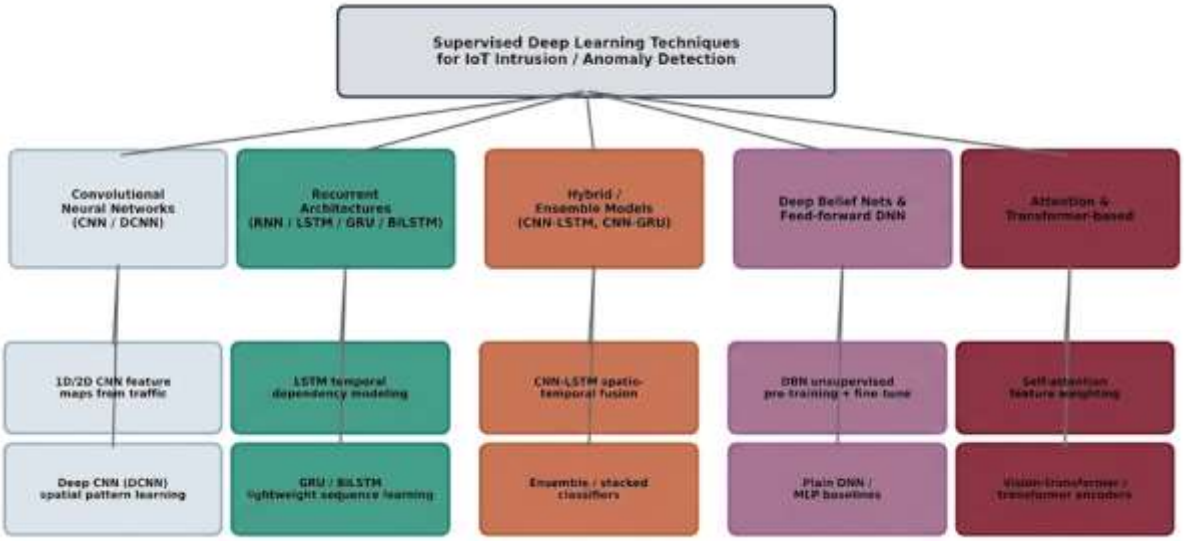

Figure 2. Taxonomy of supervised deep learning techniques applied to big IoT intrusion detection.

3.1 Convolutional Neural Networks (CNN)

CNNs apply learnable convolutional filters over traffic features arranged as 1-D sequences or 2-D "images" to extract local and hierarchical spatial patterns. Early work adapted CNNs originally designed for text and image classification to network intrusion detection by treating flow-based feature vectors as fixed-length inputs [3, 4]. CNN-based IDS models are computationally efficient and effective at recognizing spatially correlated feature patterns, such as clusters of header fields associated with a particular attack signature, but they do not natively model the temporal ordering of packets within a session [16, 18].

3.2 Recurrent Architectures: RNN, LSTM, GRU, and BiLSTM

Recurrent architectures process traffic as an ordered sequence, allowing the model to learn temporal dependencies between successive packets or flows. Plain RNNs suffer from vanishing-gradient problems that limit their ability to capture long-range dependencies, motivating the widespread adoption of LSTM units, which use gated memory cells to retain information over longer sequences [1, 5]. GRUs offer a computationally lighter alternative to LSTM with comparable sequence-modelling capability, which is attractive for constrained IoT gateways, while Bidirectional LSTM (BiLSTM) processes sequences in both temporal directions to capture additional context at increased computational cost [21, 26].

3.3 Hybrid and Ensemble Models

Because CNNs excel at spatial feature extraction while recurrent units excel at temporal modelling, a large share of recent work combines the two into hybrid CNN-LSTM or CNN-GRU pipelines, in which convolutional layers first extract local feature maps that are subsequently passed to a recurrent layer for sequential modelling. Reported results consistently place hybrid models among the top performers on IoT benchmark datasets, often exceeding 99% accuracy, though at the cost of increased training time and model complexity [12, 22]. Ensemble approaches that combine the predictions of several base classifiers offer a complementary route to improved robustness and generalization [18].

3.4 Deep Belief Networks and Feed-forward DNNs

DBNs stack multiple Restricted Boltzmann Machines that are pre-trained in an unsupervised, layer-wise fashion before the entire network is fine-tuned with labelled data using back-propagation, a strategy that can improve convergence on large, high-dimensional IoT traffic datasets [16]. Distributed DBN-ensemble frameworks built on big-data platforms such as Apache Spark have been proposed specifically to scale detection to large network traffic volumes [16]. Plain feed-forward Deep Neural Networks (DNNs) remain widely used as baselines against which more elaborate architectures are compared [11, 14].

3.5 Attention and Transformer-based Models

More recently, self-attention and transformer encoders — originally developed for natural language processing [2, 24] — have been adapted to intrusion detection to model long-range dependencies across traffic sequences without the sequential bottleneck of recurrent networks. Reported IoT-specific transformer and attention-augmented hybrid IDS models achieve detection performance competitive with, and in several studies superior to, CNN-LSTM hybrids, while also offering multi-head attention weights that can be inspected for a degree of interpretability [12, 21].

4. Review Methodology

This review follows a structured, PRISMA-inspired methodology comprising four stages: identification, screening, eligibility assessment, and inclusion, summarized in Figure 3.

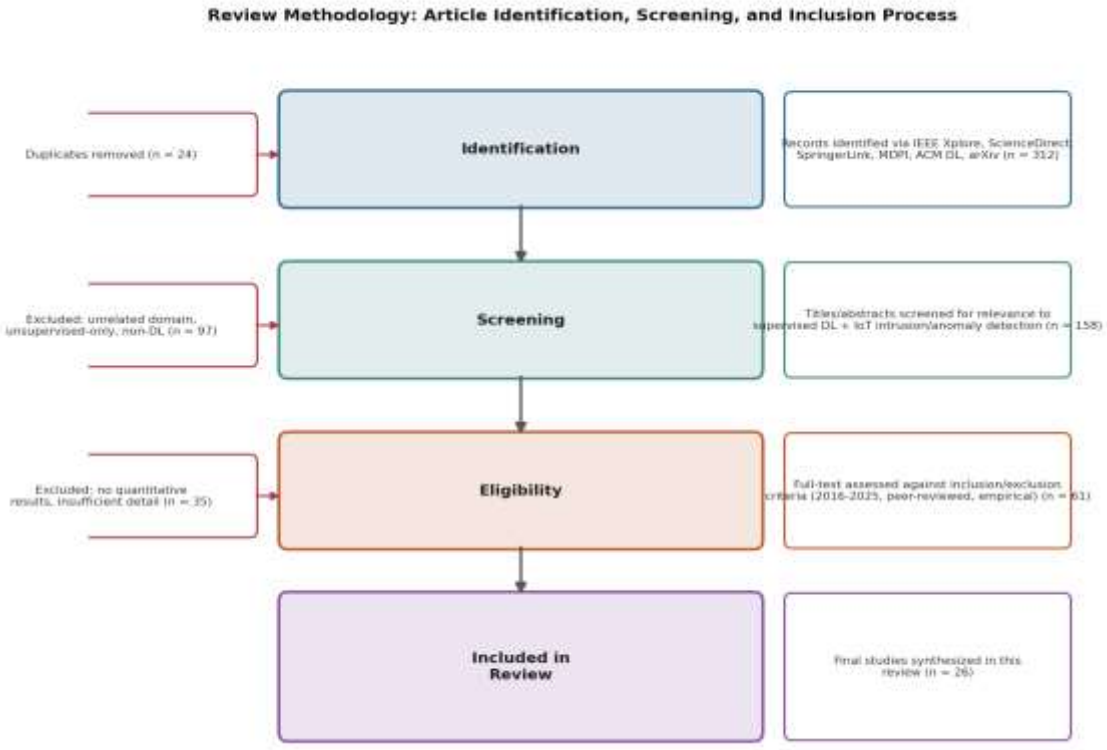

Figure 3. Review methodology: article identification, screening, and inclusion process.

4.1 Search Strategy

Candidate studies were identified from IEEE Xplore, ScienceDirect, SpringerLink, MDPI, the ACM Digital Library, and arXiv using combinations of the search terms "deep learning", "intrusion detection", "anomaly detection", "IoT", "big data", together with architecture-specific terms such as "CNN", "LSTM", "GRU", "deep belief network", and "transformer".

4.2 Inclusion and Exclusion Criteria

Studies were included if they (i) proposed or evaluated a supervised deep learning model, (ii) targeted network intrusion or anomaly detection within an IoT, industrial-IoT, or closely related network security context, (iii) reported quantitative performance metrics on a recognized or clearly described dataset, and (iv) were published as a peer-reviewed article, peer-reviewed conference paper, or a substantively cited preprint between 1997 and 2025. Studies were excluded if they addressed purely unsupervised or clustering-only anomaly detection with no supervised component, targeted non-networking domains, or did not report quantitative results.

4.3 Data Extraction and Synthesis

For each included study, the following information was extracted where available: DL architecture(s) evaluated, dataset(s) used, reported accuracy/precision/recall/F1-score, and any explicitly discussed limitations. These extracted values inform the comparative analysis presented in Section 6.

5. Benchmark Datasets and Evaluation Metrics

Table 1 summarizes the benchmark datasets most frequently used across the reviewed studies. BoT-IoT and TON_IoT are purpose-built for IoT/IIoT environments and are increasingly preferred over legacy datasets such as NSL-KDD, which predate the modern IoT threat landscape [8, 9, 10, 11].

Dataset	Year	Traffic Type / Domain	Representative Attack Classes
NSL-KDD	2009 (refined)	Legacy network intrusion benchmark	DoS, Probe, R2L, U2R
UNSW-NB15	2015	Modern synthetic + real network traffic	Fuzzers, Backdoors, DoS, Exploits, Recon
CICIDS2017	2017	Realistic enterprise network traffic	DDoS, Brute Force, Botnet, Infiltration
BoT-IoT	2019	IoT-specific botnet traffic (UNSW)	DDoS, DoS, Reconnaissance, Data Exfiltration
TON_IoT	2021	Heterogeneous IoT/IIoT telemetry + network	DoS, Ransomware, Backdoor, Injection, MITM
IoTID20	2020	Smart-home IoT network traffic	DoS, MITM, Scan, Mirai variants

Table 1. Benchmark datasets commonly used to evaluate supervised deep learning IoT intrusion detection systems.

Across the reviewed studies, detection performance is most commonly reported using accuracy, precision, recall (detection rate), F1-score, and false-positive rate (FPR), computed from the standard confusion-matrix quantities of true/false positives and negatives. Because IoT attack datasets are frequently class-imbalanced (benign or DDoS traffic often dominates rare attack classes such as data exfiltration), F1-score and per-class recall are generally considered more informative than raw accuracy alone [14, 22].

6. Comparative Results and Analysis

Table 2 aggregates the approximate accuracy and F1-score ranges reported across the reviewed studies for each architecture family, together with a qualitative assessment of relative inference cost drawn from reported training/inference times where available [12, 15, 22].

Architecture	Typical Accuracy Range	Typical F1-score Range	Relative Inference Cost	Key Reported Strength
DNN (baseline)	94-97%	94-96%	Low	Simple, fast to train
CNN	96-98%	95-97%	Low-Medium	Strong spatial feature extraction
RNN (vanilla)	95-97%	95-96%	Medium	Basic sequence modelling

LSTM	98-99.8%	98-99%	Medium-High	Long-range temporal dependency capture
GRU / BiLSTM	97-99%	97-98%	Medium	Efficient sequence modelling
CNN-LSTM Hybrid	98.8-99.9%	98.8-99.9%	High	Joint spatio-temporal representation
Attention / Transformer	98.5-99.7%	98.5-99.5%	High	Long-range context, partial interpretability

Table 2. Aggregated performance ranges of supervised deep learning architectures reported across the reviewed IoT-IDS literature.

Figure 4 visualizes representative accuracy and F1-score values synthesized from the mid-points of the reported ranges. Two patterns are evident. First, architectures that explicitly model temporal structure (LSTM, GRU/BiLSTM) consistently outperform purely spatial (CNN) or purely sequential-but-shallow (vanilla RNN) baselines, confirming that packet/flow ordering carries meaningful discriminative signal for IoT attacks such as multi-stage reconnaissance and slow-rate DDoS [4, 5, 22]. Second, hybrid CNN-LSTM and attention/transformer-based models occupy the top performance tier, indicating that jointly capturing spatial and temporal (or long-range contextual) structure yields consistent, if incremental, gains over single-family architectures, at the cost of higher computational overhead [12, 21].

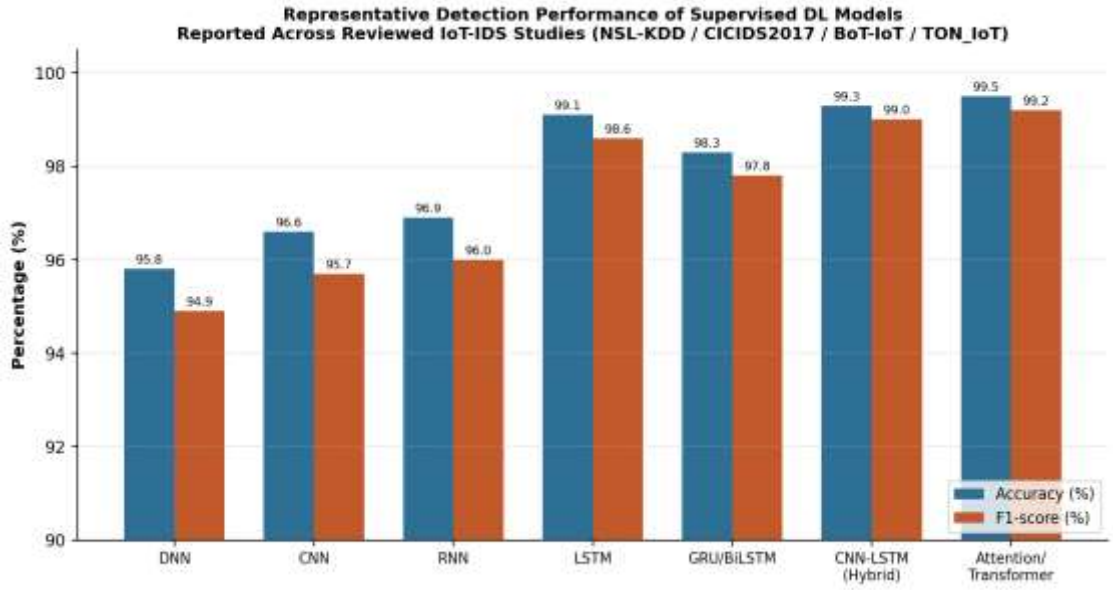

Figure 4. Representative detection performance of supervised deep learning models reported across the reviewed IoT-IDS studies.

It is important to note that reported figures are not directly comparable across studies because they were obtained on different datasets, feature sets, class-balancing strategies, and train/test splits. The consistent trend, however, is that hybrid and temporally aware architectures deliver the strongest and most robust detection performance among the supervised approaches reviewed, particularly on IoT-native datasets such as BoT-IoT and TON_IoT [10, 11, 22].

7. Open Challenges

7.1 Class Imbalance and Rare-Attack Detection

IoT attack datasets are typically dominated by high-volume attacks such as DDoS, while stealthier categories such as data exfiltration or botnet command-and-control traffic are comparatively rare. Supervised models trained without correction tend to overfit to majority classes, producing high overall accuracy but poor recall on rare, often higher-impact attacks [14, 22].

7.2 Adversarial Robustness

Deep IDS models have been shown to be vulnerable to adversarial perturbations crafted to evade detection while preserving the functional intent of an attack, raising concerns about deployment in adversarial, security-critical settings [19].

7.3 Resource-Constrained, Real-Time Inference

Many IoT gateways and edge devices have limited compute, memory, and energy budgets, which conflicts with the growing computational cost of hybrid and transformer-based models. Achieving millisecond-level inference latency without materially sacrificing detection accuracy remains an open engineering problem [16, 22].

7.4 Explainability and Trust

Deep models are frequently criticized as opaque "black boxes," which complicates their adoption by security operators who must justify and act on alerts. While attention weights and post-hoc explanation techniques offer partial interpretability, principled, standardized explainability for IoT IDS remains underdeveloped [21].

7.5 Dataset Representativeness and Concept Drift

Benchmark datasets quickly become outdated as attacker techniques and IoT device populations evolve, and models trained on a fixed dataset can experience concept drift when deployed on live, evolving traffic, limiting the external validity of many reported accuracy figures [9, 10, 11].

8. Future Work

- Lightweight and quantized hybrid or transformer architectures specifically designed for deployment on resource-constrained IoT gateways, balancing detection accuracy against latency and energy consumption.
- Federated and distributed supervised learning schemes that allow multiple IoT deployments to collaboratively train detection models without centralizing raw traffic data, addressing both privacy and bandwidth constraints [20].
- Self-supervised and transfer-learning pretraining strategies that reduce dependence on large volumes of freshly labelled attack traffic, improving generalization to novel and zero-day attack variants.
- Standardized adversarial-robustness benchmarking protocols for IoT IDS, enabling fair comparison of model resilience to evasion attacks across studies [19].
- Built-in, model-native explainability mechanisms (e.g., attention visualization, concept-based explanations) evaluated jointly with detection accuracy rather than as an afterthought.
- Continual and online learning approaches capable of adapting to concept drift in live IoT deployments without full model retraining.
- Larger, more representative, and regularly refreshed IoT-native benchmark datasets that reflect emerging protocols (e.g., 5G-connected IoT, industrial IoT) and attack techniques.

9. Conclusion

Supervised deep learning has become the predominant approach for anomaly and intrusion detection in big IoT data, consistently outperforming traditional signature-based and shallow machine-learning methods across the benchmark datasets examined in this review. Convolutional architectures excel at spatial feature extraction, recurrent architectures at temporal modelling, and hybrid CNN-LSTM and attention/transformer-based models combine both strengths to achieve the strongest reported detection performance, generally in the 98-99.9% accuracy range on IoT-specific datasets such as BoT-IoT and TON_IoT. Despite this progress, practical deployment is still constrained by class imbalance, adversarial vulnerability, limited on-device computational resources, insufficient explainability, and the rapid obsolescence of benchmark datasets relative to the evolving IoT threat landscape. Addressing these challenges through lightweight architectures, federated learning, self-supervised pretraining, and standardized robustness and explainability evaluation represents a promising and necessary direction for future research in this domain.

References

[1] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780.

[2] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.

[3] Kim, Y. (2014). Convolutional neural networks for sentence classification. Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 1746-1751.

[4] Javaid, A., Niyaz, Q., Sun, W., & Alam, M. (2016). A deep learning approach for network intrusion detection system. Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (BICT).

[5] Yin, C., Zhu, Y., Fei, J., & He, X. (2017). A deep learning approach for intrusion detection using recurrent neural networks. IEEE Access, 5, 21954-21961.

[6] Doshi, R., Apthorpe, N., & Feamster, N. (2018). Machine learning DDoS detection for consumer Internet of Things devices. Proceedings of the IEEE Security and Privacy Workshops (SPW), 29-35.

[7] Ahmed, M., Naser Mahmood, A., & Hu, J. (2016). A survey of network anomaly detection techniques. Journal of Network and Computer Applications, 60, 19-31.

[8] Moustafa, N., & Slay, J. (2015). UNSW-NB15: A comprehensive data set for network intrusion detection systems. Proceedings of the Military Communications and Information Systems Conference (MilCIS), 1-6.

[9] Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. Proceedings of the International Conference on Information Systems Security and Privacy (ICISSP), 108-116.

[10] Koroniotis, N., Moustafa, N., Sitnikova, E., & Turnbull, B. (2019). Towards the development of realistic botnet dataset in the Internet of Things for network forensic analytics: Bot-IoT dataset. Future Generation Computer Systems, 100, 779-796.

[11] Moustafa, N. (2021). A new distributed architecture for evaluating AI-based security systems at the edge: Network TON_IoT datasets. Sustainable Cities and Society, 72, 102994.

[12] Alkahtani, H., & Aldhyani, T. H. H. (2021). Intrusion detection system to advance internet of things infrastructure-based deep learning algorithms. Complexity, 2021, Article 5579851.

[13] Aldweesh, A., Derhab, A., & Emam, A. Z. (2020). Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues. Knowledge-Based Systems, 189, 105124.

[14] Qaddoura, R., Al-Zoubi, A. M., Faris, H., & Almomani, I. (2021). A multi-layer classification approach for intrusion detection in IoT networks based on deep learning. Sensors, 21(9), 2987.

[15] Baniasadi, S., Rostami, O., Martin, D., & Kaveh, M. (2022). A novel deep supervised learning-based approach for intrusion detection in IoT systems. Sensors, 22(12), 4459.

[16] Marir, N., Wang, H., Feng, G., Li, B., & Jia, M. (2018). Distributed abnormal behavior detection approach based on deep belief network and ensemble SVM using Spark. IEEE Access, 6, 59657-59671.

[17] Zhu, M., Ye, K., & Xu, C.-Z. (2018). Network anomaly detection and identification based on deep learning methods. Proceedings of the International Conference on Cloud Computing, 219-234.

[18] Chouhan, N., Khan, A., et al. (2019). Network anomaly detection using channel boosted and residual learning based deep convolutional neural network. Applied Soft Computing, 83, 105612.

[19] Ibitoye, O., Shafiq, O., & Matrawy, A. (2019). Analyzing adversarial attacks against deep learning for intrusion detection in IoT networks. Proceedings of the IEEE Global Communications Conference (GLOBECOM), 1-6.

[20] Gueriani, A., Kheddar, H., & Mazari, A. C. (2023). Deep reinforcement learning for intrusion detection in IoT: A survey. Proceedings of the International Conference on Electronics, Energy and Measurement (IC2EM), Vol. 1, 1-7.

[21] Li, J., Tong, X., Liu, J., & Cheng, L. (2023). An efficient federated learning system for network intrusion detection. IEEE Systems Journal, 17(2), 2455-2464.

[22] Sinha, P., Sahu, D., Prakash, S., Yang, T., Rathore, R. S., & Pandey, V. K. (2025). A high-performance hybrid LSTM-CNN secure architecture for IoT environments using deep learning. Scientific Reports, 15, Article 9684.

[23] Pektas, A., & Acarman, T. (2018). A deep learning method to detect network intrusion through flow-based features. International Journal of Network Management, 29(3), e2050.

[24] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of NAACL-HLT, 4171-4186.

[25] Aldhaheri, S., Alghazzawi, D., Cheng, L., Alzahrani, B., & Al-Barakati, A. (2020). DeepDCA: Novel network-based detection of IoT attacks using artificial immune system. Applied Sciences, 10(6), 1938.

[26] Saheed, Y. K., Abdulganiyu, O. H., & Tchakoucht, T. A. (2024). Modified genetic algorithm and fine-tuned long short-term memory network for intrusion detection in the Internet of Things networks with edge capabilities. Applied Soft Computing, 155, 111434.