Computer Modeling in Engineering & Sciences

Tech Science Press

https://doi.org/10.32604/cmes.2026.085472

REVIEW

Deep Reinforcement Learning-Based Intrusion Detection in IoT Networks: A Systematic Mapping and Literature Review

Maryam Omar Abdullah Sawad[1], Said Jadid Abdulkadir[1,2,*], Hitham Seddig Alhussian[1,2] and Majdy Mohamed Eltayeb Eltahir[3]

[1]Department of Computing, Universiti Teknologi PETRONAS, Seri Iskandar, Perak, Malaysia
[2]Center for Research in Data Science (CeRDaS), Universiti Teknologi PETRONAS, Seri Iskandar, Perak, Malaysia
[3]Unit of Specialization of Technology and Engineering, Applied College of Muhayil Asir, King Khalid University, Muhayil Asir, Saudi Arabia

*Corresponding Author: Said Jadid Abdulkadir. Email: saidjadid.a@utp.edu.my

Received: 11 May 2026; Accepted: 29 June 2026; Published: 28 August 2026

ABSTRACT: The increasing complexity and heterogeneity of cyberattacks targeting Internet of Things (IoT) environments, driven by the diversity of interconnected nodes and communication channels, necessitate the development of more advanced and intelligent cyber defence techniques. However, the most effective methods are Machine Learning (ML)-based and Deep Learning (DL)-based intrusion detection systems (IDS), which perform well but still face significant limitations and challenges. To address these issues, Deep Reinforcement Learning (DRL) has been proposed in recent years to automatically resolve the issues by detecting attacks in IoT environments. Therefore, this Systematic Literature Review (SLR) presents an up-to-date review by analyzing the existing studies on DRL-based IDS models that detect intrusions in IoT networks. To achieve this goal, this review focuses on and scrutinizes scientific journals and articles extracted from 2020 to 2026 across multiple databases, identifying 267 articles. A systematic mapping procedure was then carried out using Rayyan and Mendeley to screen the articles based on nine well-defined inclusion and exclusion criteria covering publication year, language, publication type, full-text availability, explicit use of DRL techniques, relevance to IoT attack detection, minimum page count, duplication, and open access, which collectively reduced the pool to 26 high-quality studies. The majority of excluded articles 241 in total were removed primarily because they did not explicitly employ DRL techniques in an IoT intrusion detection context, were duplicate records, or did not meet the minimum quality thresholds defined in the eligibility assessment. The review reveals that the most used algorithm for DRL-based IDS is Deep Q-Network (DQN), appearing in 8 studies (30.8%). Three studies utilized feature selection methods, including LightGBM and Mutual Information Feature Selection (MIFS), and the most frequently targeted attacks are DoS, DDoS, Backdoors, Mirai, Reconnaissance, Scan, and Torii. Finally, this research highlights the open issues and challenges for future research in DRL-based IDS models, to enhance IoT network security.

KEYWORDS: Cyber attacks; deep reinforcement learning (DRL); Internet of Things (IoT); intrusion detection systems (IDS); network security

1 Introduction

The Internet of Things (IoT) is a rapidly emerging field that has transformed traditional environmental sensing by connecting physical objects to the Internet [1]. IoT devices, such as cameras, door locks, medical sensors, smart TVs, and fitness trackers, are used in various industries [2,3]. These devices collect, measure, and understand the environment, improving people's quality of life. With an estimated 27 billion IoT devices

projected by 2025, IoT is one of the fastest-growing fields in computing history, playing a crucial role in real-life applications such as healthcare, homes, transportation, and education [2,4–6].

However, IoT faces several challenges, including security, interoperability, efficiency, and scalability [7,8]. Security is the primary challenge, as IoT systems are integrated across critical industries like military and smart grids. The growing attack surface makes IoT devices prime targets for hackers, potentially resulting in property losses. The widespread and interconnected nature of IoT systems also puts users' personal information at risk of theft or unauthorized access [7,9].

Traditional security measures like encryption, authentication, and access control are insufficient for IoT systems, exposing them to vulnerabilities exploited by sophisticated attacks like Mirai botnets [10,11]. In contrast, intrusion detection systems (IDS) are seen as an efficient way to monitor network packets and identify malicious traffic [12]. IDS acts as a second line of defense, catching attackers before they can cause damage. Existing security mechanisms and conventional IDSs detect sinkhole and wormhole attacks [13] but do not support identifying complex and larger attacks in the real IoT environment. Therefore, developing intelligent security mechanisms capable of detecting attacks in dynamic environments is crucial.

Reinforcement Learning (RL) is a machine learning (ML) technique utilized for developing IoT security solutions [14]. RL algorithms have unique characteristics, making them suitable for developing robust solutions [10,15]. They can adjust their behavior dynamically and learn through experience, minimizing complexities in IoT scenarios where data collection is challenging, and traditional datasets are scarce [16]. However, RL is not a universal solution for all IoT security challenges, and its applicability varies based on specific use cases and threat models. For example, RL may require more computational resources than alternatives, posing challenges in increasingly complex and expansive IoT environments [17].

Researchers like [18] have combined Deep Learning (DL) and RL techniques to improve the efficacy of RL algorithms. This approach has been applied in healthcare and cloud computing, enhancing agent intelligence, optimizing policies, and efficiently handling high-dimensional data with scalability. The combination of DL and RL is called Deep Reinforcement Learning (DRL). The use of DRL in IoT security has gained significant attention in the past few years [19–21]. Recent advances include the D3O-IIoT framework [22], which employs a Dueling Deep Q-Network agent for dynamic deception orchestration in IIoT security, and hybrid Convolutional Neural Network–Long Short-Term Memory (CNN-LSTM) architectures, combined with DQN and PPO for securing Internet of Medical Things (IoMT) healthcare networks [23]. However, the proposed DRL-based IDS models still need extensive testing and review in real-world IoT environments to be widely adopted.

1.1 Motivation for the Study

Due to the increasing complexity of IoT network systems, DRL-based intrusion detection systems must meet various requirements. As smart devices connect more heterogeneously, these systems become more susceptible to attacks. A comprehensive review of these systems is crucial to understanding and analyzing each component of the intrusion detection process, ensuring best practices during the development of the detection engine.

Most recent surveys and reviews on DRL-based IDSs concentrate on taxonomic classification without systematically examining the specific DRL algorithms employed, the feature engineering strategies adopted, or the IoT-specific evaluation metrics applied [18,24–29]. While feature selection has been addressed in the intra-vehicle IDS literature [30], its application within the narrower and more constrained context of DRL-based IDS models designed specifically for IoT environments remains critically underexplored. Notably, among the 26 primary studies identified in this review, only three explicitly incorporated feature selection (two using LightGBM and one using MIFS), which underscores the limited adoption of systematic feature engineering practice in this domain rather than contradicting it. This gap, combined with the lack

of comprehensive analysis of datasets, attack types, and network performance metrics in existing reviews, motivates the present study. Table 1 summarizes the related works and positions this study relative to prior surveys.

1.2 Contribution and Structure of the Study

Despite the growing adoption of DRL in IoT security, relatively few reviews have provided a unified, PRISMA-guided synthesis that jointly examines DRL algorithm selection, feature engineering practices, attack taxonomies, dataset suitability, and evaluation metric comprehensiveness within IoT intrusion detection. This review consolidates these dimensions through the following contributions:

1. A systematic and reproducible review of DRL-based IDS models for IoT environments, covering 26 rigorously selected primary studies published between 2020 and 2026, conducted in accordance with PRISMA guidelines and a structured nine-criterion inclusion and exclusion framework.
2. A critical comparative analysis of nine DRL algorithm categories employed in IoT IDS development, examining not only their algorithmic formulations but also their adaptation strategies, performance characteristics, and practical limitations within resource-constrained IoT deployment contexts.
3. A systematic assessment of feature engineering practices in DRL-based IDS models, identifying the critical underutilization of feature selection and its implications for model efficiency and deployability on constrained IoT devices.
4. A taxonomically structured analysis of IoT attack types targeted by existing DRL-based IDS models, distinguishing between attack strategies, attack mechanisms, and IoT-specific malware families.
5. A weighted critical evaluation of datasets and evaluation metrics used across reviewed studies, including a cross-study performance comparison, overfitting risk assessment, and identification of the persistent imbalance between detection-oriented and deployment-oriented evaluation frameworks.
6. The identification of six higher-order research gaps not previously synthesised in the literature, providing a structured agenda for future research in DRL-based IoT security.

Table 1 systematically positions this review against eleven prior surveys across eleven comparative dimensions. As summarized, few existing reviews collectively address the range of dimensions covered in this study. Specifically, this review extends prior works in five respects. First, unlike [25–27,29], which focus exclusively on DRL taxonomy without examining feature engineering, this review provides systematic assessment of feature selection practices within DRL-based IDS for IoT, revealing a critical underestimation that directly impacts model deployability. Second, unlike [24,28], which address RL broadly without IoT-specific algorithmic analysis, this review critically examines nine DRL algorithms with explicit discussion of their adaptation strategies and limitations in resource-constrained IoT contexts. Third, in contrast to most prior reviews, this study provides a cross-study quantitative performance comparison with overfitting risk assessment across shared benchmark datasets, addressing the absence of quantitative synthesis identified as a recurring limitation. Fourth, this review applies a taxonomically rigorous classification of IoT attack types, distinguishing attack strategies, attack mechanisms, and IoT-specific malware families, a distinction rarely made explicit in prior surveys. Fifth, this review provides a weighted critical evaluation of dataset suitability for IoT deployment, explicitly identifying methodological misalignment between benchmark selection and IoT operational contexts, and extends coverage to emerging research published through early 2026.

The remainder of this article is structured as follows. Section 2 provides the necessary background on IoT, IDS classifications, DRL algorithms, and feature selection methods. Section 3 reviews and critically positions related surveys. Section 4 details the systematic review methodology. Section 5 presents the analysis and synthesis across five research questions. Section 6 identifies research gaps and future directions, and Section 7 presents the conclusions.

Table 1: Evaluation of the related works.

Study	Type of Study	IoT	Security Techniques		Type of Algorithm	Focus on DRL	Feature Eng.	Attacks	Datasets		Evaluation Metrics	
			IDS	Others					Mention	Criticism	Data Mining	Network Perf.
[29]	Review	✓	✓	×	DRL	✓	×	×	✓	×	×	×
[18]	Review	×	✓	✓	DRL	✓	×	✓	×	×	×	×
[24]	SLR	✓	✓	✓	RL, DRL	×	×	×	✓	✓	✓	×
[25]	SLR	×	✓	×	DRL	✓	×	×	✓	×	✓	×
[26]	Survey	✓	✓	×	DRL	✓	×	✓	✓	×	×	×
[28]	Critical Review	✓	✓	×	ML, DL, RL, DRL	×	×	✓	✓	✓	✓	×
[27]	Survey	✓	✓	×	DRL	✓	×	×	✓	×	×	×
[31]	Survey	✓	✓	×	DRL	✓	×	✓	✓	×	✓	×
[32]	SLR	✓	✓	×	DRL	✓	×	✓	✓	✓	✓	×
[33]	Review	✓	×	×	FL	×	×	✓	✓	✓	✓	×
[30]	Review	×	✓	×	ML, DL	×	✓	✓	✓	✓	✓	×
This study	Review	✓	✓	✓	DRL	✓	✓	✓	✓	✓	✓	✓

Note: Symbol (✓) shows the paper covers that specific topic, and (×) shows the paper does not cover the topic.

Fig. 1 presents a visual overview of the overall structure of this review, summarizing the seven-section organization, the PRISMA-based study selection process that reduced 267 candidate studies to 26 included primary studies, the five research questions addressed in the analysis, and the six research gaps identified for future work.

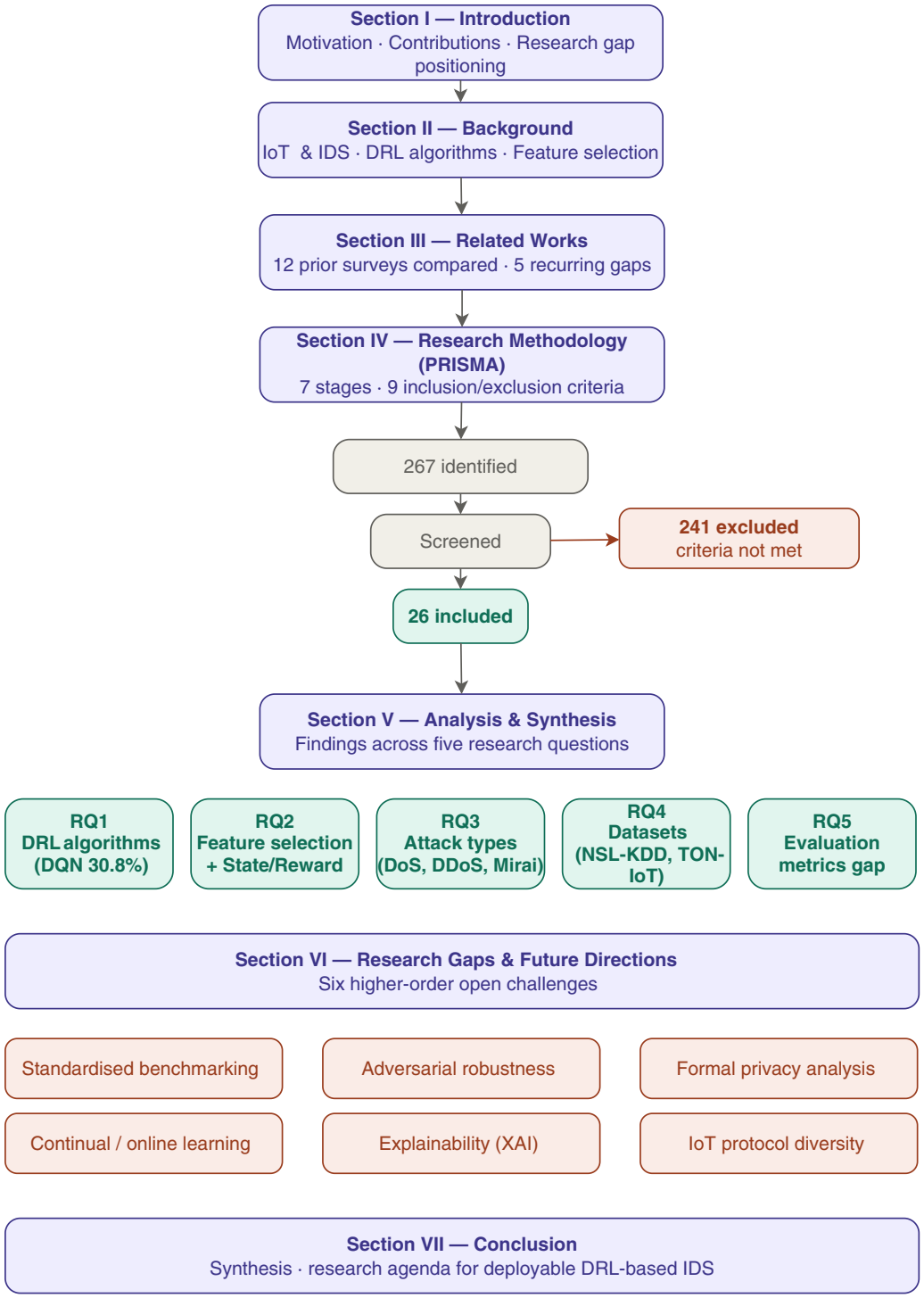

Figure 1: Overall structure of the systematic literature review.

2 Background

2.1 IoT and Its Framework

The Internet of Things (IoT) was first proposed in 1999 by a member of the Radio Frequency Identification (RFID) development community [34]. It is a ubiquitous network that allows and controls large geographically spread devices, collecting, gathering, and processing data using smart devices [35]. An IoT system consists of sensors communicating via the internet and gateways, generating automated alerts and allowing sensors to evaluate data without human intervention [35]. A user interface will enable users to input or check the system [35]. The IoT system operates on five components: sensors, devices, connectivity, data processing, and the user interface (phone) [35]. Fig. 2 illustrates the fundamentals of how IoT works.

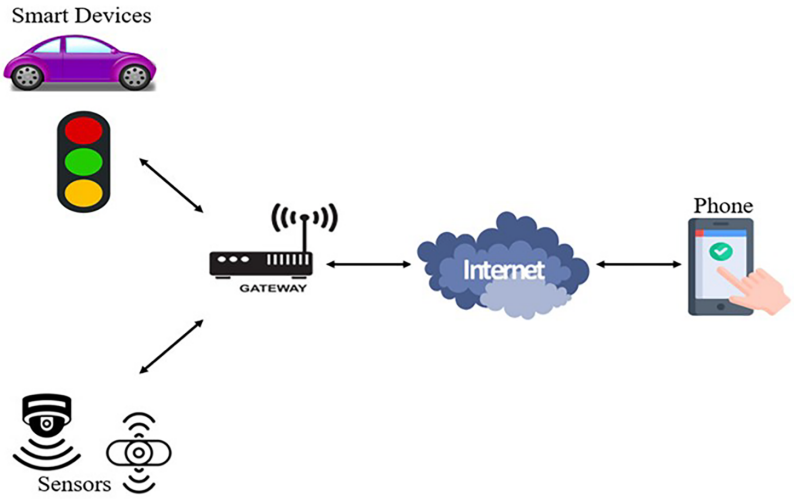

Figure 2: Fundamentals of how IoT works [5].

Recent high-quality studies have further advanced anomaly detection across IoT ecosystems, including the Internet of Medical Things (IoMT) and consumer IoT. One study [36] proposed a collaborative, federated SRU-based network with dynamic behaviour aggregation, reduced communication overhead, and explainable features for privacy-preserving anomaly detection in smart healthcare (IoMT) networks. Another study [37] introduced a federated, boosting-powered cyber-attack detection scheme that enhances security and privacy for consumer IoT. Collectively, these contributions highlight the growing momentum of intelligent IDS research and reinforce the relevance and timeliness of this review.

IoT Challenges

IoT offers numerous benefits but also presents challenges, including security. IoT security is crucial to protect against cyber threats and attacks on people, companies, and governments. IoT devices are often linked to vital infrastructure, making them attractive targets for hackers. Additionally, they are often placed in hostile or remote locations, making them susceptible to assault due to physical access and maintenance challenges [38]. Recent work by [39] has further demonstrated the importance of resource-efficient anomaly detection frameworks at the edge for enhancing IoT resilience in streaming data environments. The vast number of IoT devices and the lack of standardized security protocols make securing these systems complex and challenging. Potential risks include data theft, privacy violations, and physical harm [40,41]. These security challenges can be categorized into technological challenges and security challenges [42,43]. Technological challenges arise from the heterogeneous nature of IoT devices, while security challenges relate to the principles and functionalities needed to achieve a secure network and prevent security attacks.

2.2 IDS for IoT

An IDS is a crucial tool in securing IoT networks, acting as a second line of defense against intruders [12]. IDS monitors network packets to identify malicious traffic, and in an environment with numerous interconnected devices, incorporating IDS is essential for ensuring security and privacy. IDS can be classified into three types based on the intrusion data sources, detection techniques, and placement strategy. Incorporating IDS into IoT networks ensures the security and privacy of interconnected devices and their transmitted data [12,44,45]

Recent studies published in this domain have explored deep learning and federated approaches for IoT intrusion detection, including anomaly-based deep learning models for IoT networks [46], privacy-aware federated learning frameworks [47], and lightweight detection methods suited to resource-constrained devices [48], reflecting the growing emphasis on both detection accuracy and deployment efficiency [49]. Complementary to these detection-focused efforts, deep learning has also been applied to intelligent traffic scheduling for mobile edge computing in IoT, optimizing resource utilization and supporting efficient operation in constrained edge environments [50].

2.2.1 IDS Based on Intrusion Data Source

IDS is divided into host-based IDSs (HIDS) and network-based IDSs (NIDS) categories based on intrusion data sources [28]. HIDS analyzes data from host systems and audit sources such as operating system, Windows server logs, firewall logs, application system audits, or database logs, identifying insider attacks without network traffic. Network-based IDS monitors network traffic through packet capture and NetFlow, allowing for the monitoring of multiple connected computers. Both types of IDS can identify insider attacks, while network-based IDS additionally monitors network traffic.

2.2.2 IDS Based on Detection Techniques

IDS can be categorized into four classes: signature-based, anomaly-based, specification-based, and hybrid-based [44,45]. Signature-based IDS detects attacks by comparing activity signatures against pre-installed signatures in the IDS database. If a match occurs, the system raises an alarm [44,45]. Anomaly-based IDS trains IDS to detect network anomalies by analyzing their behavior. If activity exceeds a specific threshold, an attack has occurred [44,45]. Specification-based approach, also known as the rule-based approach, involves IDS checking network activity against predefined rules and settings to detect misbehaving intruders. This method detects activities that do not meet system specifications [44,45]. The hybrid approach combines multiple approaches to maximize their advantages and minimize drawbacks [44,45].

2.2.3 IDS Based on Placement Strategies

Based on their location, IDS can be categorized into three types: centralized, distributed, and hybrid [51–54]. Centralized IDS Placement is a type of IDS that uses data traffic to identify attacks, performing extensive security checks within a powerful node. It protects networks from Internet-side attacks and botnets, serving as a firewall. However, it has the disadvantage of difficulty monitoring the network during an attack itself [12]. Distributed IDS Placement is a network security method where every node in the network has a full IDS implementation, enabling efficient attack detection at any stage. Collaboration among nodes allows early detection of threats, but it results in high energy and resource consumption [52,54]. Hybrid IDS placement is a method that combines the benefits of centralized and distributed IDS placement. Central nodes with ample resources handle computationally intensive tasks like data analysis and decision-making, while normal nodes perform lightweight duties like monitoring neighbor nodes and responding to mitigation control messages.

This hybrid approach improves the speed of attack detection and reduces resource consumption compared to the centralized approach [52,54].

2.3 Deep Reinforcement Learning (DRL)

Reinforcement learning (RL) is a machine learning approach where an agent interacts with the environment to maximize numerical rewards [14]. It can be formalized using dynamical systems theory, specifically as the optimal control of incompletely known Markov decision processes (MDP). MDP captures the most important aspects of a learning agent's problem and aims to achieve a goal. In RL, an agent interacts discretely with the environment, receiving observations and selecting actions to maximize the reward. The agent must also have a goal related to the environment's state. The framework for RL is illustrated in Fig. 3.

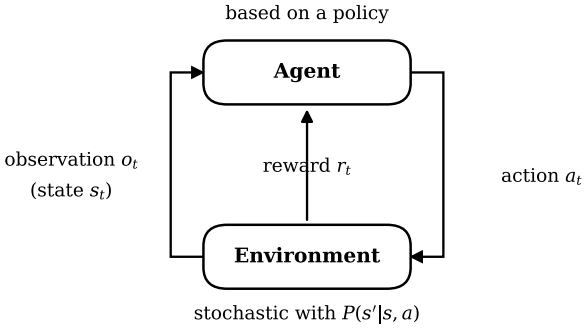

Figure 3: Reinforcement learning framework.

The environment is modeled by a Markov decision process (MDP), consisting of (S, A, Pr, R, γ), where [16]:

i. $S \in \{s_i, i = 1, \ldots, t+1\}$ is a state set describing the environment and the agent.
ii. $A \in \{a_i, i = 1, \ldots, t\}$ is an action set of the agent.
iii. $Pr : S \times A \times S \to [0,1]$ is the transition function, which specifies the probability of transitioning to a new state given the current state and action a.
iv. $R \in \{r_i, i = 1, \ldots, t+1\}$ is the reward function, which specifies the immediate reward that the agent receives after taking action in each state.
v. $\gamma \in [0,1]$ is the discount factor, which determines the relative importance of future rewards compared to immediate rewards.

$$Pr(s_{t+1}|s_t, a_t) = Pr(s_{t+1}|s_t, a_t, \ldots, s_1, a_1) \tag{1}$$

RL aims to find an optimal policy $\pi : S \times A \to [0,1]$ that maximizes the expected return from each state. This can be learned directly [55] or inferred by acting near-greedily according to learned action values [56]. However, RL requires agents to explore the entire system, which is not feasible in many scenarios due to the time it takes to converge and obtain an optimal policy. Traditional RL also suffers from the curse of dimensionality, as the environment becomes complex [17]. DRL is a combination of DL and RL that has been successfully applied to overcome scalability challenges in large-scale environments. DRL can implement various tasks requiring high-dimensional raw input and policy control, solve complex computational tasks, and accelerate learning speed in large-scale problems [55]. This has led to increased efforts in the application of DRL in decision-making, control problems, and IoT security. There are two types of DRL algorithms: Model-based and Model-Free. Model-based methods involve the agent building its internal transition model from environment feedback, using it to understand the effect of actions on states

and rewards. This approach offers higher quality at lower sample complexity, but dealing with uncertainty and model bias is challenging [55]. Model-free algorithms, on the other hand, derive policies without relying on building environmental models or maps. Instead, agents interact directly with the environment and improve performance based on explored samples through trial-and-error learning. This approach is valuable in dynamic or complex environments where modeling transitions can be impractical or impossible [55]. DRL-based IDS is a feasible scheme for the dynamic IoT environment.

2.4 DRL-Based IDS for IoT

This study focuses on IDS using DRL algorithms for IoT scenarios, with notable examples provided.

2.4.1 Deep Q-Networks (DQN)

DeepMind developed DQN in 2015, used to solve Atari games [57]. It combines Q-learning with deep neural networks (DNN) and experience replay [57]. DQN estimates a complex, nonlinear Q-value function using a neural network, as shown in Eq. (2).

$$Y_k^Q = r + \gamma \max_{\acute{a} \in A} Q(\acute{s}, \acute{a}; \bar{\theta}_k) \tag{2}$$

The max operator in Eq. (2) is the source of DQN's defining limitation: because the same network both selects and evaluates the greedy action, estimation noise is systematically propagated as an upward bias. In IoT intrusion detection this is not a benign artifact. Under the heavy class imbalance typical of IoT traffic datasets, the rare attack classes generate the noisiest Q-estimates, so the max-induced overestimation falls disproportionately on minority intrusions, inflating their apparent value and encouraging the agent to misclassify normal traffic as attacks, raising the false-positive rate precisely where detection matters most. Compounding this, the experience-replay buffer and separate target network that stabilize training each demand persistent memory proportional to buffer size and model duplication, a cost that conflicts directly with the kilobyte-scale RAM budgets of edge IoT nodes. DQN's design thus offers a natural fit to the discrete normal-vs.-malicious action space, but its stability mechanisms are the very features that impede on-device deployment.

2.4.2 Double Deep Q-Networks (DDQN)

The Double DQN (DDQN) function in Q-learning uses a similar amount of Q-value to identify actions that cause overestimated values and upward bias [58]. The target value function in DDQN is described as Eq. (3).

$$Y_k^{DDQN} = r + \gamma Q\left(\acute{s}, \arg\max_{a \in A} Q(\acute{s}, a; \theta_k); \bar{\theta}_k\right) \tag{3}$$

By substituting the single max operator with a decoupled estimator, in which the online network selects the action and the target network evaluates it, Eq. (3) removes the structural source of DQN's overestimation. For IoT IDS this decoupling is most consequential in multi-class settings: the minority-class value inflation that drives DQN's false positives is suppressed, yielding more trustworthy detection of rare attacks such as U2R or R2L without additional inference-time cost, since the second network already exists for stabilization. The benefit is therefore obtained essentially for free in memory terms relative to DQN. However, the correction is purely statistical; DDQN inherits the same replay-buffer footprint and the same discrete-action assumption, leaving it confined to classification-style IDS and equally exposed to the high-dimensional state spaces produced by raw IoT packet features.

2.4.3 Multiagent Deep Deterministic Policy Gradient (MADDPG)

Multiagent Deep Deterministic Policy Gradient (MADDPG) is a variant of the DDPG algorithm designed for multi-agent environments [59]. It allows multiple agents to learn optimal policies in a decentralized and cooperative manner [60]. In IoT intrusion detection, MADDPG is particularly suited for distributed network architectures, such as Software-Defined Networking (SDN)-based IoT environments, where multiple detection agents must coordinate to identify distributed attacks such as DDoS. Its centralized training with decentralized execution paradigm enables agents to leverage global network state during training while operating independently at inference time. However, MADDPG suffers from scalability limitations as the number of agents increases, and its convergence becomes unstable in highly dynamic IoT environments with non-stationary traffic patterns.

The mathematical signature of MADDPG is its centralized critic, which conditions on the joint observations and actions of all agents during training while each actor retains only local information at execution. This structure maps elegantly onto distributed SDN-IoT detection, where per-switch agents can be trained with global network visibility yet deployed to act on local flow statistics alone, an arrangement well-suited to detecting coordinated, distributed attacks such as DDoS. The same joint-action conditioning, however, is the root of two IoT-specific weaknesses. First, the critic's input dimension grows with the number of agents, so the training cost scales poorly across the many nodes of a realistic IoT fabric. Second, because every agent updates concurrently, each agent perceives a non-stationary environment induced by the shifting policies of its peers, a problem that is amplified by the already non-stationary nature of live IoT traffic and that manifests as unstable or oscillating convergence. MADDPG therefore trades deployability of the individual actor against a training process whose stability degrades exactly as the IoT deployment scales.

2.4.4 Advantage Actor-Critic (A2C)

The Actor-Critic (AC) architecture involves updating a policy distribution using gradients, while the critic estimates the value function for the current policy [61], as shown in Eq. (4).

$$\nabla_\omega V^{\pi_\omega}(s_0) = \mathbb{E}_{s\sim\rho^{\pi_\beta}, a\sim\pi_\beta}\left[\nabla_\omega \left(\log \pi_\omega(s,a) Q^{\pi_\omega}(s,a)\right)\right] \tag{4}$$

The defining mechanism in Eq. (4) is the advantage baseline, which rescales each action's value by the state's expected value and thereby reduces the variance of the policy-gradient estimate. For IoT IDS this variance reduction is a genuine design advantage: A2C is on-policy and so dispenses with the replay buffer entirely, giving it a markedly lighter memory footprint than DQN-family methods and making it the more edge-amenable of the value- and policy-based options. Its synchronous parallel updates further accelerate training when multiple simulated traffic environments are available. The same baseline, however, becomes a liability under the sparse and imbalanced reward signals of IoT intrusion data: when attack events are rare, the advantage estimate for minority classes is computed from few samples and inherits high bias, which can suppress the policy's sensitivity to exactly the infrequent intrusions an IDS must catch. A2C thus trades the memory cost of replay for a dependence on reward density that IoT traffic distributions rarely satisfy.

2.4.5 Deep Deterministic Policy Gradient (DDPG)

The deep deterministic policy gradient (DDPG) algorithm is a model-free, actor-critic approach for continuous action spaces [62], consisting of four neural networks: a policy network, a Q network Q, a target Q network Q^{tar}, and a target policy network. DDPG soft updates target network parameters θ_q, θ_π, θ'_q, and θ'_π [63].

$$\text{Actor: } \theta_{\pi'} \leftarrow \alpha\theta_{\pi} + (1-\alpha)\theta_{\pi'} \tag{5}$$

$$\text{Critic: } \theta_{q'} \leftarrow \alpha\theta_{q} + (1-\alpha)\theta_{q'} \tag{6}$$

For IoT IDS, DDPG's continuous action space enables non-classification formulations such as adaptive detection thresholds and trust scoring. However, its four-network architecture and well-documented sensitivity to hyperparameters raise both memory footprint and tuning cost, and its convergence is unstable under the non-stationary traffic conditions typical of live IoT deployments, limiting its suitability for direct edge deployment.

2.4.6 Twin Delayed Deep Deterministic Policy Gradient (TD3)

Twin Delayed Deep Deterministic Policy Gradients (TD3) was designed as a successor to DDPG [64,65], addressing the overestimation bias in the critic network through clipped double Q-learning, as shown in Eq. (7).

$$\mathbb{E}_{\tau\sim d_{\pi_\theta}}\left[\sum_{t=1}^{T} r(s_t,a_t) + \gamma \min\left(Q_1(s_t,a_t+\epsilon), Q_2(s_t,a_t+\epsilon)\right)\sum_{t=1}^{T}\nabla_\theta \log \pi_\theta(s_t,a_t)\right] \tag{7}$$

where $\epsilon \sim \text{clip}(N(0,\sigma,-c,c))$, σ is the standard deviation and c is a constant for clipping. TD3 improves upon DDPG by employing delayed policy updates and target policy smoothing, which collectively reduce variance and prevent premature convergence [66,67]. In IoT IDS applications, these properties translate to more reliable detection performance across diverse traffic conditions. Nevertheless, TD3's increased architectural complexity relative to DDPG incurs higher computational costs, which may limit its feasibility on edge IoT devices with stringent resource constraints.

2.4.7 Proximal Policy Optimization (PPO)

The Proximal Policy Optimization (PPO) algorithm [68] addresses the overhead of Trust Region Policy Optimization (TRPO) by incorporating the constraint into the objective function as a penalty, enabling simple stochastic gradient descent optimization, as shown in Eq. (8).

$$\mathop{\mathbb{E}}_{\tau\sim\pi_\theta}\left[\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)}\hat{A}_t\right] - C\cdot\overline{KL}_{\pi_{\theta_{\text{old}}}}(\pi_\theta) \tag{8}$$

PPO's clipped surrogate objective prevents excessively large policy updates, making it more stable than earlier policy gradient methods. In IoT intrusion detection, this stability is advantageous when training on noisy or non-stationary network traffic data. However, the selection of the clipping coefficient C remains non-trivial, and suboptimal values can result in either overly conservative updates or unstable policies, both of which reduce detection effectiveness in dynamic IoT attack scenarios [65].

2.4.8 Proximal Policy Optimization 2 (PPO2)

Proximal Policy Optimization 2 (PPO2) is an enhanced and computationally optimized implementation of PPO, available through the OpenAI Baselines library [69]. PPO2 introduces vectorized environment support and GPU-accelerated training, substantially reducing wall-clock training time. In the context of IoT IDS, PPO2's efficiency gains are significant, as the high-volume and high-dimensional nature of IoT network traffic demands scalable training pipelines. Nonetheless, PPO2 inherits PPO's sensitivity to hyperparameter

tuning, and its performance in detecting low-frequency or zero-day IoT attacks remains an open challenge, particularly when training data exhibit severe class imbalance.

2.5 Comparative Theoretical Analysis of DRL Algorithms for IoT IDS

While the preceding subsections describe each algorithm individually, a deeper theoretical comparison is necessary to understand their relative suitability for IoT intrusion detection. Table 2 contrasts the reviewed algorithm families along five dimensions that directly govern their practical viability in IoT environments: convergence stability, sample efficiency, the exploration exploitation tradeoff, adaptation to non-stationary traffic, and edge-deployment feasibility.

Table 2: Comparative theoretical analysis of DRL algorithm families for IoT intrusion detection.

Algorithm	Convergence Stability	Sample Efficiency	Exploration–Exploitation	Non-Stationary Adaptation	Edge Feasibility
DQN	Moderate; prone to Q-value overestimation	Low; large replay buffer needed	ϵ-greedy; coarse, slow to anneal	Weak; static target network lags drift	Moderate; replay buffer raises memory cost
DDQN	Higher than DQN; decoupled selection/ evaluation	Low; inherits replay overhead	ϵ-greedy; same as DQN	Weak–moderate; reduced overestimation aids stability	Moderate; similar footprint to DQN
A2C	Moderate; advantage reduces variance	Moderate; on-policy, no replay	Entropy-regularized stochastic policy	Moderate; synchronous updates track change	Higher; no replay buffer, lighter memory
PPO/PPO2	High; clipped objective bounds updates	Moderate; on-policy sample reuse limited	Stochastic policy with entropy bonus	Moderate–strong; stable under noisy traffic	Moderate; PPO2 GPU-optimized but compute-heavy
DDPG	Low; sensitive to hyperparameters	Moderate; off-policy replay	Action-noise exploration in continuous space	Weak; unstable under non-stationarity	Low; four networks raise compute/ memory
TD3	High; twin critics + delayed updates	Moderate; off-policy	Target-policy smoothing exploration	Moderate; variance reduction aids robustness	Low; higher complexity than DDPG
MADDPG	Low–moderate; degrades as agents scale	Low; multi-agent joint training costly	Per-agent noise; coordination-dependent	Weak; non-stationary from co-adapting agents	Low; centralised training resource-intensive

A clear tradeoff emerges between training stability and deployment cost. Value-based methods (DQN, DDQN) dominate the reviewed literature owing to their conceptual simplicity and natural fit to the discrete classification actions of intrusion detection [57,70]; however, their reliance on experience-replay buffers imposes a memory burden that conflicts with the constrained resources of edge IoT nodes, and their ϵ-greedy exploration adapts slowly to the rapidly shifting traffic distributions characteristic of IoT networks.

The decoupling of action selection from evaluation in DDQN mitigates the Q-value overestimation of DQN and yields more stable convergence [58], though it inherits the same replay-driven sample inefficiency. Policy-gradient methods with trust-region or clipping mechanisms (PPO, PPO2) offer the strongest convergence stability and the most graceful behaviour under noisy, non-stationary traffic [68,69], but their on-policy nature limits sample reuse, which is problematic when labelled IoT attack data are scarce. The A2C advantage formulation reduces the high variance of pure policy gradients and supports synchronous parallel updates, partially improving adaptation to changing traffic [61]. Continuous-control algorithms (DDPG, TD3) extend DRL to non-classification IDS tasks such as adaptive thresholding [62,63], yet their multi-network architectures and hyperparameter sensitivity make them the least feasible for direct edge deployment; TD3's twin-critic design and delayed policy updates improve stability at the cost of further computational overhead [64]. Multi-agent approaches (MADDPG) are uniquely suited to distributed SDN-IoT detection but suffer from a fundamental non-stationarity problem, as each agent's policy shifts the effective environment of the others, degrading convergence as the agent population grows [59,60]. Critically, none of these properties is captured by the accuracy-centric evaluation prevalent in the reviewed studies; the exploration–exploitation balance, adaptation to concept drift, and on-device resource cost remain largely unquantified, reinforcing the evaluation gap identified in Section 5 and the need for deployment-aware benchmarking discussed in Section 6.

2.6 Feature Selection

Intrusion detection relies on monitoring data exchange between nodes and using a dataset of records representing different objects. Each case is defined by attributes or features that quantify its properties. The detection model development process uses these features to differentiate between normal behaviors and anomalies [30]. However, using every feature from a dataset can lead to suboptimal results due to redundancy and irrelevant features. Therefore, careful feature selection is crucial for better generalization capabilities [17]. Three main categories of feature selection methods include: Filter-based methods use rating system scores to identify important features, offering fast implementation and versatility. Examples include information gain, correlation coefficient, chi-square test, and Fisher score, which use a rating system score to distinguish irrelevant ones [30]. Wrapper-based methods optimize the learning algorithm to select the optimal subset of features, but the increasing number of features presents a challenge as the selection process requires continuous evaluation of the algorithm's performance [17,30]. Search and genetic algorithms are commonly used to implement this method. Embedded methods combine wrapper and filter-based techniques with built-in feature selection strategies like XGBoost, Random Forest, and Decision Trees, reducing computational time for subset dataset classification [30].

3 Related Works

Reinforcement Learning (RL) is a machine learning paradigm where an agent maximizes rewards by interacting with its environment [16]. Inspired by human brain functions, RL shows promise in decision-making [14,15]. However, traditional RL struggles with scalability, facing issues like the curse of dimensionality and slow convergence, limiting its use in complex environments [30]. Deep Reinforcement Learning (DRL) merges Deep Learning's feature extraction with RL's decision-making, overcoming these limitations. DRL is ideal for high-dimensional inputs and complex tasks [17], prompting research into its applications in decision-making, control systems, and IoT security. There is a lack of comprehensive reviews on DRL-based IDSs for IoT networks. Although several studies exist, they often provide limited discussion of the methodological limitations, evaluation metrics, and dataset coverage in this field. This section reviews existing surveys to understand the current state of DRL-based IDS research and identify remaining gaps. One

review examined DRL-based IDSs for IoT networks, focusing on detection methods and effectiveness [29]. It highlighted key challenges, including IoT device heterogeneity and resource constraints, which increase exposure to advanced attacks such as zero-day exploits and adversarial machine learning tactics. The review also emphasized the need for adaptable IDSs against evolving threats, covering DRL-based approaches alongside conventional ML and DL methods. A study examined integrating RL with DL to improve IDS in IoT [18]. It reviewed DRL-based security for cyber-physical systems and intrusion detection, suggesting realistic environments for DRL. Combining host and network-based IDSs was recommended. The study focused on broader applications beyond IoT, limiting details on DRL-based IDSs for IoT. A comprehensive survey explored anomaly detection in IoT security using ML and DL [71], addressing evolving threats and the need for adaptive solutions, but offered limited insights into DRL for IoT anomaly detection. A survey of RL applications in cybersecurity focused on IDS, IPS, IoT, and IAM, identifying key datasets such as NSL-KDD and CICIDS used in RL-based IDS from 2010 to 2021 [24]. It lacked a focus on DRL-based IDSs for IoT. A systematic review examined DRL applications for anomaly detection from 2017 to 2022 [25], showing DRL's strengths in complex anomaly detection, including IoT, but lacked IoT security specifics. A survey of DRL approaches in IoT intrusion detection highlighted DRL's role in addressing IoT challenges [26]. The study covered lessons, best practices, and future directions but lacked systematic evaluation of metrics and datasets. Another review analyzed IDSs in IoT, covering detection methods, attack types, and public datasets, but discussed DRL only as part of a broader IDS review without a dedicated systematic analysis of DRL-based approaches [28]. A further study reviewed machine learning-based IDSs for IoT, categorizing solutions into supervised learning, DL, and federated learning, but included fewer than five DRL contributions, providing limited insights [27].

Complementary recent work outside the DRL paradigm further underscores the importance of feature engineering and class-imbalance handling in IoT intrusion detection. Ref. [72] proposed real-time IoT IDS frameworks based on Kolmogorov–Arnold Networks combined with SMOTE oversampling, reporting high multi-class detection accuracy on the RT-IoT2022 dataset and highlighting the value of imbalance correction for minority attack classes. Ref. [73] introduced a hybrid ANOVA–Recursive Feature Elimination (ANOVA-RFE) feature-selection model coupled with machine and deep learning classifiers, achieving strong IoT and IIoT attack detection using only a small subset of selected features. While these approaches do not employ DRL, their emphasis on systematic feature selection and imbalance mitigation reinforces the feature-engineering gap identified in this review for DRL-based IDS models.

Further recent contributions have advanced IoT and IIoT intrusion detection through zero-day attack detection using optimisation-driven deep learning [74], deep ensemble methods for IoMT security [75], machine learning for SDN-based intrusion detection [76], and comparative dataset analyses across ToN-IoT and UNSW-NB15 [77]. While these works do not employ DRL, they reinforce the methodological themes of feature engineering, imbalance handling, and deployment efficiency examined in this review. Current literature on DRL-based IDSs for IoT exhibits five identifiable and recurring gaps that the present review explicitly addresses. First, existing surveys overlook IoT-specific challenges such as resource constraints, heterogeneous device types, and protocol diversity, treating IoT environments as generic network settings. Second, no prior review provides a standardized quantitative synthesis of evaluation metrics, leaving researchers without a comparative performance baseline across studies. Third, the role of feature engineering in DRL-based IDS development has received negligible systematic attention, despite its direct impact on model efficiency and deployability on constrained IoT devices. Fourth, limited IoT-representative datasets impede meaningful benchmarking and restrict the generalizability of proposed models to real-world deployments. Fifth, advanced DRL paradigms such as multi-agent DRL and hybrid DRL architectures

remain underexplored relative to their potential for addressing the distributed and dynamic nature of IoT threat landscapes.

DQN was the most frequently employed algorithm, appearing in 8 studies (30.8%), followed by Double DQN in 5 studies (19.2%), hybrid methods in 3 studies (11.5%), and MADDPG, DL + Q-Learning, PPO2, and DDPG each in 2 studies (7.7%), with TD3 and Federated DRL each appearing in 1 study (3.8%). Regarding feature selection, only 3 studies (11.5%) explicitly applied feature engineering (two using LightGBM and one using MIFS), 2 studies (7.7%) referred to feature selection without specifying the method, and the remaining 21 studies (80.8%) reported no feature selection procedure. In terms of datasets, NSL-KDD was the most widely used, appearing in 7 studies (26.9%), followed by TON-IoT in 4 studies (15.4%). With respect to evaluation, accuracy was reported in 23 studies (88.5%), while network performance metrics such as energy consumption and latency were assessed in only 6 studies (23.1%), revealing a critical under representation of deployment-relevant metrics in the existing literature. These quantitative observations are elaborated in the analysis presented in Section 5, and collectively demonstrate that the present review provides a more systematic, comprehensive, and empirically grounded contribution than prior surveys in this domain.

Table 3 presents a comparative analysis of the review studies in the literature.

Table 3: Comparative analysis of the review studies presented in the literature.

Study	Advantages	Limitations	Findings
[29]	A comprehensive review of DRL-based IDs for IoT, taxonomy, and approaches.	Lacks critical analysis of DRL-based IDs for IoT systems; also includes non-DRL IDSs, deviating from the IoT-specific focus.	Proposes DRL-based IDS for IoT, emphasizing adversarial robustness.
[18]	Explores DRL methods for cybersecurity, including intrusion detection and defense strategies.	Overlooks key aspects of DRL-based IDSs for IoT network security, including limitations, datasets, and applications.	Recommends realistic environments for DRL in intrusion detection and combining host- and network-based IDSs.
[24]	Reviews RL in cybersecurity for IDS, IPS, IoT, IAM.	Offers limited discussion on DL-RL integration, neglecting feature selection and optimization.	Emphasizes exploring DRL to reduce computation and convergence times.
[25]	The first systematic literature review on the employment of DRL for anomaly detection.	Examines DRL frameworks for anomaly detection but overlooks limitations in DRL-based techniques.	Encourages research on advanced DRL techniques, alternative datasets, and novel applications.
[26]	Reviews DRL applications in IoT-IDS, trends, challenges, lessons, and benchmarking datasets.	Fails to classify DRL-based IDS for IoT, assess performance metrics, and include relevant studies.	Proposes leveraging threat intelligence and recent datasets to enhance DRL-based IDSs.

(Continued)

Table 3 (continued)

Study	Advantages	Limitations	Findings
[28]	Critically reviews IDS approaches for IoT, including machine learning, deep learning, reinforcement learning, and DRL-based methods.	Discusses DRL only as part of a broader IDS review and does not provide a dedicated systematic analysis of DRL algorithms, feature engineering, or network performance metrics.	Emphasizes the importance of intelligent IDS models and highlights dataset and evaluation challenges in IoT intrusion detection.
[27]	Discusses the use of supervised and deep learning techniques for IoT intrusion detection.	Limited focus on federated learning and reinforcement learning techniques.	Supervised learning approaches demonstrate high accuracy; emerging technologies offer potential.
[33]	Recommends Federated Learning for IoT anomaly detection.	Lacks focus on advanced FL aggregation methods and diverse, realistic IoT datasets.	Suggests using advanced datasets and enhancing FL tool documentation.
[30]	Reviews intra-vehicle IDSs, categorizing them as flow-based, payload-based, or hybrid.	Emphasizes technical categorizations but lacks broader dataset evaluations.	Emphasizes CAN vulnerabilities and the need for lightweight, real-time IDSs.

4 Research Methodology

This study adopts the review methodology adopted in [78], combining PRISMA guidelines [79–81] with the systematic mapping process employed in [82]. The methodology consists of seven main stages, as shown in Fig. 4: preliminary study, research question formulation, search criteria identification, literature search and screening, eligibility and quality assessment, data extraction and compilation, and final article selection with results reporting.

The completed PRISMA 2020 checklists are provided as separate supplementary materials.

4.1 First Stage: Preliminary Studies

The preliminary study focuses on IDSs using DRL for IoT environments. Popular internet search engines are searched using the domain-specific keyword "deep reinforcement learning-based intrusion detection systems in IoT." Relevant search terms are selected from the results, allowing for a comprehensive review of the available studies.

4.2 Second Stage: Research Questions

This review focuses on the development of DRL-based IDSs for IoT environments, examining their advantages and limitations. It also investigates feature selection algorithms used to reduce dimensionality, the types of attacks these technologies detect, the datasets used along with their advantages and limitations, and the evaluation metrics commonly employed to assess DRL algorithm performance in IDS. Table 4 outlines the main research questions.

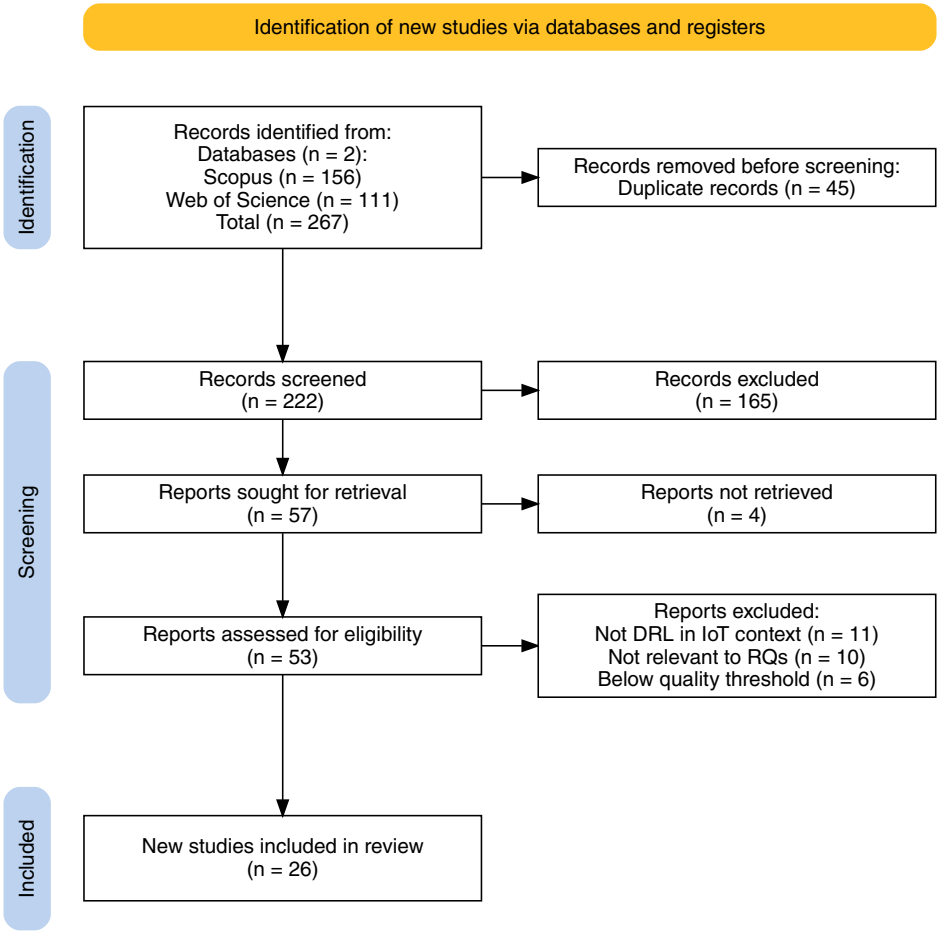

Figure 4: PRISMA flow diagram illustrating the study selection process for the systematic literature review.

Table 4: The fundamental research questions.

Question	Motivation
RQ1: Which DRL algorithms are used in developing IDSs to improve the security of IoT networks?	The study aims to assess the effectiveness of DRL algorithms in creating IDSs for IoT network security, identifying their strengths, weaknesses, and suitability.
RQ2: What feature selection methods are applied in conjunction with DRL-based IDSs?	The study explores the effectiveness of feature selection methods in enhancing the efficiency, effectiveness, and interpretability of DRL-based IDSs for securing IoT networks.
RQ3: What types of attacks in IoT were these technologies designed to detect?	The study aims to categorize and analyze IoT attacks that used DRL technologies in IDSs, providing a comprehensive understanding of threats mitigated by these systems.
RQ4: What datasets are widely used with DRL-based IDS, their advantages, and limitations?	This study examines datasets used with DRL-based IDSs, highlighting their advantages and limitations for training and evaluating DRL-based IDSs in IoT security.
RQ5: What evaluation metrics are commonly used to assess the performance of DRL-based IDSs for IoT networks?	The study aims to investigate the evaluation metrics used to evaluate the performance of DRL algorithms in IDSs, focusing on detection accuracy, false positives, false negatives, and other relevant metrics.

4.3 Third Stage: Search Criteria

This study performed an initial search on DRL-based IDSs for IoT contexts using Google Scholar. Relevant keywords and search terms were then found using the research questions. Using the finalized keywords, two scientific databases and locations were chosen for a more thorough search. The search parameters utilized are shown in Table 5, and the keywords and corresponding search strings applied across the databases are listed in Table 6.

Table 5: Identified search criteria.

SN	Database	Search Scheme	Role	Publications
1	Scopus	Title, abstract, keywords	Systematic search	Journal papers, conference proceedings, book chapters, technical reports
2	Web of Science	Title, abstract, keywords	Systematic search	
3	Google Scholar	Full text	Preliminary scoping only	

Table 6: Keywords and search strings.

Domain	Keywords	Search Strings
DRL-based intrusion detection in IoT	Deep reinforcement learning intrusion detection system IoT, Deep reinforcement learning attack detection IoT	(“intrusion detection” AND “deep reinforcement learning” AND (“IoT” OR “internet of things”))

4.4 Fourth Stage: Inclusion and Exclusion Criteria (Screening Process)

To choose relevant papers for the study topics, the database was searched using pre-identified keywords and search strings, and all the retrieved articles were filtered according to the inclusion and exclusion criteria. Table 7 lists the criteria. Duplicate removal and title–abstract screening were performed using the Rayyan systematic review platform, while Mendeley was used for reference management and full-text organisation throughout the screening process.

It should be noted that the initial open access restriction has been revised in the present version of this review. Subscription-based articles accessible through institutional repositories, author-shared preprints, or interlibrary access were subsequently included to mitigate publication access bias and ensure comprehensiveness. Consequently, four additional high-quality studies were incorporated, including a DQN-based open-set intrusion detection solution for Industrial IoT [83], a DQN-based heuristic intrusion detection approach against edge-based SIoT zero-day attacks [84], a double deep Q-network strategy with stochastic games for IIoT intrusion detection [85], and a multi-teacher knowledge distillation framework for privacy-aware unlearning in IIoT intrusion detection [86].

Table 7: Description of inclusion and exclusion criteria.

Criteria	Inclusion Criteria	Exclusion Criteria
Publication year	2020–2026	Not between 2020–2026
Language	English	Non-English
Publication Type	Journal articles, conference proceedings, book chapters, preprint servers	Review articles, seminars, letters, tutorials, interviews, blogs, posters
Availability	The study is available in full text	Unavailable studies in full text
Technique	Explicitly mentioned the employed DRL techniques	Did not mention the employed DRL techniques explicitly
Relevance	Used DRL techniques in detecting attacks in IoT and important to the RQs	Did not use DRL techniques in detecting attacks in IoT
Number of pages	Articles more than 5 pages	Articles less than 5 pages
Article Duplication	Non-duplicated articles	Duplicated articles
Access Type	Open access and subscription-based articles accessible through institutional repositories or author-shared manuscripts	Articles entirely unavailable through any access channel

4.5 Fifth Stage: Eligibility and Quality Assessment

Three quality rating scores adapted from [87] were used in the survey to evaluate the eligibility and applicability of the primary studies. The grading system is 0 for "no," 0.5 for "partially," and 1 for "yes." Table 8 lists the quality evaluation results.

Table 8: Scoring-based eligibility criteria for assessing the quality of each paper.

Criteria	Score	Description
Clearly defined objectives?	1	Yes, articles with clear objectives and goals
	0.5	Articles with clear objectives and unclear goals
	0	No, articles with unclear objectives
Clearly detailed methodology?	1	Yes, comprehensive and well-structured methodology
	0.5	Unclear methodology that requires further details
	0	No, fails to provide a clear methodology
Declared limitations?	1	Yes, the article has clear limitations
	0.5	The limitations are not fully described
	0	No, the article does not mention any limitations
Clear research findings?	1	Clear findings with suitable visualizations
	0.5	Findings without clear statements or visualizations
	0	Research findings are unclear

Fig. 5 visually summarizes the cross-study accuracy comparison on the three most frequently shared benchmark datasets. Among the seven studies employing NSL-KDD, reported accuracy values range from 82.4% [88] to 99.10% [70], a variance of approximately 17 percentage points that cannot be attributed solely to algorithmic differences. Similarly, among the four studies using TON-IoT, accuracy values range from 93.4% [89] to 98.0% [90]. The figure underscores that dataset partitioning and preprocessing choices influence reported outcomes as significantly as the DRL algorithm itself, reinforcing the need for standardised benchmarking protocols identified in Section 6.

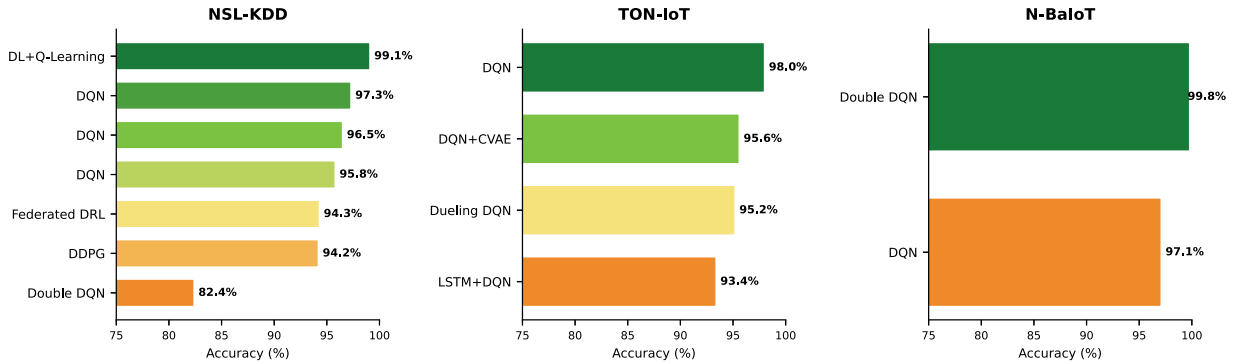

Figure 5: Cross-study comparison of accuracy values reported by DRL-based IDS models evaluated on shared benchmark datasets (NSL-KDD, TON-IoT, and N-BaIoT).

4.6 Sixth Stage: Data Extraction and Compilation

The previous sections were used to classify the studies that were included in the review. Data extraction and compilation were carried out with the aid of Google Sheets, REDCap, and Microsoft Excel. The extracted data comprised the following: title, authors, year of publication, abstract, keywords, scientific databases/location, type of publication, algorithms utilized, and study type.

Fig. 6 shows the distribution of the reviewed articles by publication year.

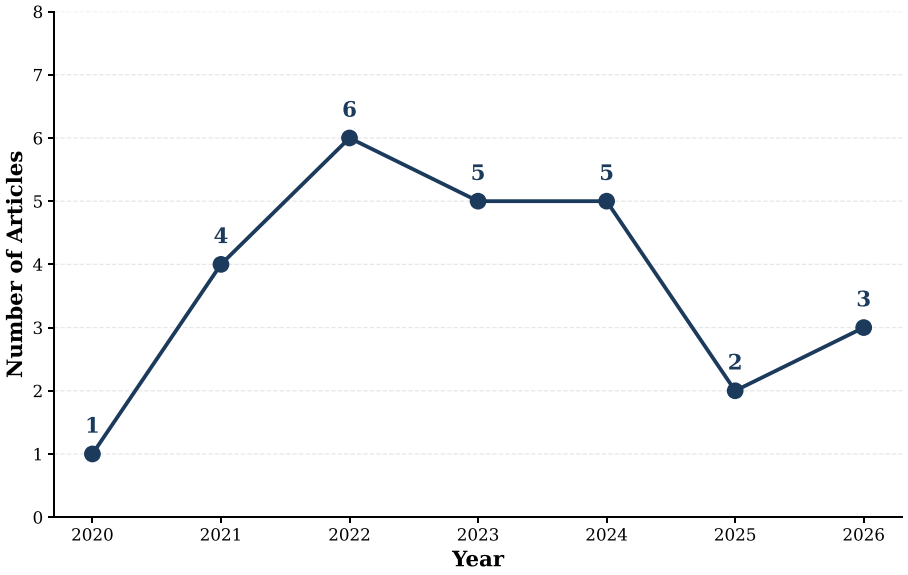

Figure 6: Number of articles per year.

4.7 Seventh Stage: Final Findings and Results

Following PRISMA principles, the updated research mapping procedure identified 267 studies from 2020 to 2026 by searching scientific databases such as Web of Science and Scopus with keywords and search strings. Using inclusion and exclusion criteria, screenings were carried out. With a preference for journal papers, conference proceedings, and book chapters relevant to DRL-based IDSs for IoT networks, 26 relevant studies were ultimately included. A total of 241 studies were disqualified for not meeting the requirements for inclusion. Fig. 7 provides a statistical presentation of the selected studies.

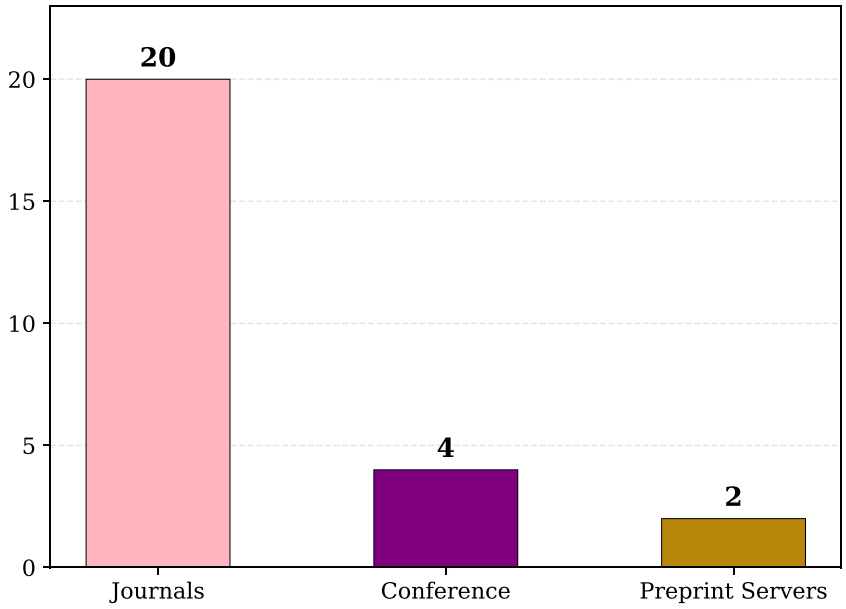

Figure 7: Studies included categorized by publication types.

5 Analysis and Synthesis of Data

This study explores the use of DRL-based IDS for securing IoT networks. It discusses the advantages and limitations of DRL algorithms, feature selection methods, attacks detected using DRL algorithms, datasets used for evaluation, and metrics for performance assessment. The data suggest that DRL use in IDS development is in its early stages, with researchers starting to focus on this direction in 2020. Fig. 6 shows the selected articles for this review from 2020 to 2026.

To provide a consolidated overview before the detailed question-by-question analysis, Fig. 8 presents a structured taxonomy of the key findings derived from the reviewed literature. The taxonomy is organized along the five research-question dimensions, namely the DRL algorithms employed, the feature selection methods applied, the categories of IoT attacks addressed, the datasets used for evaluation, and the evaluation metrics reported. Each dimension summarizes its principal findings, which converge into the overall conclusions of the review and the open research gaps discussed in Section 6. The following subsections elaborate on each dimension in turn.

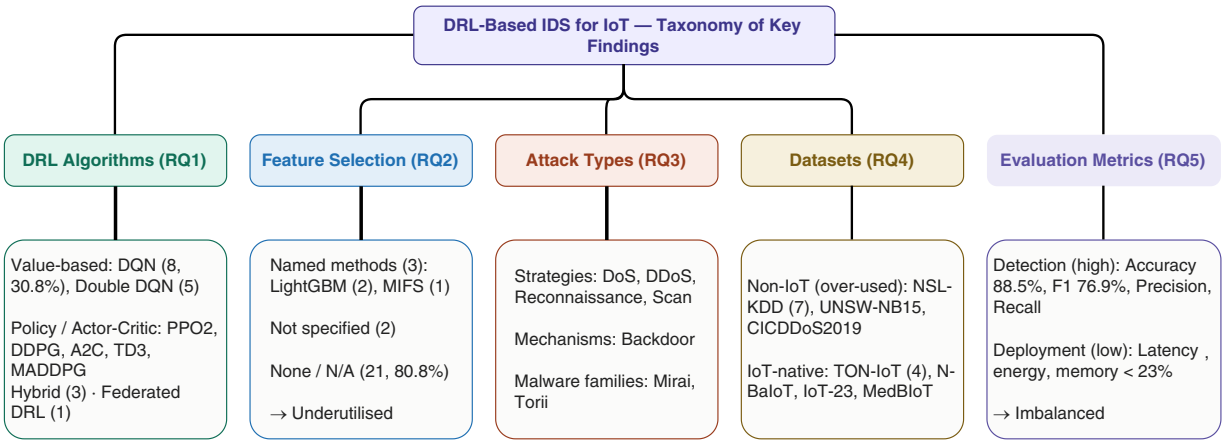

Figure 8: Structured taxonomy of the key findings of this review across the five research-question dimensions, the overall conclusion, and the identified open research gaps.

5.1 RQ1: Which DRL Algorithms Are Employed in the Development of IDS Models to Enhance the Security of IoT Networks?

This review explores techniques used in studies to develop IDS models for detecting attacks in IoT environments using DL, RL, DRL, and hybrid methods. Performance is assessed based on accuracy, precision, recall, and F1-measure. A detailed search results for the DRL algorithms is presented in Table 9. Note that Table 9 reports the raw per-database retrieval counts in Scopus and Web of Science prior to deduplication; the study counts and percentages discussed in the text refer to the 26 final included primary studies after screening, in which each study is counted exactly once.

Table 9: Search results for DRL algorithms used in IDS for IoT security.

DRL Algorithms	Scopus	Web of Science
DQN	7	5
Double DQN	4	3
Hybrid	3	2
MADDPG	2	1
DL + Q-Learning	2	1
PPO2	2	1
DDPG	2	1
TD3	1	1
Federated DRL	1	1

The utilization of DRL algorithms in IDS for IoT security remains limited, despite extensive research. Most DRL-based IDS models for IoT networks use the DQN algorithm, appearing in 8 studies (30.8%), followed by Double DQN in 5 studies (19.2%) and hybrid methods in 3 studies (11.5%). Two studies each utilized the MADDPG, DL + Q-Learning, PPO2, and DDPG algorithms, while TD3 and Federated DRL were each explored in one study for robust IDS tailored for IoT environments. Scopus has the highest count of publications, while WoS has the lowest number. Figs. 9 and 10 demonstrate the application and range of different DRL algorithms.

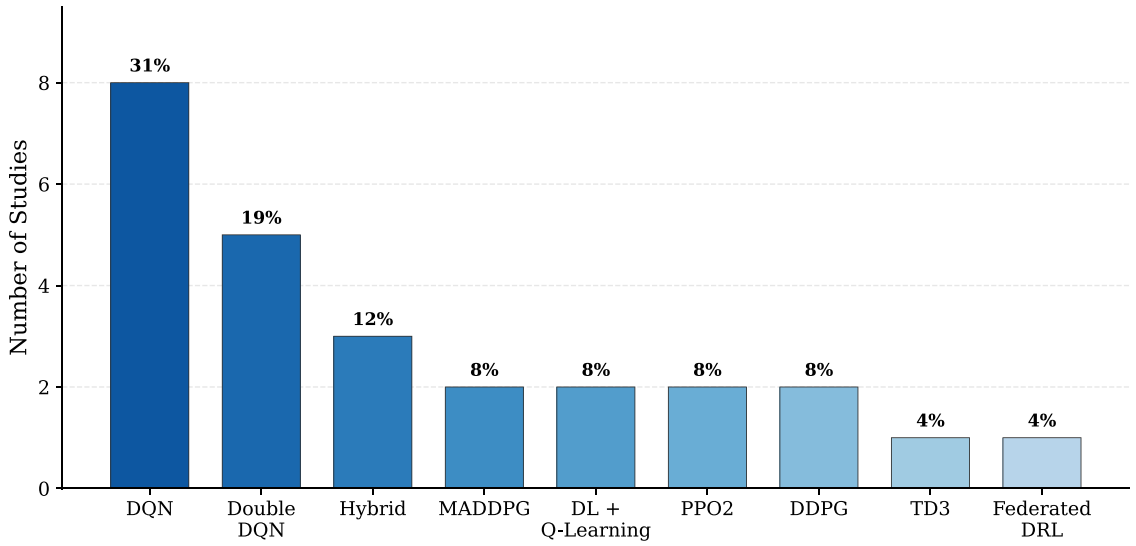

Figure 9: Distribution of the DRL algorithms used to develop IDS for IoT networks security.

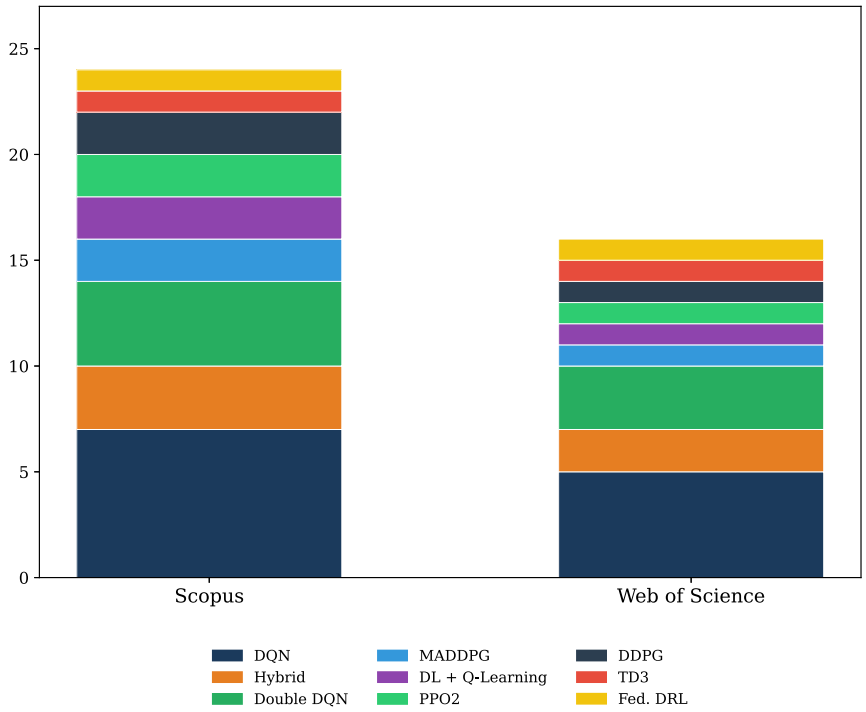

Figure 10: DRL algorithms used to develop IDS models to secure IoT networks.

5.1.1 Deep Q-Networks (DQN), Advantage Actor-Critic (A2C) and Proximal Policy Optimization (PPO)

DRL methods, such as DQN, A2C, and PPO, have been successfully applied to enhance IDSs. One study proposed an innovative IDS architecture designed to detect DoS and DDoS attacks generated by Cellular IoT (CIoT) botnets. This architecture integrated advanced attack simulations and defense mechanisms, including scenarios without attacks, attack plans inducing adversary actions, and two specific defense strategies: Label-based Semi-supervised Defense (LSD) and Clustering-based Semi-supervised Defense (CSD). The system was tested on three IoT datasets NSL-KDD, IoT-23, and NBaIoT, featuring botnets from compromised CIoT

devices. Results demonstrated that DRL-based algorithms significantly outperformed traditional machine learning and deep learning approaches in accuracy [91]. Another application of DQN focused on intrusion detection for IoT systems, utilizing the TON-IoT dataset. The study evaluated the model using metrics like accuracy, precision, recall, F1-score, false negatives (FN), false positives (FP), and geometric mean (G-Mean). Experimental results highlighted the DQN-based IDS's superior performance compared to benchmark algorithms [90]. To address the challenges posed by the complexity of IoT networks in the 5G era, a study [92] proposed a distributed DQN algorithm for IDSs. This approach enabled agents to learn continuously, improving their ability to detect normal and anomalous IoT behaviors. Evaluations conducted with the NSL-KDD dataset demonstrated the approach's superior performance in accuracy, F1-score, precision, and recall. In another advancement, research conducted by [83] combined DQN with Conditional Variational Auto-Encoders (CVAE) to develop DC-IDS, an open-set IDS for the IIoT. Experiments conducted on the TON-IoT dataset demonstrated the system's ability to balance true positive rates for unknown attacks with false positive rates for normal traffic. Finally, another study [93] employed DQN to secure Wireless Sensor Networks (WSNs) and IoT environments by monitoring and analyzing network activity in real-time. The system utilized the NSL-KDD dataset and achieved significant improvements in detection rate, accuracy, and precision.

Fig. 11 visualizes the stark disparity between detection-oriented and deployment-oriented evaluation practices. Data mining metrics such as accuracy (88%), F1-score (77%), and precision (73%) are reported by the majority of studies, well above the 50% threshold. In contrast, no network performance metric exceeds 23% coverage, with critical deployment indicators such as memory footprint (8%) and throughput (4%) reported by only one or two studies. This imbalance reveals a fundamental disconnect between academic evaluation and IoT operational realities, where nodes typically operate under strict constraints on processing power, memory, and battery life. Notably, only one reviewed study [94] validated its model on actual IoT hardware (Raspberry Pi 3), reporting inference latency, CPU utilization, and memory footprint under real operating conditions. Future research must mandate reporting of both detection performance and deployment feasibility metrics to enable credible assessment of real-world IoT viability.

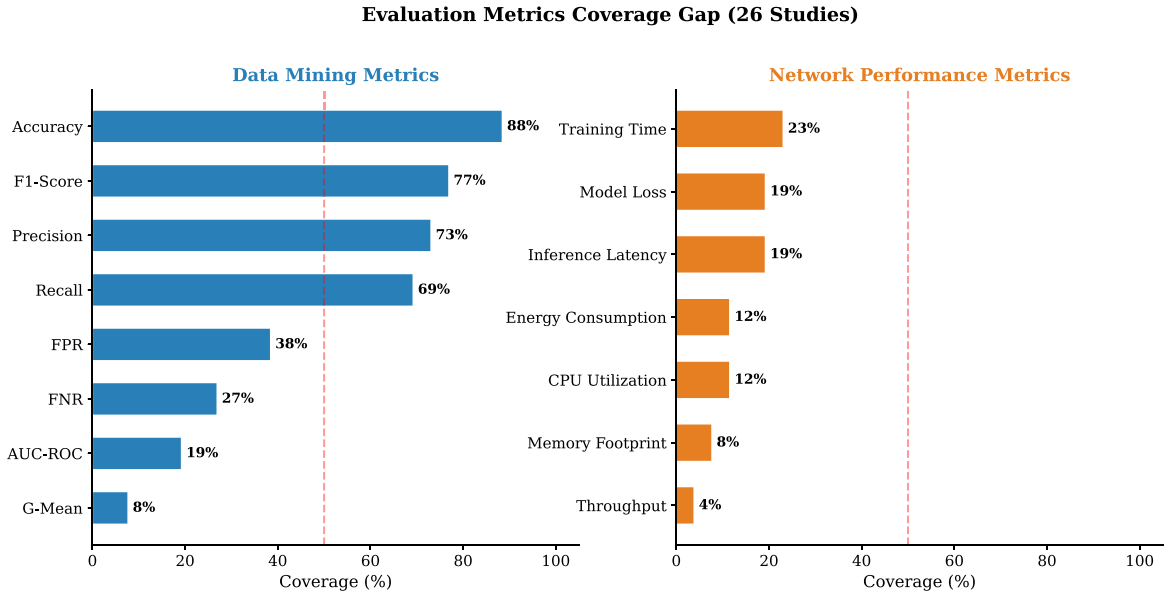

Figure 11: Evaluation metrics coverage gap across the 26 reviewed DRL-based IDS studies, contrasting data mining metric coverage (left) with network performance metric coverage (right). The dashed red line indicates the 50% threshold.

5.1.2 Hybrid Methods

A hybrid attack detection model combining LSTM networks and DQN was developed to improve anomaly detection in IoT networks. The model used the SIBRO algorithm to optimize LSTM weights. The model's performance was evaluated using UNSW-NB15 and TON-IOT datasets [89]. The authors in [95] developed a hierarchical Federated Learning (FL) methodology for anomaly detection in IIoT, enabling decentralized model training while preserving privacy. A trust metric-based anomaly detection framework was developed using a DDPG algorithm and belief networks (DDPG-BN). The model was reported by its authors to outperform the baseline models considered in that study on the NSL-KDD dataset, although its 94.20% accuracy is mid-range among the NSL-KDD studies compared in Table 10 [96].

Table 10: Cross-study comparison of accuracy on shared datasets among reviewed DRL-based IDS models.

Study	Dataset	Algorithm	Acc. (%)	F1	C1	C2	C3	Overfitting Risk
[70]	NSL-KDD	DL + Q-Learning	99.10	0.991	×	×	×	High†
[91]	NSL-KDD, N-BaIoT	DQN	97.30	0.971	×	×	✓	Moderate
[92]	NSL-KDD	DQN	96.50	0.963	×	✓	×	Moderate
[93]	NSL-KDD	DQN	95.80	0.956	×	×	×	High
[96]	NSL-KDD	DDPG	94.20	0.940	×	×	×	High
[88]	NSL-KDD	Double DQN	82.40	0.824	×	✓	×	Moderate
[90]	TON-IoT	DQN	98.00	0.978	×	×	×	High
[83]	TON-IoT	DQN + CVAE	95.60	0.954	×	✓	×	Moderate
[89]	TON-IoT, UNSW-NB15	LSTM + DQN	93.40	0.932	×	×	✓	Moderate
[98]	N-BaIoT, MedBIoT	Double DQN	99.80	0.998	×	×	✓	High†
[85]	IIoT sim.	Double DQN	96.80	0.965	×	✓	×	Moderate
[104]	NSL-KDD	Federated DRL	94.30	0.893	×	✓	×	Moderate
[23]	CICIoMT2024	Hybrid CNN-LSTM + DQN + PPO	99.58	0.995	×	×	×	High†
[94]	IoT-DH	Adaptive DQN	98.50	0.985	×	✓	✓	Low
[22]	CIC-IIoT2025	Dueling DQN	95.20	0.950	×	✓	✓	Low

Note: Accuracy values are as reported in the original studies. Overfitting risk is assigned by a fixed rule from three observable safeguards: C1 = cross-validation reported; C2 = independent, held-out, or cross-dataset test set used; C3 = evaluation on more than one dataset (✓ = present, × = absent). Score = C1 + C2 + C3: a score of 0 = High, 1 = Moderate, ≥2 = Low. †Override: a study reporting near-perfect accuracy (>99%) without cross-validation and without an independent external or cross-dataset validation protocol is classified as High risk regardless of score. The classification reflects validation-reporting completeness, not a directly measured generalization gap.

5.1.3 Double Deep Q-Networks (Double DQN)

A study by [88] developed an anomaly detection framework that combines adversarial environment reinforcement learning (AE-RL) and Synthetic Minority Over-sampling Technique (SMOTE) to address class imbalance issues. The framework uses the NSL-KDD dataset. The AESMOTE model demonstrated superior predictive performance with an accuracy of over 82% and an F1-score exceeding 0.824. A novel framework, DDQN-Prioritized Experience Replay (PER), has been developed to improve anomaly detection in smart environments. The framework demonstrated adaptability by learning directly from data. It outperformed traditional models in both tasks, achieving 92.6% accuracy with a 70.5% F1-score in fall detection (the comparatively low F1 reflecting the severe class imbalance of rare fall events), and 98.2% accuracy in occupancy detection, for which no F1-score was reported [97]. Another study introduced MalBoT-DRL, a DDQN-based technique for detecting malware botnets in IoT networks. Validated using MedBIoT and N-BaIoT datasets, MalBoT-DRL achieved average detection rates of 99.80% in the early phase and 99.40% in the late phase [98].

5.1.4 Proximal Policy Optimization 2 (PPO2)

An innovative DRL-based IDS was proposed to improve the security of the IIoT. The PPO2 algorithm was used to train an intelligent detection agent, while LightGBM was used for feature selection. The IDS achieved precise and efficient cyber threat detection, attaining an accuracy of 99.09% [99]. A study has developed a stability-oriented routing protocol for IIoT environments, enhancing SDN and improving trustworthiness, QoS, and power efficiency. The DRL-based Graphical Detection System showed superior results and achieved 95% accuracy [100].

5.1.5 Multiagent Deep Deterministic Policy Gradient (MADDPG)

A Multi-Agent Reinforcement Learning (MARL) architecture, using the MADDPG algorithm, has been proposed to detect and mitigate DDoS attacks in high-bandwidth SDN-IoT networks. The MADDPG-based architecture achieved faster convergence and superior performance compared to Single-Agent DDPG (SAD-DPG) [101]. A new MARL framework has been developed to optimize multipath routing and defend against malicious DDoS traffic in SDN-IoT environments. The proposed architecture significantly outperformed traditional methods [102].

5.1.6 DL and RL

A new mechanism has been proposed by [103] to improve the efficiency and robustness of IDS in IIoT environments by combining DRL and Generative Adversarial Networks (GAN). Experiments on the DS2OS dataset showed that the DRL-GAN approach outperformed standard DRL models. A study by [70] developed a DRL-based intrusion detection model that uses the Black Widow Optimization (BWO) algorithm. The model's performance was evaluated using the NSL-KDD dataset, achieving high precision, recall, F1-score, and accuracy.

5.1.7 Deep Deterministic Policy Gradient (DDPG)

Ref. [20] proposed an anomaly-based intrusion detection algorithm using DDPG to detect DDoS attacks in green IoT environments. The algorithm achieved accurate traffic prediction and intrusion detection while maintaining a low false detection rate.

5.1.8 Twin Delayed Deep Deterministic Policy Gradient (TD3)

Research by [21] developed a Transfer Learning-based Trajectory Anomaly Detection strategy (TLTAD) to improve anomaly detection in IoT-empowered Maritime Transportation Systems (IoT-MTS). TLTAD outperformed comparison models in anomaly detection accuracy and significantly reduced training times.

5.1.9 Federated Deep Reinforcement Learning

A recent study [104] proposed a federated reinforcement learning-driven multi-task optimization framework for robust and ethical edge IoT security. The framework distributes DRL-based intrusion detection across twelve heterogeneous edge nodes, achieving 94.3% accuracy and an overall F1-score of 89.3% on the NSL-KDD dataset, with explicit attention to improving the detection of rare User-to-Root (U2R) attacks. The federated architecture reduces communication overhead while preserving data privacy across distributed IoT deployments. However, the reliance on a single benchmark dataset limits the generalizability assessment of the proposed framework.

5.1.10 Recent Advances in DQN Variants (2024–2026)

Ref. [84] proposed DQN-HIDS, a heuristic learning intrusion detection system integrating DQN with LSTM for edge-based Social IoT (SIoT) networks. The system addresses zero-day attack detection under insufficient training samples by employing a reward and penalty mechanism that gradually improves the agent's ability to independently label new network traffic samples. Experimental results demonstrated advantages over state-of-the-art deep learning methods with fewer training samples.

Ref. [85] proposed a novel intrusion detection strategy combining stochastic games with Double DQN for IIoT environments. The interaction between attackers and detectors was modeled as dynamic adversarial stochastic games with incomplete information, and Nash equilibria were theoretically analyzed. The framework achieved effective detection while balancing accuracy and computational efficiency.

Ref. [22] presented D3O-IIoT, a Dueling Deep Q-Network agent for dynamic deception orchestration in IIoT security. The defense problem was formulated as a Markov Decision Process with a multi-objective reward balancing attack mitigation, deception engagement, false positive control, and resource cost. Experiments on three IIoT datasets (CIC-IIoT2025, WUSTL-IIoT2021, TON-IoT) demonstrated a 13.7% attack mitigation rate improvement over baselines with sub-3 ms decision latency.

Ref. [23] proposed HCLR-IDS, integrating CNN, LSTM, DQN, and PPO for IoMT healthcare network security. The system employed Enhanced Mutual Information Feature Selection (MIFS) to preprocess the CICIoMT2024 dataset, achieving binary classification accuracy of 99.58% and multi-class accuracy of 77.73% across 18 attack classes. Complementary work by [105] on deep learning-driven anomaly detection for IoMT-based smart healthcare systems further underscores the critical importance of robust intrusion detection in medical IoT environments.

Ref. [106] proposed IoT-ONDDQN, a detection model combining OneR feature selection with a Noisy Double DQN architecture for IoT data security. The model was evaluated on the CIC-IoT2023 dataset, which captures contemporary IoT malicious traffic patterns across multiple attack categories. The integration of noisy layers enabled improved exploration during training, while the OneR preprocessing reduced feature dimensionality. Experimental results demonstrated that IoT-ONDDQN outperformed standard DQN and DDQN baselines in detecting diverse IoT attack types.

Ref. [94] proposed an adaptive DRL-based DDoS detection framework specifically designed for resource-constrained IoT edge devices. The framework was evaluated on the IoT-DH dataset, a real-world multi-protocol IoT traffic dataset collected via a honeypot, achieving F1-scores of 0.98–0.99 on in-distribution data while maintaining low false positive rates. Cross-dataset evaluation demonstrated stable performance with F1-scores above 0.96 under distribution shifts. Notably, the model was implemented and validated on a Raspberry Pi 3, confirming feasibility with low inference latency and acceptable resource usage, making it the only reviewed study to explicitly evaluate deployment on actual IoT hardware.

5.1.11 Cross-Study Performance Comparison and Overfitting Analysis

A critical limitation observed across the reviewed studies is the reporting of performance metrics in isolation, without cross-study comparison on identical datasets. To address this, Table 10 presents a structured comparison of accuracy values reported by studies sharing the same benchmark dataset. Several observations warrant critical attention. Among the seven studies employing the NSL-KDD dataset, reported accuracy values range from 82% [88] to 99.10% [70], a variance of approximately 17 percentage points that cannot be attributed solely to algorithmic differences, as variations in train-test splitting strategies, feature preprocessing, and class balancing techniques introduce confounding factors that preclude direct comparison. Similarly, among the four studies using the TON-IoT dataset, accuracy values range from

93.4% [89] to 98.0% [90], suggesting that dataset partitioning and preprocessing choices influence reported outcomes as significantly as the DRL algorithm itself.

Regarding overfitting, a notable concern across the reviewed studies is the near-perfect accuracy values reported by several models, particularly those achieving accuracies exceeding 99% [23,70,98]. However, the majority of these studies neither report cross-validation results nor provide learning curves that would allow assessment of the generalization gap between training and testing performance. The absence of validation on independent or held-out datasets, combined with the known class imbalance in datasets such as NSL-KDD and N-BaIoT, raises legitimate concerns that reported performance metrics may reflect overfitting to dataset-specific distributional characteristics rather than genuine generalization capability. Furthermore, no reviewed study evaluated their model on more than two datasets simultaneously, limiting confidence in the transferability of the reported results to real-world heterogeneous IoT environments. Future research should adopt rigorous cross-dataset validation protocols and explicitly report generalization metrics to enable meaningful and credible performance comparisons.

To ensure that the overfitting-risk classification in Table 10 is reproducible rather than subjective, each study was assessed against three observable methodological safeguards: (C1) whether cross-validation was reported; (C2) whether an independent, held-out, or cross-dataset test set was used; and (C3) whether the model was evaluated on more than one dataset. Each study receives one point per safeguard present (score range 0–3), and the risk level is assigned by a fixed rule: a score of 0 corresponds to *High* risk, a score of 1 to *Moderate* risk, and a score of 2 or above to *Low* risk. As a single documented override, a study reporting near-perfect accuracy (exceeding 99%) without cross-validation and without an independent external or cross-dataset validation protocol is classified as *High* risk regardless of its score. Notably, none of the reviewed studies reported cross-validation (C1), which is itself a key finding of this review. It should be emphasized that this rubric characterises the completeness of validation reporting rather than a directly measured generalization gap, as the primary studies do not provide the information required to compute the latter.

5.2 RQ2: What Feature Selection Methods Are Applied in Conjunction with DRL-based IDS?

This review highlights machine learning feature selection techniques and their major categories. Feature selection is essential for crafting robust, efficient DRL-based intrusion detection models, particularly for IoT devices with constrained resources. Recent research has demonstrated that employing selected critical features instead of the original feature set can decrease training time and costs [99,100]. This review investigates feature selection methods used in studies applying DRL-based IDS models within secure IoT settings. Three studies explicitly employ feature selection approaches. Two utilized the LightGBM algorithm [99,100], and one employed Enhanced Mutual Information Feature Selection (MIFS) [23]. Study [99] employed a dataset from a natural gas pipeline transportation network with 26 features. By narrowing the features to three, the LightGBM algorithm lessened computational complexity while sustaining high performance. The proposed intrusion model with the PPO2 algorithm achieved a 99.09% accuracy rate. Study [100] used LightGBM to derive a reduced feature set from an industrial SDN-IoT traffic dataset comprising approximately 1.68 million instances. This method substantially reduced training time and enhanced performance metrics. Additionally, two studies [88,103] referred to feature selection without detailing the methods. The other 21 studies neither deployed feature selection methods nor mentioned the process. Fig. 12 provides a visual representation.

Study [23] employed Enhanced Mutual Information Feature Selection (MIFS) to preprocess the CICIoMT2024 dataset, selecting the most relevant features while reducing noise and computational complexity. The selected features were then processed through a hybrid CNN-LSTM architecture, demonstrating

that feature selection is particularly critical for IoMT environments where high-dimensional healthcare data must be processed under strict latency constraints.

Figure 12: Distribution of feature selection methods in DRL-based IDS for IoT. The (N/A) category denotes studies that did not apply or report any feature selection procedure.

Accordingly, the distribution in Fig. 12 comprises three categories: studies employing a named feature selection method (LightGBM or MIFS), studies that referred to feature selection without specifying the method, and studies with no feature selection, the latter shown as the (N/A) category. These correspond to 3, 2, and 21 studies respectively, summing to the 26 included primary studies.

5.2.1 State-Space and Reward-Function Design Patterns

The construction of the state vector S and the alignment of the reward R are the two engineering decisions that most directly determine the practical behavior of a DRL-based IDS, and they are closely tied to the feature-engineering concerns of this research question. Across the reviewed studies, three recurring design patterns can be identified along the dimensions of state-space representation, action-space type, and reward-function structure; these are summarized conceptually in Table 11.

Although all 26 reviewed frameworks share the common Markov Decision Process formulation of state, action, and reward introduced in Section 2, they differ substantially in how each component is instantiated; the recurring patterns observed across the full corpus are characterised below.

Three observations follow. First, the state space is overwhelmingly constructed from flow-level statistical features inherited from the benchmark dataset schema (for example, the NSL-KDD attribute set or the TON-IoT telemetry fields) rather than from IoT-protocol-aware encodings. As a consequence, the IoT-specific structure of the traffic is largely invisible to the agent, which observes only generic IP-flow abstractions. Second, the action space is predominantly discrete, framing intrusion detection as a classification decision; only a minority of studies exploit continuous-control formulations to output trust scores, adaptive thresholds, or mitigation actions. Third, and most critically, the reward is most commonly a simple binary detection signal,

which is poorly aligned with the severe class imbalance of IoT traffic, since an agent can accumulate high reward by correctly labelling the abundant normal class while neglecting rare attacks. Only a small number of studies adopted class-weighted or multi-objective rewards that explicitly account for minority-attack detection or deployment cost.

Table 11: Conceptual taxonomy of state-space, action-space, and reward-function design patterns observed across reviewed DRL-based IDS models.

Design Dimension	Observed Patterns
State-space representation	(i) Flow-level statistical features inherited directly from the benchmark dataset schema (dominant); (ii) sequential/temporal features via recurrent encoders; (iii) features augmented with auxiliary signals (e.g., trust or belief metrics). Protocol-aware encodings (MQTT, CoAP, Zigbee) were not observed.
Action space	(i) Discrete classification actions, i.e., normal-vs.-attack or multi-class labelling (dominant); (ii) continuous-control actions, e.g., trust scores, detection thresholds, or mitigation/routing decisions (minority).
Reward structure	(i) Binary detection reward tied to correct classification (dominant); (ii) class-weighted reward penalising missed minority attacks; (iii) multi-objective reward jointly encoding detection and deployment costs such as energy, QoS, or false-positive control (rare).

A further and important finding is that a substantial proportion of the reviewed studies do not fully specify their state encoding or reward function, describing their models only at the algorithmic level. This inconsistent reporting constitutes a reproducibility gap that hinders direct comparison and re-implementation. We therefore recommend that future DRL-based IDS designs (i) construct protocol-aware state representations that preserve IoT-specific semantics, (ii) adopt reward functions that jointly encode detection of rare attacks and on-device cost rather than defaulting to binary formulations, and (iii) report their full MDP formulation to support reproducibility.

5.3 RQ3: What Types of Attacks in IoT Were These Technologies Designed to Detect?

This research question seeks to categorize the types of attacks detected by DRL-based IDSs within IoT environments. A taxonomic clarification is necessary before interpreting the distribution presented in Fig. 13. The identified attacks span three distinct categories that should not be treated as a flat list. The first category comprises attack strategies, including Denial of Service (DoS), Distributed Denial of Service (DDoS), Reconnaissance, and Scanning attacks, which describe the operational objective and execution method of the adversary. The second category comprises attack mechanisms, including Backdoor attacks, which describe the technical vector exploited to establish or maintain unauthorized access. The third category comprises specific malware families, including Mirai and Torii, which are IoT-specific botnets responsible for large-scale DDoS campaigns and are named by their malware lineage rather than their attack strategy. The term Generic, which appeared in several studies sourced from the UNSW-NB15 dataset labels, does not represent a distinct attack type but rather a dataset-specific label denoting attacks that do not conform to any predefined signature category. Its inclusion in prior analyses reflects the dataset labelling conventions rather than a meaningful attack classification, and it is therefore excluded from the taxonomic framework adopted in this review. The most frequently targeted attack strategies are DoS and DDoS, reflecting the susceptibility

of resource-constrained IoT devices to volumetric exhaustion attacks [107], while the prevalence of Mirai and Torii underscores the persistent threat posed by IoT-targeting botnet malware. It is recommended that future researchers adopt a consistent taxonomic framework when categorizing IoT attacks to enable meaningful cross-study comparisons.

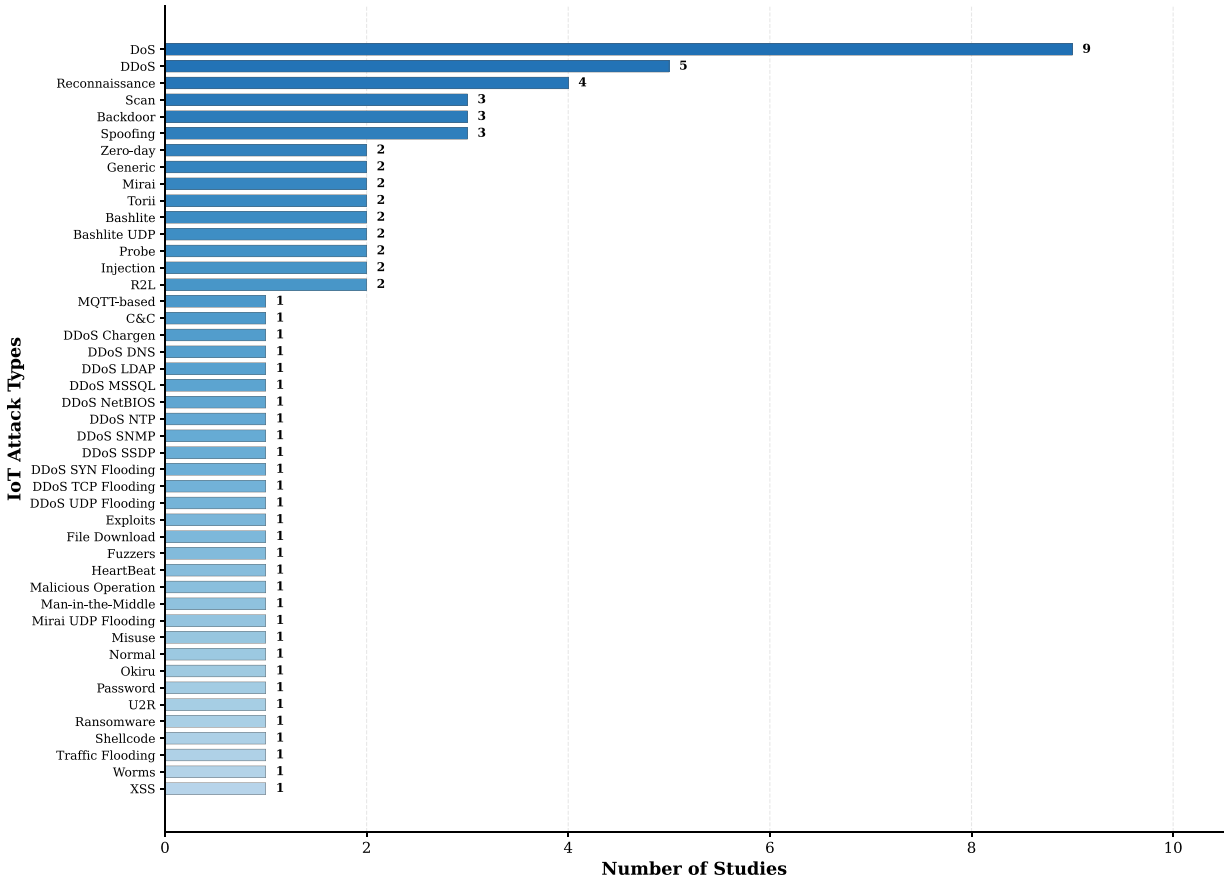

Figure 13: Distribution of different IoT attack types.

5.4 RQ4: What Datasets Are Widely Used with DRL-based IDS, Their Advantages and Their Limitations?

DRL-based IDS models strive to develop broadly applicable and efficient frameworks for accurately identifying attacks. This research identifies and compiles 20 datasets as shown in Fig. 14. Table 12 presents the overview of IoT attack datasets. This review identifies that the majority of DRL-based IDS studies rely on datasets that, while publicly accessible and widely adopted as benchmarks, present significant representational limitations when evaluated against the specific characteristics of real-world IoT environments. A critical and weighted assessment of the most frequently used datasets is therefore necessary before concluding the generalizability of the reviewed models.

NSL-KDD, the most frequently used dataset, appears in 7 studies (26.9%) [70,88,91–93,96,104], was introduced in 2009 as a refinement of the KDDCup99 dataset [108]. Despite its widespread adoption, NSL-KDD has been extensively criticised in the literature for its fundamental unsuitability as an IoT benchmark. It was generated from a simulated military network environment that bears no structural resemblance to modern IoT architectures, contains no IoT-specific protocol traffic such as MQTT, CoAP, or Zigbee, and does not include contemporary attack vectors such as botnet-driven DDoS, ransomware, or firmware exploitation.

Its attack categories, including Probe, U2R, R2L, and DoS, reflect threat models from the late 1990s that are largely obsolete in current IoT threat landscapes. Studies achieving near-perfect accuracy on NSL-KDD, such as [70] with 99.10%, should therefore be interpreted with considerable caution, as high performance on this dataset provides limited evidence of real-world IoT deployability.

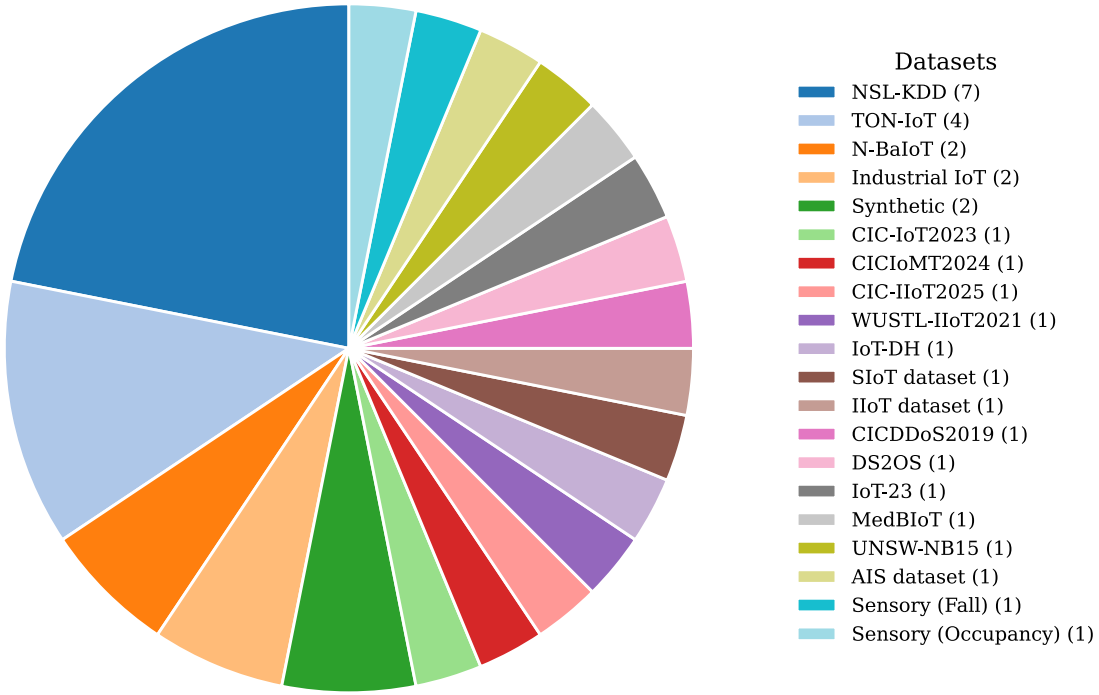

Figure 14: Dataset types included in reviewed studies.

The TON-IoT dataset, used in 4 studies (15.4%) [22,83,89,90], represents a comparatively more appropriate benchmark, as it was specifically designed to capture heterogeneous IoT and IIoT telemetry data from physical sensors and includes contemporary attack types [109]. However, TON-IoT has been criticised for lacking diversity in normal traffic patterns and for the absence of real botnet traffic, which limits its fidelity to live IoT network conditions.

The N-BaIoT dataset [110], used in 2 studies (7.7%) [91,98], was collected from real IoT devices infected with Mirai and Bashlite malware, making it one of the more ecologically valid datasets in this review. Nevertheless, its coverage is restricted to two botnet families, and its traffic features are extracted at a statistical level, precluding packet-level analysis.

Datasets such as UNSW-NB15 and CICDDoS2019, while large and structurally diverse, were generated in laboratory environments using non-IoT hardware and lack IoT protocol representation, rendering conclusions drawn from models trained on these datasets of limited applicability to constrained IoT deployments. Studies evaluating their models on these non-IoT datasets [89] should be weighted accordingly when assessing the contribution of their proposed DRL-based IDS models to practical IoT security.

Table 12: Overview of IoT attack datasets for the selected studies.

Dataset	Year	IoT Focus	Instances	Features	Unknown Attacks	Advantages	Limitations
NSL-KDD [24,108,112]	2009	No	148,517	42	No	Large number of samples; No duplicate records	Inconsistency in test/training sets; Lacks modern attacks
TON-IoT [77,109,112–115]	2019	Yes	22,339,021	46	Yes	Includes normal and attack events; Features physical IoT sensors	Lacks variety of normal traffic; No real botnet traffic
N-BaIoT [110,112,116,117]	2018	Yes	7,062,606	116	Yes	Addresses lack of public botnet datasets for IoT	Focuses on Mirai and Bashlite only
MedBIoT [115–118]	2020	Yes	37,433,560	100	No	Real botnet attack network traffic	Lacks IoT protocol-based attacks
CICDDoS2019 [24,119,120]	2019	No	50,063,112	88	No	Includes normal traffic and real-world DDoS attacks	DDoS attacks heavily emphasized; IoT context not addressed
UNSW-NB15 [24,112]	2015	No	2,540,044	49	No	Large sample count; Combines real and synthetic attacks	Class imbalance and overlap; Not IoT focused
DS2OS [121,122]	2018	Yes	357,952	13	No	Handles concept drift; Adapts to evolving attacks	Limited attack diversity and class imbalance
IoT-23 [116,123,124]	2020	Yes	325,307,990	19	Yes	Diverse IoT devices; Large labeled dataset	Severely imbalanced; Focused on DNS traffic
AIS dataset [125]	2020	Yes	Unknown	–	–	–	Lacks large abnormal traffic
Sensory data (Fall) [97]	2010	Yes	164,259	5	No	Large number of samples	High volatility; Imbalanced
Sensory data (Occupancy) [97]	2016	Yes	20,560	7	No	Clean and well-structured data	Small samples; Poor generalizability
Industrial IoT [99]	–	Yes	196,315	27	No	Large number of samples	Class imbalance
Industrial IoT [100]	–	Yes	1,679,741	27	No	Large number of samples	Class imbalance
CIC-IoT2023 [106]	2023	Yes	–	–	Yes	Contemporary IoT malicious traffic; diverse attacks	Relatively new; limited adoption
CICIoMT2024 [23]	2024	Yes	3,200,000	–	Yes	IoMT-specific; 18 attack classes	Single domain focus
CIC-IIoT2025, TON-IoT [22]	2025	Yes	–	–	Yes	Latest IIoT dataset; contemporary attacks	Limited adoption
WUSTL-IIoT2021 [22]	2021	Yes	–	–	No	Real IIoT testbed data	Limited diversity
IoT-DH [94]	2024	Yes	–	–	Yes	Real-world honeypot data; multi-protocol	Limited scale; honeypot bias
SIoT dataset [84]	–	Yes	–	–	Yes	Social IoT network traffic	Limited availability
IIoT dataset [85]	–	Yes	–	–	No	IIoT simulation traffic	Synthetic

The overrepresentation of traditional network datasets, particularly NSL-KDD, in the DRL-based IoT IDS literature, reflects a persistent methodological gap rather than an informed benchmark selection. Future research must prioritise IoT-native datasets such as TON-IoT, N-BaIoT, MedBIoT, and IoT-23, or generate new datasets that faithfully capture the heterogeneous, resource-constrained, and protocol-diverse characteristics of contemporary IoT environments. Fig. 14 illustrates the distribution of dataset usage across the reviewed studies, and Table 12 presents a structured overview of each dataset alongside its documented advantages and limitations.

The volume and velocity of the underlying dataset also directly influence the detection accuracy attainable by DRL-based IDS models. Because DRL agents learn through repeated interaction with sampled experiences, high-volume datasets such as TON-IoT, IoT-23, and CICDDoS2019 provide a richer and more diverse experience distribution than smaller benchmarks like NSL-KDD, allowing the agent to encounter rare attack patterns more frequently and thereby reducing the minority-class underrepresentation that inflates false negatives. Likewise, high-velocity streaming traffic exposes the agent to non-stationary distributions resembling live IoT deployment, encouraging detection policies that remain stable under rapidly shifting conditions rather than overfitting to static snapshots, a regime to which DRL's sequential decision-making is well suited.

These gains, however, increase the memory and computational cost of experience replay and policy updates, which conflicts with the resource constraints of edge IoT nodes. We therefore recommend that future research adopt high-volume, high-velocity IoT-native datasets to improve accuracy and temporal robustness, while concurrently reporting the associated overhead and considering energy-efficient computation-offloading strategies to mitigate on-device cost [111], so that accuracy gains can be weighed against deployment feasibility.

5.5 RQ5: What Evaluation Metrics Are Employed to Assess the Performance of Algorithms in the IDSs?

Evaluating DRL-based IDS models requires a dual-layered assessment framework that addresses both the detection capability of the model and its operational feasibility within resource-constrained IoT environments. The reviewed studies, however, reveal a pronounced and methodologically significant imbalance between these two evaluation dimensions, as illustrated in Figs. 15 and 16.

Among the 26 reviewed studies, data mining metrics were near-universally adopted. Accuracy was reported in 23 studies (88.5%) [20,21,70,91,100,103], F1-score in 20 studies (76.9%), precision in 19 studies (73.1%), and recall in 18 studies (69.2%). While these metrics are appropriate for quantifying classification performance, their exclusive use presents a critical limitation in the IoT context. Accuracy, in particular, is a misleading indicator when applied to the severely imbalanced class distributions characteristic of IoT intrusion datasets such as NSL-KDD and N-BaIoT, where a classifier that labels all traffic as normal may still achieve deceptively high accuracy. The predominance of accuracy as a primary metric across studies therefore warrants critical caution when interpreting and comparing reported results.

Network performance metrics, which are directly relevant to the practical deployability of DRL-based IDS models on constrained IoT devices, were reported in only 6 studies (23.1%). Training time was the most commonly reported network performance metric (6 studies, 23.1%), followed by model loss and inference latency (5 studies each, 19.2%). Critical deployment metrics remained scarce: energy consumption and CPU utilization were each reported in only three studies (11.5%), memory footprint in two (7.7%), and network throughput in a single study (3.8%). This represents a fundamental disconnect between the academic evaluation of DRL-based IDS models and the operational realities of IoT deployment, where nodes typically operate under strict constraints on processing power, memory, and battery life. A model achieving 99%

accuracy while consuming prohibitive computational resources is, in practice, undeployable in the majority of real-world IoT scenarios.

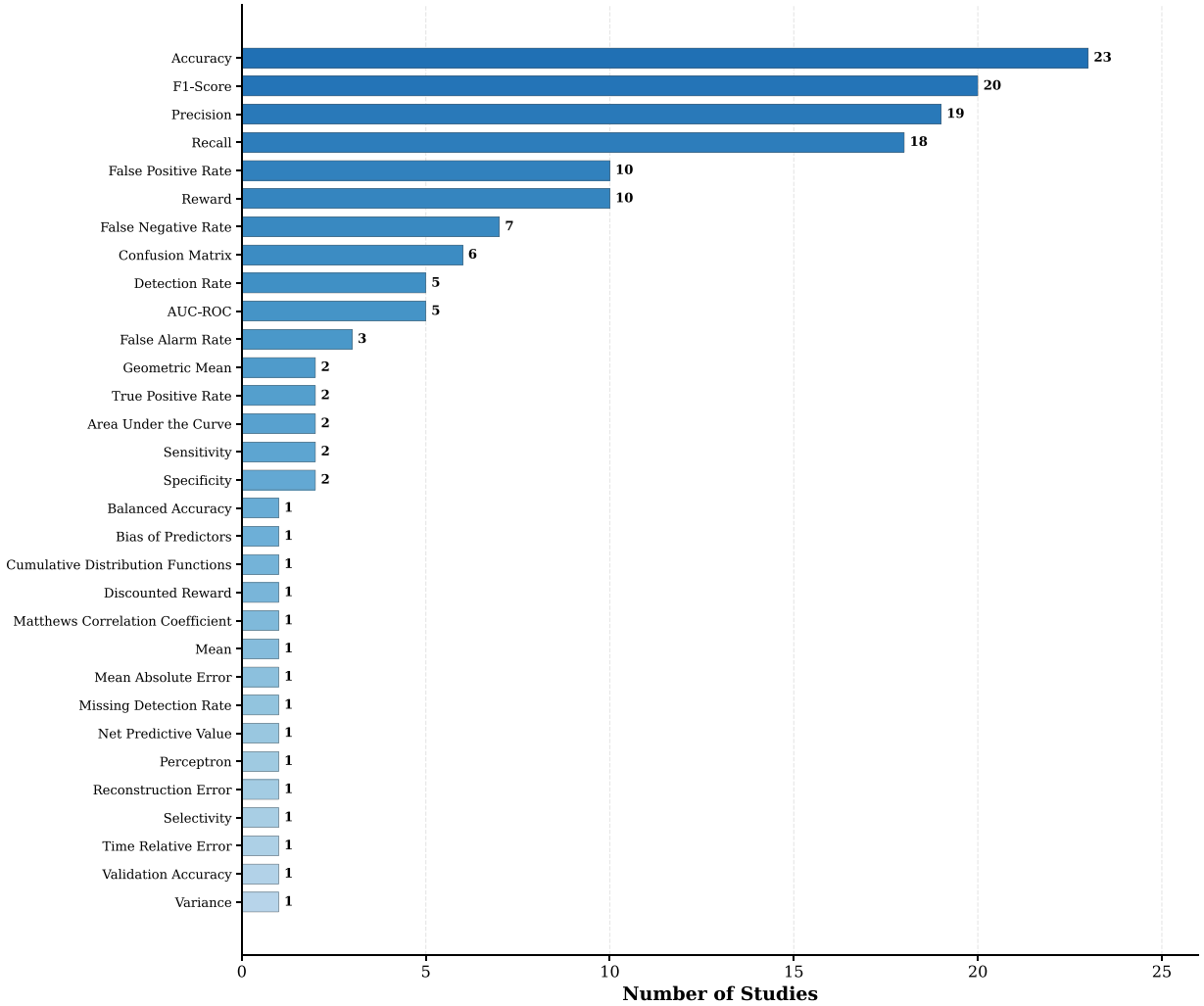

Figure 15: Metrics distribution in DRL-based IDS evaluation.

Table 13 presents a structured summary of the evaluation gap observed across the reviewed studies, distinguishing between detection-oriented and deployment-oriented metrics.

This evaluation gap has direct implications for the reproducibility and comparability of results across studies. The absence of standardised reporting guidelines for DRL-based IDS evaluation in IoT environments means that studies optimise and report metrics selectively, often favouring those that present their models most favourably. Future research must adopt a comprehensive evaluation framework that mandates reporting of both detection performance and deployment feasibility metrics, particularly energy consumption, inference latency, and memory utilization, to enable credible assessment of whether proposed DRL-based IDS models are genuinely viable for real-world IoT deployment.

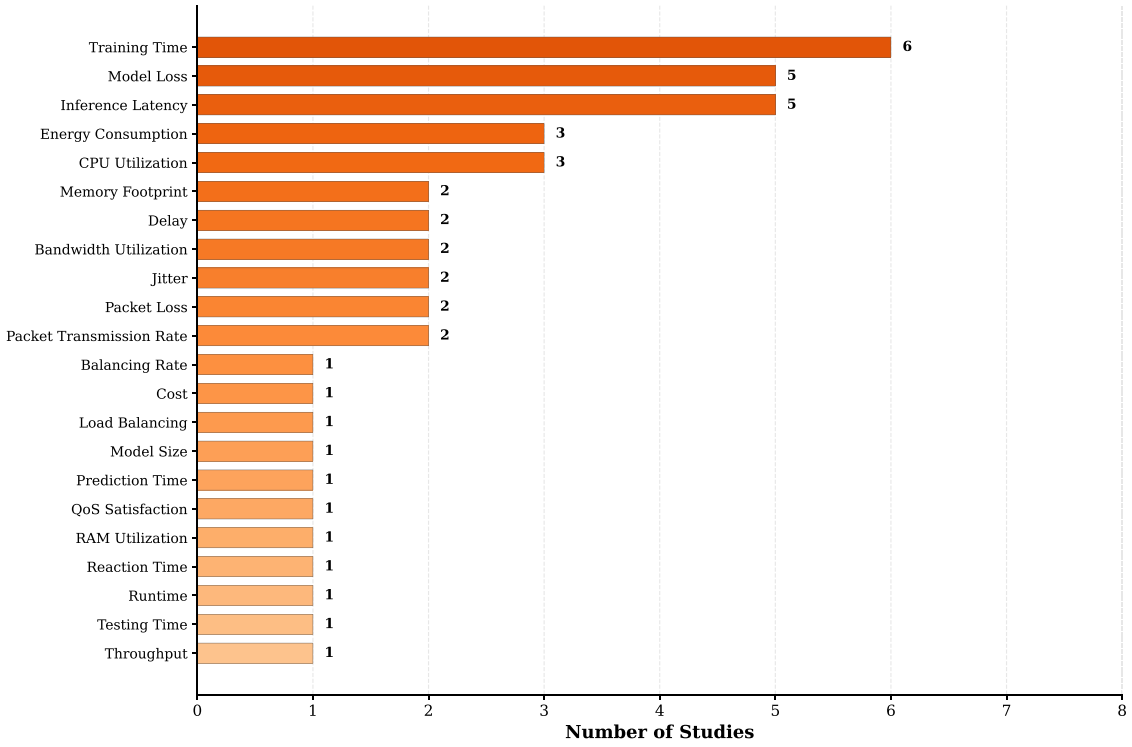

Figure 16: Network performance metrics distribution in DRL-based IDS evaluation.

Table 13: Evaluation metrics coverage across reviewed DRL-based IDS studies.

Metric	Studies Reporting	Coverage (%)
Data Mining Metrics		
Accuracy	23	88.5
F1-Score	20	76.9
Precision	19	73.1
Recall	18	69.2
False Positive Rate	10	38.5
False Negative Rate	7	26.9
AUC-ROC	5	19.2
G-Mean	2	7.7
Network Performance Metrics		
Training Time	6	23.1
Model Loss	5	19.2
Inference Latency	5	19.2
Energy Consumption	3	11.5
CPU utilization	3	11.5
Memory Footprint	2	7.7
Network Throughput	1	3.8

Note: The stark contrast between data mining metric coverage (8%–89%) and network performance metric coverage (4%–23%) reveals a critical evaluation gap in the DRL-based IoT IDS literature.

6 Research Gaps, Challenges, and Future Research Directions

While Section 5 provides a detailed empirical analysis of the reviewed studies, this section synthesizes higher-order research gaps that transcend individual study limitations and identifies forward-looking challenges and directions that the existing literature has not yet adequately addressed.

6.1 Absence of Standardised Benchmarking Protocols

No reviewed study adopted a unified benchmarking protocol that would enable direct and reproducible cross-study comparison. The heterogeneity in dataset preprocessing, train-test splitting strategies, hyperparameter configurations, and evaluation metric selection renders the reported results largely incommensurable. The establishment of a community-agreed benchmarking standard for DRL-based IDS evaluation in IoT environments, analogous to those adopted in computer vision and natural language processing, represents a critical and unaddressed research priority.

6.2 Lack of Adversarial Robustness Evaluation

None of the reviewed studies evaluated the robustness of their proposed DRL-based IDS models against adversarial attacks, including adversarial machine learning (AML) techniques such as evasion attacks, poisoning attacks, and model inversion. Given that intelligent adversaries are increasingly capable of manipulating network traffic to evade learned detection policies, the absence of adversarial robustness testing represents a significant gap between laboratory evaluation and real-world threat conditions.

6.3 Absence of Formal Privacy Analysis

IoT intrusion detection inherently involves the continuous monitoring and analysis of network traffic that may contain sensitive user data. Only one reviewed study [95] explicitly addressed privacy preservation through federated learning. No study conducted a formal privacy analysis or evaluated compliance with data protection frameworks. Future research must integrate formal privacy guarantees, such as differential privacy or secure multi-party computation, into DRL-based IDS architectures to ensure regulatory compliance and user trust.

6.4 Absence of Continual and Online Learning Frameworks

All reviewed DRL-based IDS models were trained in static, offline settings using fixed historical datasets. None investigated continual or online learning frameworks capable of adapting to concept drift, which is a fundamental characteristic of evolving IoT threat landscapes. The development of DRL-based IDS models capable of incrementally updating their detection policies in response to newly observed attack patterns, without catastrophic forgetting of previously learned behaviours, constitutes an open and largely unexplored research challenge.

6.5 Absence of Explainability and Interpretability

No reviewed study incorporated explainability mechanisms into their DRL-based IDS models. In operational IoT security environments, the ability of a detection system to provide human-interpretable justifications for its classification decisions is essential for analyst trust, regulatory compliance, and forensic investigation. The integration of explainable AI (XAI) techniques, such as SHapley Additive explanations (SHAP), Local Interpretable Model-agnostic Explanations (LIME), or attention-based attribution into DRL-based IDS architectures for IoT represents a critical and unaddressed research direction.

6.6 Insufficient Consideration of IoT Protocol Diversity

The reviewed studies predominantly treated IoT network traffic at a generic IP packet level, without accounting for the diversity of IoT-specific communication protocols such as MQTT, CoAP, Zigbee, and Z-Wave. Future DRL-based IDS models must incorporate protocol-aware feature engineering and detection logic to address the unique vulnerability surfaces presented by each protocol layer in heterogeneous IoT deployments.

7 Conclusion

This systematic literature review synthesized findings from 26 primary studies, selected from 267 candidates through a PRISMA-guided screening process, examining DRL-based IDS models for IoT environments published between 2020 and 2026. Across five structured research questions, the review established that DQN is the most widely adopted algorithm (30.8% of studies), followed by Double DQN (19.2%), feature selection remains critically underutilized with only three studies employing feature engineering methods (LightGBM and MIFS), and DoS and DDoS constitute the most frequently targeted attack strategies. Dataset analysis identified a persistent methodological misalignment, as NSL-KDD, despite its well-documented unsuitability for IoT benchmarking, remained the most frequently used dataset. Evaluation metric analysis revealed a critical imbalance between detection-oriented metrics, reported in 69%–89% of studies, and deployment-oriented metrics such as energy consumption and memory utilization, reported in fewer than 12% of studies. Six higher-order research gaps were identified, including the absence of adversarial robustness evaluation, continual learning frameworks, explainability mechanisms, and IoT protocol-aware detection logic. Emerging advances in dynamic deception orchestration, federated reinforcement learning, and hybrid DRL architectures further validate the expanding applicability of DRL in IoT security. This review provides a rigorous empirical foundation and structured research agenda for advancing deployable DRL-based IDS solutions in IoT environments.

Acknowledgement: The authors thank the Deanship of Scientific Research and Graduate Studies at King Khalid University and Yayasan Universiti Teknologi PETRONAS for their support of this work.

Funding Statement: The authors extend their appreciation to the Deanship of Scientific Research and Graduate Studies at King Khalid University for funding this work through the Large Group Research Project under grant number (RGP2-209-46). The authors also acknowledge the support of Yayasan Universiti Teknologi PETRONAS under the YUTP-FRG grant (015LC0-578).

Author Contributions: The authors confirm their contributions to the paper as follows: Maryam Omar Abdullah Sawad: conceptualization, methodology, investigation, data curation, formal analysis, writing original draft. Said Jadid Abdulkadir: conceptualization, methodology, supervision, validation, writing review & editing, project administration. Hitham Seddig Alhussian: supervision, validation, formal analysis, writing review & editing. Majdy Mohamed Eltayeb Eltahir: funding acquisition, resources, validation, writing review & editing. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: This study is a Systematic Literature Review (SLR) and does not generate any new primary datasets. All data supporting the findings of this review were derived from publicly available peer-reviewed articles sourced from the Scopus, Web of Science, and Google Scholar databases. The 26 selected studies analyzed in this review are cited within the manuscript and listed in the reference section. The publicly available datasets discussed in this review, including NSL-KDD, TON-IoT, N-BaIoT, UNSW-NB15, CICDDoS2019, MedBIoT, DS2OS, and IoT-23, can be accessed through their respective sources as referenced in Table 12. No additional data beyond what is presented in this article are required to reproduce the findings of this review.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Supplementary Materials: The supplementary material is available online at https://www.techscience.com/doi/10.32604/cmes.2026.085472/s1.

Abbreviations

The following abbreviations are used in this manuscript

Abbreviation	Definition
A2C	Advantage Actor-Critic
AC	Actor-Critic
AE-RL	Adversarial Environment Reinforcement Learning
AI	Artificial Intelligence
AML	Adversarial Machine Learning
AUC-ROC	Area Under the Receiver Operating Characteristic Curve
BWO	Black Widow Optimization
CIoT	Cellular Internet of Things
CNN	Convolutional Neural Network
CoAP	Constrained Application Protocol
CPU	Central Processing Unit
CSD	Clustering-based Semi-supervised Defense
CVAE	Conditional Variational Auto-Encoder
DDoS	Distributed Denial of Service
DDPG	Deep Deterministic Policy Gradient
DDQN	Double Deep Q-Network
DL	Deep Learning
DNN	Deep Neural Network
DNS	Domain Name System
DoS	Denial of Service
DQN	Deep Q-Network
DRL	Deep Reinforcement Learning
FL	Federated Learning
FN	False Negative
FP	False Positive
GAN	Generative Adversarial Network
G-Mean	Geometric Mean
GPU	Graphics Processing Unit
HIDS	Host-based Intrusion Detection System
IAM	Identity and Access Management
IDS	Intrusion Detection System
IIoT	Industrial Internet of Things
IoMT	Internet of Medical Things
IoT	Internet of Things
IoT-MTS	IoT-empowered Maritime Transportation Systems
IPS	Intrusion Prevention System
LDAP	Lightweight Directory Access Protocol
LightGBM	Light Gradient Boosting Machine
LIME	Local Interpretable Model-agnostic Explanations

LSD	Label-based Semi-supervised Defense
LSTM	Long Short-Term Memory
MADDPG	Multiagent Deep Deterministic Policy Gradient
MARL	Multi-Agent Reinforcement Learning
MDP	Markov Decision Process
MIFS	Mutual Information Feature Selection
ML	Machine Learning
MQTT	Message Queuing Telemetry Transport
MSSQL	Microsoft SQL Server
NIDS	Network-based Intrusion Detection System
NTP	Network Time Protocol
PER	Prioritized Experience Replay
PPO	Proximal Policy Optimization
PPO2	Proximal Policy Optimization 2
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QoS	Quality of Service
R2L	Remote-to-Local
RFID	Radio Frequency Identification
RL	Reinforcement Learning
SADDPG	Single-Agent Deep Deterministic Policy Gradient
SDN	Software-Defined Networking
SHAP	SHapley Additive exPlanations
SIBRO	Self-Improved Battle Royale Optimization
SIoT	Social Internet of Things
SLR	Systematic Literature Review
SMOTE	Synthetic Minority Over-sampling Technique
SNMP	Simple Network Management Protocol
SYN	Synchronize (TCP Synchronize Segment)
TCP	Transmission Control Protocol
TD3	Twin Delayed Deep Deterministic Policy Gradient
TON	Telemetry, Operating Systems, and Network
TRPO	Trust Region Policy Optimization
U2R	User-to-Root
UDP	User Datagram Protocol
WSN	Wireless Sensor Network
XAI	Explainable Artificial Intelligence
XGBoost	Extreme Gradient Boosting
XSS	Cross-Site Scripting

References

1. Yalli JS, Hasan MH, Badawi AA. Internet of Things (IoT): origins, embedded technologies, smart applications, and its growth in the last decade. IEEE Access. 2024;12(1):91357–82. doi:10.1109/access.2024.3418995.
2. Dastjerdi AV, Buyya R. Fog computing: helping the internet of things realize its potential. Computer. 2016;49(8):112–6. doi:10.1109/mc.2016.245.
3. Al-Amiedy TA, Anbar M, Belaton B, Kabla AHH, Hasbullah IH, Alashhab ZR. A systematic literature review on machine and deep learning approaches for detecting attacks in RPL-based 6LoWPAN of Internet of Things. Sensors. 2022;22(9):3400. doi:10.3390/s22093400.

4. Mukhaini GA, Anbar M, Manickam S, Al-Amiedy TA, Momani AA. A systematic literature review of recent lightweight detection approaches leveraging machine and deep learning mechanisms in Internet of Things networks. J King Saud Univ Comput Inf Sci. 2024;36(1):101866. doi:10.1016/j.jksuci.2023.101866.
5. Yalli JS, Hasan MH. A unique PUF authentication protocol based fuzzy logic categorization for Internet of Things (IOT) devices. In: ACM International Conference Proceeding Series. New York, NY, USA: ACM; 2023. p. 246–52.
6. Sinha S. Number of connected IoT devices growing 13% to 18.8 billion. IoT Anal. [cited 2024 Dec 8]. Available from: https://iot-analytics.com/number-connected-iot-devices/.
7. Nikoui TS, Rahmani AM, Balador A, Javadi HHS. Internet of Things architecture challenges: a systematic review. Int J Commun Syst. 2021;34(4):1–42. doi:10.1002/dac.4678.
8. Yalli JS, Hasan MH. Internet of Things (IoT): trends, challenges, simulators, emulators and test-beds. In: Proceedings of the 2024 8th International Conference on Computing, Communication, Control and Automation; 2024 Oct 17–19; Pune, India. p. 1–6.
9. Zeinab KAM, Elmustafa SAA. Internet of things applications, challenges and related future technologies. World Sci News. 2017;67(2):126–48. [cited 2026 Jan 6]. Available from: https://worldscientificnews.com/internet-of-things-applications-challenges-and-related-future-technologies/.
10. Khan MA, Salah K. IoT security: review, blockchain solutions, and open challenges. Fut Gener Comput Syst. 2018;82(15):395–411. doi:10.1016/j.future.2017.11.022.
11. Yalli JS, Hasan MH, Jung LT, Al-Selwi SM. Internet of Things (IoT) networks: a systematic review and security assessment of authentication schemes. Internet Things. 2024;30(1):101469. doi:10.1016/j.iot.2024.101469.
12. Jamalipour A, Murali S. A taxonomy of machine-learning-based intrusion detection systems for the internet of things: A survey. IEEE Internet Things J. 2022;9(12):9444–66. doi:10.1109/jiot.2021.3126811.
13. Ribera EG, Alvarez BM, Samuel C, Ioulianou PP, Vassilakis VG. An intrusion detection system for RPL-based IoT networks. Electronics. 2022;11(23):4041. doi:10.3390/electronics11234041.
14. Caminero G, Lopez-Martin M, Carro B. Adversarial environment reinforcement learning algorithm for intrusion detection. Comput Netw. 2019;159(1):96–109. doi:10.1016/j.comnet.2019.05.013.
15. Uprety A, Rawat DB. Reinforcement learning for IoT security: A comprehensive survey. IEEE Internet Things J. 2021;8(11):8693–706.
16. Ding Z, Huang Y, Yuan H, Dong H. Introduction to reinforcement learning. In: Deep reinforcement learning: fundamentals, research and applications. Berlin/Heidelberg, Germany: Springer; 2020. p. 47–123.
17. Bakhshad S, Ponnusamy V, Annur R, Waqasyz M, Alhussian H, Tux S. Deep reinforcement learning based intrusion detection system with feature selections method and optimal hyper-parameter in IoT environment. In: Proceedings of the 2022 International Conference on Computing, Information and Telecommunication Systems (CITS); 2022 Jul 13–15; Piraeus, Greece. p. 1–7.
18. Nguyen TT, Reddi VJ. Deep reinforcement learning for cyber security. IEEE Trans Neural Netw Learn Syst. 2023;34(8):3779–95. doi:10.1109/tnnls.2021.3121870.
19. Bouhamed O, Bouachir O, Aloqaily M, Al Ridhawi I. Lightweight IDS for UAV networks: a periodic deep reinforcement learning-based approach. In: Proceedings of the 2021 IFIP/IEEE International Symposium on Integrated Network Management; 2021 May 17–21; Bordeaux, France. p. 1032–7.
20. Nie L, Sun W, Wang S, Ning Z, Rodrigues JJPC, Wu Y, et al. Intrusion detection in green Internet of Things: a deep deterministic policy gradient-based algorithm. IEEE Trans Green Commun Netw. 2021;5(2):778–88. doi:10.1109/tgcn.2021.3073714.
21. Hu J, Kaur K, Lin H, Wang X, Hassan MM, Razzak I, et al. Intelligent anomaly detection of trajectories for IoT empowered maritime transportation systems. IEEE Trans Intell Transp Syst. 2023;24(2):2382–91. doi:10.1109/tits.2022.3162491.
22. Wushishi U, Hussain A, Khalid MI, Hussain N, Jamjoom M, Ullah Z. D3O-IIoT: deep reinforcement learning-driven dynamic deception orchestration for industrial IoT security. Sci Rep. 2026;16(1):2389. doi:10.1038/s41598-025-33426-4.

23. Shaikh JA, Wang C, Sima MWU, Arshad M, Owais M, Hassan DSM, et al. A deep reinforcement learning-based robust intrusion detection system for securing IoMT healthcare networks. Front Med. 2025;12:1524286. doi:10.3389/fmed.2025.1524286.
24. Adawadkar AMK, Kulkarni N. Cyber-security and reinforcement learning—A brief survey. Eng Appl Artif Intell. 2022;114:105116. doi:10.1016/j.engappai.2022.105116.
25. Arshad K, Ali RF, Muneer A, Aziz IA, Naseer S, Khan NS, et al. Deep reinforcement learning for anomaly detection: a systematic review. IEEE Access. 2022;10:124017–35. doi:10.1109/access.2022.3224023.
26. M. JFC, Rizzardi A, Sicari S, Porisini AC. Deep reinforcement learning for intrusion detection in Internet of Things: best practices, lessons learnt, and open challenges. Comput Netw. 2023;236(3):110016. doi:10.1016/j.comnet.2023.110016.
27. Haq SU, Abbas AM. Advancements in intrusion detection systems for Internet of Things using machine learning. In: Proceedings of the 2022 5th International Conference on Multimedia, Signal Processing and Communication Technologies (IMPACT); 2022 Nov 26–27; Aligarh, India. p. 1–5.
28. Khraisat A, Alazab A. A critical review of intrusion detection systems in the internet of things: techniques, deployment strategy, validation strategy, attacks, public datasets and challenges. Cybersecurity. 2021;4(1):18. doi:10.1186/s42400-021-00077-7.
29. Zhang H, Maple C. Deep reinforcement learning-based intrusion detection in IoT system: A review. IET Conf Proc. 2023;2023(14):88–97.
30. Al-Jarrah OY, Maple C, Dianati M, Oxtoby D, Mouzakitis A. Intrusion detection systems for intra-vehicle networks: a review. IEEE Access. 2019;7:21266–89. doi:10.1109/access.2019.2894183.
31. Gueriani A, Kheddar H, Mazari AC. Deep reinforcement learning for intrusion detection in IoT: A survey. In: Proceedings of the 2023 2nd International Conference on Electronics, Energy and Measurement (IC2EM); 2023 Nov 28–30; Médéa, Algeria. Vol. 1. New York, NY, USA: IEEE; 2023. p. 1–7.
32. Jamshidi S, Nikanjam A, Nafi KW, Khomh F, Rasta R. Application of deep reinforcement learning for intrusion detection in Internet of Things: a systematic review. J Netw Comput Appl. 2025. In press. doi:10.22541/au.174836360.04077727/v1.
33. Isma'ila UA, Danyaro KU, Muazu AA, Maiwada UD. Review on approaches of federated modeling in anomaly-based intrusion detection for IoT devices. IEEE Access. 2024;12(1):30941–61. doi:10.1109/access.2024.3369915.
34. Tan L, Wang N. Future internet: the Internet of Things. In: Proceedings of the 2010 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE); 2010 Aug 20–22; Chengdu, China. Vol. 5. p. 376–80.
35. Patel KK, Patel SM, Scholar PG. Internet of Things-IOT: definition, characteristics, architecture, enabling technologies, application and future challenges. Int J Eng Sci Comput. 2016;6(5):1–10.
36. Khan IA, Razzak I, Pi D, Zia U, Kamal S, Hussain Y. A novel collaborative SRU network with dynamic behaviour aggregation, reduced communication overhead and explainable features. IEEE J Biomed Health Inform. 2024;28(6):3228–35. doi:10.1109/jbhi.2024.3352013.
37. Khan IA, Pi D, Kamal S, Alsuhaibani M, Alshammari BM. Federated-boosting: a distributed and dynamic boosting-powered cyber-attack detection scheme for security and privacy of consumer IoT. IEEE Trans Consum Electron. 2024;71(2):6340–7. doi:10.1109/tce.2024.3499942.
38. Sasi T, Lashkari AH, Lu R, Xiong P, Iqbal S. A comprehensive survey on IoT attacks: Taxonomy, detection mechanisms and challenges. J Inf Intell. 2023;2(6):455–513. doi:10.1016/j.jiixd.2023.12.001.
39. G. K, Pulliyasseri A, Rajesh A, Ajayan A, Alfarhood S, Safran M, et al. Enhancing IoT resilience at the edge: a resource-efficient framework for real-time anomaly detection in streaming data. Comput Model Eng Sci. 2025;143(3):3005–31. doi:10.32604/cmes.2025.065698.
40. Ali RF, Muneer A, Dominic PDD, Taib SM, Ghaleb EAA. Internet of Things (IoT) security challenges and solutions: a systematic literature review. In: Communications in computer and information science. Vol. 1487. Berlin/Heidelberg, Germany: Springer; 2021. p. 128–54.
41. Mehdipour F. A review of IoT security challenges and solutions. In: Proceedings of the 8th International Japan-Africa Conference on Electronics, Communications, and Computations (JAC-ECC); 2020 Dec 14–15; Alexandria, Egypt. p. 1–6.

42. Zhang ZK, Cho MCY, Wang CW, Hsu CW, Chen CK, Shieh S. IoT security: ongoing challenges and research opportunities. In: Proceedings of the IEEE 7th International Conference on Service-Oriented Computing and Applications (SOCA); 2014 Nov 17–19; Matsue, Japan. p. 230–4.
43. Yakubu MM, Hassan F B, Danyaro M, Junejo KU, Siraj AZ, Yahaya M, et al. A systematic literature review on blockchain consensus mechanisms' security: applications and open challenges. Comput Syst Sci Eng. 2024;48(6):1437–81. doi:10.32604/csse.2024.054556.
44. Zarpelão BB, Miani RS, Kawakani CT, de Alvarenga SC. A survey of intrusion detection in Internet of Things. J Netw Comput Appl. 2017;84(3):25–37. doi:10.1016/j.jnca.2017.02.009.
45. Liao HJ, Lin CHR, Lin YC, Tung KY. Intrusion detection system: a comprehensive review. J Netw Comput Appl. 2013;36(1):16–24. doi:10.1016/j.jnca.2012.09.004.
46. Alsoufi MA, Siraj MM, Ghaleb FA, Al-Razgan M, Al-Asaly MS, Alfakih T. Anomaly-based intrusion detection model using deep learning for IoT networks. Comput Model Eng Sci. 2024;141(1):823–45. doi:10.32604/cmes.2024.052112.
47. Alahmari S, Alkharashi A. Privacy-aware federated learning framework for IoT security using chameleon swarm optimization and self-attentive variational autoencoder. Comput Model Eng Sci. 2025;143(1):849–73. doi:10.32604/cmes.2025.062549.
48. Wang Z, Chen H, Yang S, Luo X, Li D, Wang J. A lightweight intrusion detection method for IoT based on deep learning and dynamic quantization. PeerJ Comput Sci. 2023;9(19):e1569. doi:10.7717/peerj-cs.1569.
49. Banaamah MA, Ahmad I. Intrusion detection in IoT using deep learning. Sensors. 2022;22(21):8417. doi:10.3390/s22218417.
50. Yun S, Chen Y. Intelligent traffic scheduling for mobile edge computing in IoT via deep learning. Comput Model Eng Sci. 2023;134(3):1815–35. doi:10.32604/cmes.2022.022797.
51. Raoof A, Matrawy A, Lung CH. Routing attacks and mitigation methods for RPL-based Internet of Things. IEEE Commun Surv Tutor. 2019;21(2):1582–606. doi:10.1109/comst.2018.2885894.
52. Tabassum A, Erbad A, Mohamed A, Guizani M. Privacy-preserving distributed IDS using incremental learning for IoT health systems. IEEE Access. 2021;9:14271–83. doi:10.1109/access.2021.3051530.
53. Gendreau AA, Moorman M. Survey of intrusion detection systems towards an end to end secure internet of things. In: Proceedings of the IEEE 4th International Conference on Future Internet of Things and Cloud (FiCloud); 2016 Aug 22–24; Vienna, Austria. p. 84–90.
54. Butun I, Morgera SD, Sankar R. A survey of intrusion detection systems in wireless sensor networks. IEEE Commun Surv Tutor. 2014;16(1):266–82. doi:10.1109/surv.2013.050113.00191.
55. François-Lavet V, Henderson P, Islam R, Bellemare MG, Pineau J. An Introduction to deep reinforcement learning. Found Trends Mach Learn. 2018;11(3–4):219–354. doi:10.1561/2200000071.
56. Lavanya P, Sangeetha A, Krishnan S. Intrusion detection using machine learning. Int J Recent Technol Eng. 2019;8(2):832–7. doi:10.35940/ijrte.b1154.0782s619.
57. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. Human-level control through deep reinforcement learning. Nature. 2015;518(7540):529–33. doi:10.1038/nature14236.
58. Hasselt HV, Guez A, Silver D. Deep reinforcement learning with double Q-learning. Proc AAAI Conf Artif Intell. 2016;30(1):2094–100. doi:10.1017/9781108955652.016.
59. Huang L, Fu M, Qu H, Wang S, Hu S. A deep reinforcement learning-based method applied for solving multi-agent defense and attack problems. Expert Syst Appl. 2021;176:114896. doi:10.1016/j.eswa.2021.114896.
60. Lowe R, Wu YI, Tamar A, Harb J, Abbeel OP, Mordatch I. Multi-agent actor-critic for mixed cooperative-competitive environments. Adv Neural Inf Process Syst. 2017;30:6382–93. doi:10.48550/arxiv.1706.02275.
61. Mnih V, Badia AP, Mirza M, Graves A, Lillicrap T, Harley T, et al. Asynchronous methods for deep reinforcement learning. In: Proceedings of the 33rd International Conference on Machine Learning; 2016 Jun 19–24; New York City, NY, USA. p. 1928–37.
62. Lillicrap TP, Hunt JJ, Pritzel A, Heess N, Erez T, Tassa Y, et al. Continuous control with deep reinforcement learning. In: Proceedings of the International Conference on Learning Representations (ICLR); 2016 May 2–4; San Juan, PR, USA.

63. Zhang H, Xu J, Zhang J, Liu Q. Network architecture for optimizing deep deterministic policy gradient algorithms. Comput Intell Neurosci. 2022;2022:1–10. doi:10.1109/icme46284.2020.9102834.
64. Fujimoto S, Hoof HV, Meger D. Addressing function approximation error in actor-critic methods. In: Proceedings of the 35th International Conference on Machine Learning (ICML); 2018 Jul 10–15; Stockholm, Sweden. Vol. 4. p. 2587–601.
65. Han D, Mulyana B, Stankovic V, Cheng S. A survey on deep reinforcement learning algorithms for robotic manipulation. Sensors. 2023;23(7):3762. doi:10.3390/s23073762.
66. Mock JW, Muknahallipatna SS. A comparison of PPO, TD3 and SAC reinforcement algorithms for quadruped walking gait generation. J Intell Learn Syst Appl. 2023;15(1):36–56. doi:10.4236/jilsa.2023.151003.
67. Joshi T, Makker S, Kodamana H, Kandath H. Twin actor twin delayed deep deterministic policy gradient (TATD3) learning for batch process control. Comput Chem Eng. 2021;155(9):107527. doi:10.1016/j.compchemeng.2021.107527.
68. Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O. Proximal policy optimization algorithms. arXiv:1707.06347 2017. doi:10.48550/arxiv.1707.06347.
69. Dhariwal P, Hesse C, Klimov O, Nichol A, Plappert M, Radford A, et al. OpenAI baselines. GitHub. 2017 [cited 2026 Jan 6]. Available from: https://github.com/openai/baselines.
70. Praveena V, Vijayaraj A, Chinnasamy P, Ali I, Alroobaea R, Alyahyan SY, et al. Optimal deep reinforcement learning for intrusion detection in UAVs. Comput Mater Contin. 2022;70(2):2639–53.
71. Rokhade AA, Vithapnavar AS, Amruth S, Shettigar AJV, Supreetha S, Honnavalli PB. Anomaly detection for IoT security: comprehensive survey. In: Proceedings of the IEEE International Conference on Advances in Electronics, Communication, Computing and Intelligent Information Systems (ICAECIS); 2023 Apr 19–21; Bengaluru, India. p. 84–92.
72. Mohammed AB, Chamseddine E, ElAdel A. Enhancing real-time IoT intrusion detection using KAN-based frameworks with SMOTE. J Netw Comput Appl. 2026;249(1):104461. doi:10.1016/j.jnca.2026.104461.
73. Yassen MS, Raghdah AA, Mohammed AB. Employing hybrid ANOVA-RFE with machine and deep learning models for enhanced IoT and IIoT attack detection and classification. Ing Des Syst D'information. 2023;28(4):1003. doi:10.18280/isi.280420.
74. Almofarreh M, Alshahrani A, Alharbi NH, Ahmed AO, Alshahrani H. Boosting cybersecurity: a zero-day attack detection approach using equilibrium optimiser with deep learning model. Comput Model Eng Sci. 2025;145(2):2631–56. doi:10.32604/cmes.2025.070545.
75. Naeem H, Alsirhani A, Alserhani FM, Ullah F, Krejcar O. Augmenting internet of medical things security: deep ensemble integration and methodological fusion. Comput Model Eng Sci. 2024;141(3):2185–223. doi:10.32604/cmes.2024.056308.
76. Kumar G, Alqahtani H. Machine learning techniques for intrusion detection systems in SDN: recent advances, challenges and future directions. Comput Model Eng Sci. 2023;134(1):89–119. doi:10.32604/cmes.2022.020724.
77. Tareq I, Elbagoury BM, El-Regaily S, El-Horbaty EM. Analysis of ToN-IoT, UNSW-NB15, and Edge-IIoT datasets using deep learning in cybersecurity for IoT. Appl Sci. 2022;12(19):9572. doi:10.3390/app12199572.
78. Talpur N, Abdulkadir SJ, Alhussian H, Hasan MH, Aziz N, Bamhdi A. Deep neuro-fuzzy system application trends, challenges, and future perspectives: a systematic survey. Artif Intell Rev. 2023;56(2):865–913. doi:10.1007/s10462-022-10188-3.
79. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg. 2010;8(5):336–41. doi:10.1016/j.ijsu.2010.02.007.
80. Kitchenham B, Brereton OP, Budgen D, Turner M, Bailey J, Linkman S. Systematic literature reviews in software engineering—a systematic literature review. Inf Softw Technol. 2009;51(1):7–15. doi:10.1016/j.infsof.2008.09.009.
81. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA, 2020 statement: an updated guideline for reporting systematic reviews. Int J Surg. 2021;88:n71.
82. Hussain K, Salleh MNM, Cheng S, Shi Y. Metaheuristic research: a comprehensive survey. Artif Intell Rev. 2019;52(4):2191–233. doi:10.1007/s10462-017-9605-z.

83. Yu S, Zhai R, Shen Y, Wu G, Zhang H, Yu S, et al. Deep Q-network-based open-set intrusion detection solution for industrial Internet of Things. IEEE Internet Things J. 2024;11(7):12536–50. doi:10.1109/jiot.2023.3333903.
84. Shen S, Cai C, Li Z, Shen Y, Wu G, Yu S. Deep Q-network-based heuristic intrusion detection against edge-based SIoT zero-day attacks. Appl Soft Comput. 2024;150:111080. doi:10.1016/j.asoc.2023.111080.
85. Yu S, Wang X, Shen Y, Wu G, Yu S, Shen S. Novel intrusion detection strategies with optimal hyper parameters for industrial Internet of Things based on stochastic games and double deep Q-networks. IEEE Internet Things J. 2024;11(17):29132–45. doi:10.1109/jiot.2024.3406386.
86. Shen S, Niu J, Shen Y, Dong J, Ke W, Wang T, et al. MT-KD: multi-teacher knowledge distillation for privacy-aware unlearning in IIoT intrusion detection systems. IEEE Trans Cogn Commun Netw. 2026;12:6344–57. doi:10.1109/tccn.2026.3665917.
87. Hordri NF, Samar A, Yuhaniz SS, Shamsuddin SM. A systematic literature review on features of deep learning in big data analytics. Int J Adv Soft Comput Its Appl. 2017;9(1):32–49.
88. Ma X, Shi W. AESMOTE: adversarial reinforcement learning with SMOTE for anomaly detection. IEEE Trans Netw Sci Eng. 2021;8(2):943–56. doi:10.22215/etd/2020-14010.
89. Rekha H, Siddappa M. Hybrid deep learning model for attack detection in internet of things. Serv Oriented Comput Appl. 2022;16(4):293–312. doi:10.1007/s11761-022-00342-8.
90. Rookard C, Khojandi A. Applying deep reinforcement learning for detection of Internet-of-Things cyber attacks. In: Proceedings of the 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC); 2023 Mar 8–11; Las Vegas, NV, USA. p. 389–95.
91. Baby R, Pooranian Z, Shojafar M, Tafazolli R. A heterogenous IoT attack detection through deep reinforcement learning: a dynamic ML approach. In: Proceedings of the IEEE International Conference on Communications; 2023 May 28–Jun 1; Rome, Italy. p. 479–84.
92. Moudoud H, Cherkaoui S. Empowering security and trust in 5G and beyond: a deep reinforcement learning approach. IEEE Open J Commun Soc. 2023;4:2410–20. doi:10.1109/ojcoms.2023.3313352.
93. Benaddi H, Ibrahimi K, Benslimane A, Qadir J. A deep reinforcement learning based intrusion detection system (DRL-IDS) for securing wireless sensor networks and internet of things. In: Lecture notes of the institute for computer sciences, social informatics and telecommunications engineering (LNICST), Vol. 317. Berlin/Heidelberg, Germany: Springer; 2020. p. 73–87.
94. Saif S, Widyawan W, Ferdiana R. Adaptive deep reinforcement learning: a novel framework for DDoS detection on resource-constrained edge devices. Eng Technol Appl Sci Res. 2026;16(2):32962–70. doi:10.48084/etasr.16594.
95. Wang X, Garg S, Lin H, Hu J, Kaddoum G, Piran MJ, et al. Toward accurate anomaly detection in industrial Internet of Things using hierarchical federated learning. IEEE Internet Things J. 2022;9(10):7110–9. doi:10.1109/jiot.2021.3074382.
96. Shruthi N, Siddesh GK. Trust metric-based anomaly detection via deep deterministic policy gradient reinforcement learning framework. Int J Comput Netw Commun. 2023;15(6):1–25. doi:10.5121/ijcnc.2023.15601.
97. Fahrmann D, Jorek N, Damer N, Kirchbuchner F, Kuijper A. Double deep Q-learning with prioritized experience replay for anomaly detection in smart environments. IEEE Access. 2022;10:60836–48. doi:10.1109/access.2022.3179720.
98. Al-Fawa'Reh M, Abu-Khalaf J, Szewczyk P, Kang JJ. MalBoT-DRL: malware botnet detection using deep reinforcement learning in IoT networks. IEEE Internet Things J. 2024;11(6):9610–29. doi:10.1109/jiot.2023.3324053.
99. Tharewal S, Ashfaque MW, Banu SS, Uma P, Hassen SM, Shabaz M. Intrusion detection system for industrial Internet of Things based on deep reinforcement learning. Wirel Commun Mob Comput. 2022;2022(1):9023719. doi:10.1155/2022/9023719.
100. Ghaly MA, Hannan SA. Protecting software defined networks with IoT and deep reinforcement learning. Int J Intell Syst Appl Eng. 2024;2024(8s):138–47. [cited 2026 Jan 6]. Available from: https://ijisae.org/index.php/IJISAE/article/view/4103.
101. Dake DK, Gadze JD, Klogo GS. DDoS and flash event detection in higher bandwidth SDN-IoT using multiagent reinforcement learning. In: Proceedings of the 2021 International Conference on Computing, Computational Modelling and Applications (ICCMA); 2021 Jun 25–27; Melbourne, VIC, Australia. p. 16–20.

102. Dake DK, Gadze JD, Klogo GS, Nunoo-Mensah H. Multi-agent reinforcement learning framework in SDN-IoT for transient load detection and prevention. Technologies. 2021;9(3):44. doi:10.3390/technologies9030044.
103. Benaddi H, Jouhari M, Ibrahimi K, Othman JB, Amhoud EM. Anomaly detection in industrial IoT using distributional reinforcement learning and generative adversarial networks. Sensors. 2022;22(21):8085. doi:10.3390/s22218085.
104. Li Y, Wang H, Xu G. Federated reinforcement learning-driven multi-task optimization for robust and ethical edge internet of things security. Sci Rep. 2026;16(1):5278. doi:10.1038/s41598-025-34879-3.
105. Khan A, Rizwan M, Bagdasar O, Alabdulatif A, Alamro S, Alnajim A. Deep learning-driven anomaly detection for IoMT-based smart healthcare systems. Comput Model Eng Sci. 2024;141(3):2121–41. doi:10.32604/cmes.2024.054380.
106. Hu Y, Feng Y, Zhao Y, Mao X. IoT-ONDDQN: a detection model based on deep reinforcement learning for IoT data security. Comput Commun. 2025;238(2):108263. doi:10.1016/j.comcom.2025.108263.
107. Ghali AA, Ahmad R, Alhussian H. A framework for mitigating DDoS and DoS attacks in IoT environment using hybrid approach. Electronics. 2021;10(11):1282. doi:10.3390/electronics10111282.
108. Ozdogan E. A comprehensive analysis of the machine learning algorithms in IoT IDS systems. IEEE Access. 2024;12(2):46785–811. doi:10.1109/access.2024.3382539.
109. Alsaedi A, Moustafa N, Tari Z, Mahmood A, Anwar A. TON-IoT telemetry dataset: a new generation dataset of IoT and IIoT for data-driven intrusion detection systems. IEEE Access. 2020;8:165130–50.
110. Meidan Y, Bohadana M, Mathov Y, Mirsky Y, Breitenbacher D, Shabtai A, et al. N-BaIoT—network-based detection of IoT botnet attacks using deep autoencoders. IEEE Pervasive Comput. 2018;17(3):12–22. doi:10.1109/mprv.2018.03367731.
111. Pan M, Li Z, Qian J. Energy-efficient multiuser and multitask computation offloading optimization method. Intell Converg Netw. 2023;4(1):76–92.
112. Saranya T, Sridevi S, Deisy C, Chung TD, Khan MKAA. Performance analysis of machine learning algorithms in intrusion detection system: a review. Procedia Comput Sci. 2020;171(4):1251–60. doi:10.1016/j.procs.2020.04.133.
113. Moustafa N. A new distributed architecture for evaluating AI-based security systems at the edge: network TON_IoT datasets. Sustain Cities Soc. 2021;72(6):102994. doi:10.1016/j.scs.2021.102994.
114. Ekolle ZE, Ochiai H, Kohno R. Collabo: a collaborative machine learning model and its application to the security of heterogeneous medical data in an IoT Network. IEEE Access. 2023;11:142663–75. doi:10.36227/techrxiv.22714864.
115. Clinton UB, Hoque N. MU-IoT: a new IoT intrusion dataset for network and application layer attacks analysis. IEEE Access. 2024;12(6):166068–92. doi:10.1109/access.2024.3494052.
116. Fenanir S, Semchedine F. Smart intrusion detection in IoT edge computing using federated learning. Rev D'Intell Artif. 2023;37(5):1133–45. doi:10.18280/ria.370505.
117. Azimjonov J, Kim T. Designing accurate lightweight intrusion detection systems for IoT networks using fine-tuned linear SVM and feature selectors. Comput Secur. 2024;137:103598. doi:10.1016/j.cose.2023.103598.
118. Guerra-Manzanares A, Medina-Galindo J, Bahsi H, Nomm S. MedBIoT: generation of an IoT Botnet dataset in a medium-sized IoT network. In: Proceedings of the International Conference on Information Systems Security and Privacy; 2020 Feb 25–27; Valletta, Malta. p. 207–18.
119. Hnamte V, Hussain J. An extensive survey on intrusion detection systems: datasets and challenges for modern scenario. In: Proceedings of the 2021 International Conference on Electrical, Control and Instrumentation Engineering (ICECIE); 2021 Nov 27; Kuala Lumpur, Malaysia. p. 1–10.
120. Sharafaldin I, Lashkari AH, Hakak S, Ghorbani AA. Developing realistic distributed denial of service (DDoS) attack dataset and taxonomy. In: Proceedings of the International Carnahan Conference on Security Technology; 2019 Oct 1–3; Chennai, India. p. 1–8. doi:10.1109/CCST.2019.8888419.
121. Hizal S, Çavuşoğlu Ü., Akgün D. IoT-based smart home security system with machine learning models. Acad Platf J Eng Smart Syst. 2024;12(1):28–36.

122. Pahl MO, Aubet FX. All eyes on you: distributed multi-dimensional IoT microservice anomaly detection. In: Proceedings of the 14th International Conference on Network and Service Management (CNSM); 2018 Nov 5–9; Rome, Italy. p. 72–80.
123. Sharma A, Babbar H. Understanding IoT-23 dataset: a benchmark for IoT security analysis. In: Proceedings of the 2024 4th International Conference on Intelligent Technologies (CONIT); 2024 Jun 21–23; Bangalore, India. p. 1–5.
124. Ullah I, Mahmoud QH. Design and development of a deep learning-based model for anomaly detection in IoT networks. IEEE Access. 2021;9:103906–26. doi:10.1109/access.2021.3094024.
125. Wang Y. Application of neural network in abnormal AIS data identification. In: Proceedings of the 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA); 2020 Jun 27–29; Dalian, China. p. 173–9.