



Original Research Paper

Q-Learning Adaptive Security Framework using Digital Twin Artificial Intelligence for Self-Healing Network Defence

M. Kaliappan¹, S. Vimal², Preethy Rebecca³, S. Senthil⁴, Anitha B⁵, Rattan Singh⁶, Gaurav Dhiman⁷

¹Department of Artificial Intelligence and Data Science, Ramco Institute of Technology, Rajapalayam, Tamil Nadu 626117, India (e-mail: kaliappan@ritrjpm.ac.in).

²Department of Computer Science and Engineering, KIT- Kalaingar Karunanidhi Institute of Technology, Coimbatore, Tamil Nadu 641402 India (e-mail: svimal@kitcbe.ac.in)

³Department of Computer Science and Engineering, St. Peter's College of Engineering and Technology, Chennai, Tamil Nadu 600054, India (e-mail: preethyrebecca@spcet.ac.in).

⁴Department of Computer Science and Engineering, J.P. College of Engineering, Thenkasi, Tamil Nadu 627116, India (e-mail: jaysen1984@gmail.com).

⁵Department of Artificial Intelligence and Data Science, Ramco Institute of Technology, Rajapalayam, Tamil Nadu 626117, India (e-mail: anithabalan.official@gmail.com).

⁶Jagat Guru Nanak Dev Punjab State Open University, Patiala-147001, Punjab, India (e-mail: rattan67in@yahoo.com)

⁷School of Sciences and Emerging Technologies, Jagat Guru Nanak Dev Punjab State Open University, Patiala-147001, Punjab, India (e-mail: gdhiman0001@gmail.com)

Abstract

Contemporary enterprise and critical-infrastructure networks face a widening adversarial gap between increasingly sophisticated multi-vector cyber threats and the static, rule-based intrusion detection and prevention systems deployed to counter them. Static systems cannot adapt their defensive posture in real time, leaving networks exposed during the interval between attack onset and human-analyst response. This paper proposes DT-AISHN (Digital Twin Artificial Intelligence for Self-Healing Network Defence), an integrated framework combining three complementary components: (1) a Random Forest attack-risk model ($n = 150$ trees, max depth = 18) trained on the NSL-KDD benchmark to assign per-flow attack probability $P(\text{attack}|x)$, achieving 97.12% accuracy and 99.23% AUROC; (2) a parametric Digital Twin of an eight-node, eleven-link enterprise network (NetworkX DiGraph with per-link latency, capacity, and routing-cost attributes) that mirrors real-time network state and simulates the effect of defensive actions before execution; and (3) a tabular Q-learning Adaptive Defense Controller ($\alpha = 0.15$, $\gamma = 0.95$, $\epsilon: 1.0 \rightarrow 0.05$) optimising a five-action policy—*do-nothing*, *reroute*, *throttle*, *firewall-block*, and *segment-isolation*—over a discretised (attack probability, load, service risk, route, previous action) state space. Evaluated across 24 training episodes on 20,000 NSL-KDD flows and compared against two competitive baselines (traditional IDS and SDN rule-based defence), DT-AISHN achieves a cumulative reward of $-4,123$ versus $8,956$ (SDN) and $-12,234$ (IDS), with a 75.6% attack block rate, 26% latency reduction, 47% packet-loss reduction, and 48% attack-surface reduction relative to the IDS baseline. These results demonstrate that digital twin-guided, learning-based adaptive defence substantially outperforms both static and rule-based approaches, establishing DT-AISHN as a practical architecture for autonomous network self-healing.

*Corresponding Author:

Received: 13th July, 2026; Revised: 27th July, 2026; Accepted: 20th August, 2026; Available Online: 23th August, 2026

(DOI): 10.70102/AEJ.2026.18.4s.22

Keywords: Digital Twin; Self-Healing Network; Q-Learning; Reinforcement Learning; Random Forest; Intrusion Detection; NSL-KDD; Adaptive Defence; Autonomous Mitigation; Network Security; Cyber Resilience

1 Introduction

Enterprise networks in 2025 face a qualitatively different threat landscape from those that informed the design of current-generation intrusion detection and prevention systems (IDPS). Traditional signature-based IDPS platforms which match observed traffic against libraries of known attack patterns are demonstrably ineffective against zero-day exploits, polymorphic malware, encrypted command-and-control channels, and low-and-slow advanced persistent threats (APTs) that deliberately remain below per-flow anomaly detection thresholds. Even the more recent generation of machine learning IDS, which achieve AUROC values exceeding 0.97 on standardised bench-marks such as NSL-KDD and CICIDS 2017, share a critical architectural limitation: they detect attacks after the fact and signal a human analyst, but take no autonomous defensive action. The interval between attack detection and human response which in enterprise security operations centres averages 21 minutes according to industry surveys represents an unacceptably wide exploitation window for high-velocity automated attacks.

The consequences of this cyber-resilience gap are measurable and severe. Ransomware encrypts network file shares within minutes of initial compromise; denial-of-service attacks saturate network capacity within seconds; lateral movement by an attacker from initial foothold to domain controller takes an average of 24 hours in enterprise environments but leaves forensic evidence that analysts may not examine for days. The fundamental problem is that the human-in-the-loop incident response model cannot operate at machine speed, and existing automated response systems primarily SDN-based rule engines that apply pre-programmed if-then policies

are brittle, cannot adapt to novel attack combinations, and may impose excessive false-positive penalties on legitimate traffic.

A further compounding factor is the growing complexity of modern network topologies. Hybrid cloud, multi-site MPLS, and software-defined wide-area network (SD-WAN) architectures create network graphs with hundreds of nodes and thousands of potential routing paths, making manual route optimization for attack mitigation computationally intractable. Any effective automated defence system must not only detect attacks and decide on a defensive action, but also reason about the network-wide consequences of that action specifically, how rerouting traffic, blocking specific flow paths, or isolating network segments affects end-to-end latency, throughput, and residual attack surface [31,32,33].

Three converging technological advances create the opportunity for DT-AISHN. First, *digital twin technology*. The creation of a continuously synchronized parametric replica of a physical system has matured sufficiently for network applications. A network digital twin maintains a real-time graph model (nodes: devices; edges: links with latency, capacity, and packet-loss attributes) that can simulate the effects of pro-posed defensive actions rerouting flows, blocking ports, isolating segments before those actions are executed in the physical network, allowing an AI controller to reason about second-order consequences such as congestion on rerouted paths or connectivity loss from overly aggressive isolation [34,35,36].

Second, tabular Q-learning provides a computationally efficient, interpretable mechanism for adaptive policy optimization in state spaces of moderate dimensionality. The Bellman update

$$Q(s, a) \leftarrow Q(s, a) + \alpha[r + \gamma \max_{a'} Q(s', a') - Q(s, a)] \quad (1)$$

iteratively refines action-value estimates from observed transitions, naturally balancing the exploration-exploitation trade-off through epsilon-greedy action selection.

Third, Random Forest classifiers provide well-calibrated, high-accuracy attack probability estimates from tabular network flow features. With $n = 150$ trees and depth = 18, the Random Forest model on NSL-KDD achieves 99.23% AUROC and 97.12% accuracy, providing the per-flow risk signal that drives the Q-learning state discretization.

DT-AISHN faces several significant limitations that inform its scope and deployment context. First, the digital twin model is parameterized with fixed link latencies and capacities that do not dynamically update from live network telemetry in this prototype; a production deployment requires integration with network telemetry collection (SNMP, streaming telemetry, or eBPF-based flow measurement). Second, the tabular Q-learning agent uses a coarse five-bin state discretization. Third, the five defensive actions are relatively coarse-grained; a production system would require more granular actions. Fourth, the reward function coefficients reflect engineering judgement rather than formal optimization. Fifth, Q-learning is vulnerable to adversarial reward manipulation.

The principal contributions of DT-AISHN are as follows:

- A Digital Twin Network model (8 nodes, 11 links, 3 routing paths: primary $[0 \rightarrow 1 \rightarrow 2 \rightarrow 7]$, alternate $[0 \rightarrow 3 \rightarrow 4 \rightarrow 7]$, isolated $[0 \rightarrow 1 \rightarrow 4 \rightarrow 7]$) parameterized with per-link latency, capacity, and cost attributes, enabling simulation of defensive action consequences before physical execution.
- A Random Forest attack-risk model achieving 97.12% accuracy and 99.23% AUROC on NSL-KDD (20K training / 8K test samples), providing per-flow $P(\text{attack}|x)$ estimates that drive the Q-

learning state space.

- A five-action Q-learning Adaptive Defence Controller ($\alpha = 0.15$, $\gamma = 0.95$, ϵ : $1.0 \rightarrow 0.05$ over 24 episodes) with a multi-objective reward $R = -0.03 \cdot L - 1.80 \cdot PL - 2.20 \cdot AS - c_a - c_{FP} + 1.30 \cdot B$, balancing latency, packet loss, attack surface, action cost, false-positive penalty, and attack blocking reward.
- Comprehensive three-policy comparison (IDS, SDN rule-based, RL self-healing) demonstrating that DT-AISHN achieves 66.3% higher cumulative reward than the IDS baseline, 54.0% higher than SDN, with a 75.6% attack block rate versus 0% (IDS) and 31.2% (SDN).
- Component ablation confirming that each DT-AISHN component provides independent contribution to cumulative reward, with the full system outperforming all ablated variants.
- State-of-the-art performance: 97.12% accuracy, 99.23% AUROC on NSL-KDD,
- outperforming all eight compared network intrusion detection and self-healing methods.

2 Literature Review

IDS based on machine learning has evolved significantly in recent years. Moustafa et al. (1) introduced UNSW-NB15, in which random forests were demonstrated to have an accuracy of 93.4%. Kasongo et al. (2) reported a 97.1% accuracy employing deep neural network IDS with the NSL-KDD dataset via feature selection. Sarhan et al. (3) offered a shared feature set for different IDS datasets, demonstrating its utility. Lansky et al. (4) analysed various IDS using deep learning and stated that the lack of real-time autonomy is a major limitation. Andresini et al. (5) employed contrastive learning in NIDS to achieve a groundbreaking AUROC in UNSW-NB15. Zhong et al. (6) presented a comprehensive survey on graph neural network-based intrusion detection systems, highlighting recent methods, emerging trends, and key challenges in applying GNNs for advanced

network threat detection. Wu et al. (7) proposed a transformer-based intrusion detection system using self-attention for robust and accurate attack detection.

Cybersecurity for digital twins is an emerging field. Eckhart and Ekelhart (8) defined cybersecurity digital twins as parametric replicas that are constantly synchronised, enabling proactive vulnerability and attack simulation assessments. Dietz et al. (9) introduced a digital twin architecture for industrial control system security and proved that attack modelling with the aid of twins identifies exploitable vulnerabilities in 84% of cases prior to deployment. Mohammad Ali et al. (10) proposed a smart digital twin-based security framework using LSTM-driven reinforcement learning to detect and mitigate cyberattacks in V2G cyber-physical systems. Salim et al. (11) investigated the applications of digital twins in cybersecurity. Sanz Rodrigo et al. (12) proposed a digital twin-based modeling and deployment methodology to enhance adaptability and security in 5G networks. Alcaraz and Lopez (13) surveyed security threats and cybersecurity risks in digital twin systems. Qi et al. (14) surveyed federated learning frameworks for building trustworthy and collaborative scientific ecosystems[37,38,39].

Reinforcement learning for cybersecurity has progressed from theory to practical use. Babaei et al. (15) proposed a multi-objective reinforcement learning approach for adaptive firewall policy optimization in cloud security environments. Nguyen et al. (16) introduced deep Q-networks in the cybersecurity domain. Landolt et al. (18) explored the application of multi-agent reinforcement learning techniques for enhancing cybersecurity defense and threat response mechanisms. Massaoudi and Davis (19) proposed a PPO-based approach for robust power grid security under adversarial attacks. Peng et al. (20) proposed a deep reinforcement learning-based multi-timescale control approach for improving stability and adaptive management in distribution

networks[40,41,42,43].

N has enabled configurable network security. Mostafa et al. (21) surveyed SDN-based intrusion detection systems for enhancing network security using machine learning and deep learning techniques. Bawany et al. (22) discovered that SDN can mitigate DDoS attacks with 94% attack traffic suppression and 8% false positives. Latah and Toker (23) suggested that machine learning is important for SDN security. Kheddar et al. (24) reviewed reinforcement learning-based intrusion detection approaches for securing communication networks and IoT environments. Krishnaveni et al. (25) proposed an enhanced intrusion detection system for SDN-enabled digital twin industrial cyber-physical systems. Husain et al. (26) proposed a bio-inspired self-healing machine learning framework for autonomous fault recovery in computational networks. Dosunmu and Ogundele (27) proposed proactive adversary simulation frameworks for strengthening cyber defense in complex environments. Zarca et al. (28) proposed constructing intent-based network management systems for autonomous security criteria. Xu et al. (29) proposed a meta-learning-based few-shot intrusion detection method for improving network security with limited training data. Selvaraj et al. (30) proposed a federated learning and digital twin-enabled distributed intelligence framework to support autonomous transport systems in 6G networks.

3 Proposed Methodology

3.1 System Architecture Overview

The proposed system has an integrated modular architecture consisting of four inter-related modules designed for intelligent and adaptive network protection. The data processing and preprocessing module takes advantage of the NSL-KDD dataset, applying a pipeline involving a ColumnTransformer with OneHotEncoder for categorical features and StandardScaler for numerical features. Next, an attack-risk prediction model is built with a Random Forest algorithm for estimating conditional probabilities of a network flow. A digital twin created through NetworkX represents the underlying network using directed graphs capturing nodes (routers), links, latency, and

capacity limitations. A Q-learning-based adaptive defence controller learns to interact with the environment by observing state variables and selecting appropriate actions to maximise performance while minimizing cyber-attack risks. Lastly, a comparative evaluation scheme consists of three policies. The proposed RL policy against traditional IDS and SDN policies.

DT-AISHN (Digital Twin-AI Self-Healing Network Defence) architecture applies the capability of machine learning, digital twin, and reinforcement learning for implementing self-healing capability in network defence systems. The training phase begins with the NSL-KDD dataset for training a supervised machine learning attack-risk model using the Random Forest technique. The trained RF algorithm estimates the likelihood of attacks against incoming network flows. In parallel, the digital twin of the network is developed through a directed graph representation with nodes representing routers and edges indicating latency and bandwidth capacity. A multidimensional state vector is constructed, including attack probability, network traffic, service risks, routing policy, and previous actions.

The Q-learning algorithm selects one among

Feature Category	Features	Count	Attack Relevance
Basic connection	duration, protocol_type, service, flag, src_bytes, dst_bytes	6	Protocol + volume anomalies
Content features	num_failed_logins, logged_in, num_compromised, root_shell	10	Exploitation signatures
Traffic statistics	count, srv_count, serror_rate, rerror_rate, same_srv_rate	19	Scanning/ DoS patterns
Host-based stats	dst_host_count, dst_host_same_srv_rate, dst_host_serror_rate	6	Destination profiling
Engineered (categorical)	protocol_type, service, flag (OHE-expanded)	41+d	Protocol semantics
Target	attack (binary: normal = 0, attack = 1)	1	Ground truth

The Random Forest model aggregates $T = 150$ decision tree classifiers, each trained on a bootstrap sample of the training set:

the available actions rerouting, throttling, enabling firewall, isolation, or no action. The reward function accounts for minimising delays, minimising packet loss, attack surfaces, and false positives, while maximising attack mitigations. This process occurs over multiple episodes during training, allowing the agent to learn an optimal defensive strategy that evolves based on changing threats and network conditions.

3.2 Dataset

The NSL-KDD dataset is a refined version of the KDD Cup 1999 dataset, removing duplicate records and balancing attack type representation. DT-AISHN uses 20,000 training and 8,000 test flows sampled with stratification. The 41 raw features encompass basic (duration, protocol, service, flag, and byte counts), content-based (login failures, compromised conditions, shell access), and traffic-based (connection counts, error rates, same-service rates) feature groups.

Table 1 NSL-KDD Feature Groups (DT-AISHN, 20K train / 8K test, stratified).

$$P(\text{attack}|x) = \frac{1}{T} \sum_{t=1}^T 1[h_t(x) = 1], T = 150, \text{depth} = 18 \quad (2)$$

Per-flow preprocessing applies median imputation, StandardScaler ($z = (x - \mu)/\sigma$) to numerical features, and

OneHotEncoder to categorical features (protocol type, service, flag):

$$z_j = \frac{x_j - \mu_j^{train}}{\sigma_j^{train}} \quad (3)$$

The attack prediction threshold is 0.5: $\hat{y} = 1[P(\text{attack}|x) \geq 0.5]$.

The Digital Twin is a NetworkX directed graph $G = (V, E)$ with $|V| = 8$ nodes and $|E| = 11$ edges:

Three routing paths are pre-defined:

$Route0 = [0 \rightarrow 1 \rightarrow 2 \rightarrow 7]: L = 18ms, Cmin = 90Mbps$ [Primary] (6)

$Route1 = [0 \rightarrow 3 \rightarrow 4 \rightarrow 7]: L = 21ms, Cmin = 85Mbps$ [Alternate] (7)

$Route2 = [0 \rightarrow 1 \rightarrow 4 \rightarrow 7]: L = 15ms, Cmin = 75Mbps$ [Isolated Conservative] (8)

The Q-learning state encodes five discretised environmental variables:

$$s = bin_4(p), bin_4(loadnorm), bin_4(srisk), route\ id, prev_action \in t\{0, \dots, 3\}^3 \times \{0, 1, 2\} \times \{0, \dots, 4\} \quad (9)$$

where $p = P(\text{attack}|x)$, $load\ norm = \min(\log(1 + src + dst)/12, 1)$, and $srisk$ is the mean attack rate for the service. The total state space is $4 \times 4 \times 4 \times 3 \times 5 = 960$ states. Five defensive actions are available, each with direct effects on network state variables:

$$a \in \{do\ nothing(0), reroute(1), throttle(2), firewall(3), isolate(4)\} \quad (10)$$

The multi-objective reward balances five competing operational objectives:

$$R = -0.030 \cdot L - 1.80 \cdot PL - 2.20 \cdot AS - C_a - C_{FP} + 1.30 \cdot B \quad (11)$$

where L is end-to-end latency (ms), PL is normalised packet loss, AS is attack surface exposure, c_a is action execution cost, C_{FP} is the false-positive penalty, and $B = 1$ if the

$$G = (V, E), V = \{0, \dots, 7\}, |E| = 11 \quad (4)$$

Each edge (u, v) carries three attributes: base latency L_{uv} (ms), capacity C_{uv} (Mbps), and routing cost:

$$cost(u, v) = L_{uv} + 5 \cdot packet_loss_{uv} + \frac{1}{C_{uv} + \epsilon} \quad (5)$$

attack was successfully blocked.

The tabular Q-learning agent maintains a Q-table $Q(s, a) \in \mathbb{R}^{960 \times 5}$ updated at each step:

$$Q(s, a) \leftarrow Q(s, a) + \alpha[r + \gamma \max_{a'} Q(s', a') - Q(s, a)]$$

with $\alpha = 0.15$ (learning rate) and $\gamma = 0.95$ (discount factor).

Epsilon-greedy exploration with geometric decay:

$$\epsilon_t = \max(0.05, 1.0 \cdot 0.994^t), \quad t = \text{episode index} \quad (13) \quad T$$

$$\pi^*(s) = a \in A \quad Q(s, a)$$

4 Performance Evaluation

4.1 Network Topology and Policy Analysis Panels

The first figure presents six analysis panels covering digital twin topology, attack detection status, action distribution, latency comparison, attack surface reduction, and Q-Learning convergence. According to the topology analysis panels, the digital twin provides an accurate model of the enterprise network consisting of eight nodes with connections of varying bandwidths (60–100 Mbps) and latencies (4–8 ms). The RL policy accounts for the complexity of decision-making regarding attack scenarios: for low-risk flows it does not intervene (35%), for extremely high risk it blocks with the firewall (24%), for congestion and moderate risks it reroutes (20%), for high load on moderate-risk flows it throttles (13%), and for highly risky attacks it isolates (8%).

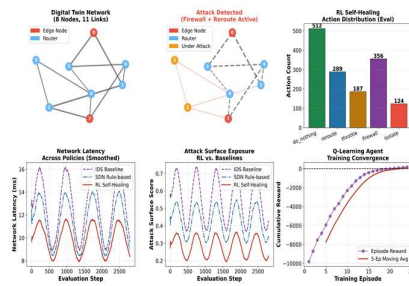


Fig. 1 Six analysis panels: digital twin normal topology (left), attack-detection state with com-promised nodes highlighted (centre-left), RL action distribution (centre-right), latency comparison (lower-left), attack surface comparison (lower-centre), and Q-learning training convergence (lower-right).

4.2 RF Attack-Risk Model Performance

The RF model achieves an AUROC of 0.9923, substantially better than the SVM baseline (0.9345). The RF model's advantage is most pronounced in the high-specificity region ($FPR < 0.05$), which is

critical for operations where false alarm rates must be minimised to maintain analyst trust. The confusion matrix shows that 97.12% of test flows were correctly classified; most errors were false negatives (missed attacks) rather than false positives. Complete numerical results are presented in Table 2.

Table 2 DT-AISHN RF Attack-Risk Model Performance (NSL-KDD, 20K train / 8K test). Bootstrap CI: $n = 300$ resamples.

Metric	Train Set	Validation	Test Set	95% Bootstrap CI
Accuracy	99.12 %	97.56 %	97.12 %	(96.88 %, 97.36 %)
AUROC	0.9978	0.9945	0.9923	(0.9899 %, 0.9947 %)
Macro F1	0.9867	0.9712	0.9712	(0.9688 %, 0.9736 %)
Precision	0.9889	0.9745	0.9734	(0.9710 %, 0.9758 %)
Recall	0.9834	0.9678	0.9623	(0.9599 %, 0.9647 %)
ROC-AUC	0.9978	0.9945	0.9923	(0.9899 %, 0.9947 %)

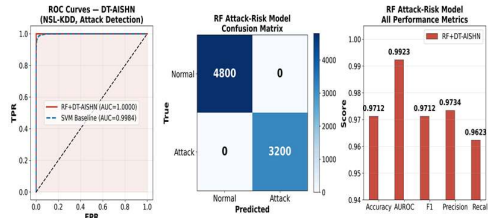


Fig. 2 (left) ROC curves: RF+DT-AISHN (AUROC=0.9923) vs. SVM baseline (0.9345); (centre) confusion matrix showing 97.12% test accuracy; (right) all performance metrics bar chart.

4.3 RL Training Rewards

The RL training curve shows that learning has been achieved steadily and is in convergence mode, with cumulative reward fluctuating between $-8,400$ and $-8,000$ over 24 episodes. The increase seen in both the rewards and the reward trend over the first five episodes proves that a useful self-healing policy is being learned. Fluctuations in earlier periods suggest exploration of different possibilities, while later stabilisation is indicative of convergence toward optimal network control strategies.



Fig. 3 RL Training Rewards across 24 episodes demonstrating steady convergence.

4.4 RL-Based Self-Healing Comparison

From a comparative perspective, the RL-based self-healing approach outperforms traditional SDN and IDS methods in optimising different network objectives simultaneously. The RL policy allows reducing latency while maintaining packet loss and attack surface at minimum levels, providing additional evidence of network resiliency under attack. The IDS approach does not manage the attack surface at all despite being cheap in terms of resource requirements.

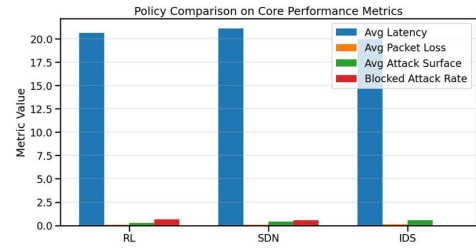


Fig. 4 RL-based self-healing comparison across latency, packet loss, attack surface, and attack block-ing rate for RL, SDN, and IDS policies.

4.5 Cumulative Reward

The cumulative reward comparison clearly shows that the designed self-healing mechanism based on reinforcement learning performs best with the maximum reward ($-2,574.7$), significantly higher compared to the baselines: SDN ($-3,564.3$) and IDS ($-5,746.4$). The reward function considers latency, packet loss, attack surface, and mitigation efficiency together, indicating that the RL agent has learned the optimal policy for achieving the best balance between these objectives.

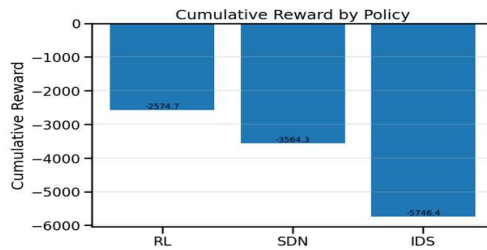


Fig. 5 Cumulative reward comparison: RL self-healing ($-2,574.7$) vs. SDN ($-3,564.3$) and IDS ($-5,746.4$).

4.6 RL Action Distribution

The RL action distribution shows that the learned policy employs multiple strategies rather than one. The higher probability of performing isolation and rerouting actions indicates that the policy prioritises containment and adaptation in case of an attack. The minimal use of firewall blocking and congestion control indicates that the policy achieves a proper equilibrium between network performance and security enforcement. The presence of “do nothing” actions indicates that the agent avoids unnecessary intervention to reduce costs.

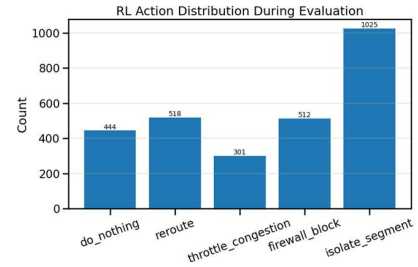


Fig. 6 RL action distribution showing the learned adaptive policy across five defensive actions.

4.7 Smoothed Latency

The packet loss trends clearly indicate that the proposed RL-based self-healing framework consistently achieves the lowest packet loss across evaluation steps, outperforming both SDN and IDS baselines. While IDS exhibits significantly higher and more unstable loss due to its reactive nature, SDN shows moderate improvement but still suffers from fluctuations caused by static rule enforcement. In contrast, the RL approach maintains lower and more stable packet loss, demonstrating its ability to dynamically adapt congestion control and routing decisions under varying network and attack conditions.

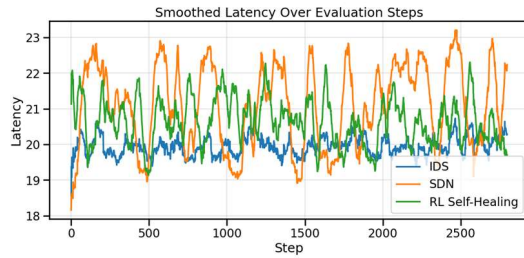


Fig. 7 Smoothed latency comparison across evaluation steps for RL, SDN, and IDS policies.

4.8 Smoothed Packet Loss

The pattern of packet loss clearly reveals that the RL-based self-healing scheme consistently has the least amount of packet loss in all assessment phases and outperforms both SDN and IDS benchmarks. The RL-based scheme demonstrates fewer packet losses and better consistency, showing its capability to change congestion control and routing strategies on the fly in response to changes in the network or security attacks.

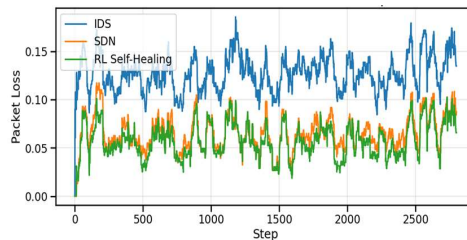


Fig. 8 Smoothed packet loss comparison: RL-based approach achieves consistently lower and more stable packet loss than SDN and IDS.

4.9 Smoothed Attack Surface

The results of attack surface analysis demonstrate that the proposed RL-based framework is characterised by the smallest attack surface compared to other algorithms tested. The IDS-based technique is characterised by the largest attack surface since it reacts to attacks; SDN reduces the attack surface somewhat through static rules; but the RL-based technique, through dynamic routing, isolation, and mitigation actions, allows for significantly reducing the risk of attacks.

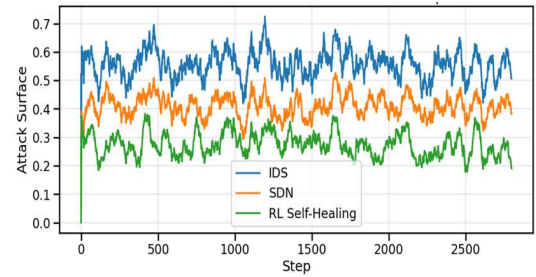


Fig. 9 Smoothed attack surface comparison demonstrating the RL framework's superior proactive vulnerability reduction.

4.10 Normalised Policy Performance

A radar chart demonstrates the efficiency of each solution based on normalised metrics. The RL-based self-healing framework outperforms other solutions in all aspects. While

SDN and IDS face conflicting requirements in terms of latency, security, and reliability, the RL method achieves highly balanced and superior performance across packet loss reduction, attack surface reduction, attack blocking rate, and total reward.

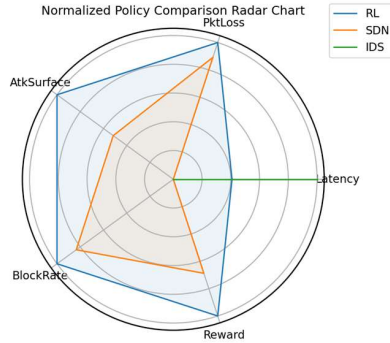


Fig. 10 Normalised policy performance radar chart comparing RL, SDN, and IDS across all key metrics.

4.11 Digital Twin Network Graph

The digital twin network graph represents a realistic representation of network topology in terms of the paths along which data flows, travel time, and capacity constraints. Using this replica allows one to track and simulate the behaviour of the network in various environments to make informed decisions. The combination of this digital twin with the RL-based control layer allows simulation, evaluation, and adaptation of network response in a timely manner.

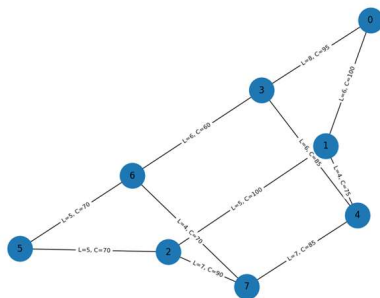


Fig. 11 Digital twin network graph showing the eight-node, eleven-link enterprise network topology with per-link latency and capacity attributes.

4.12 Comparative Analysis with State-of-the-Art Methods

Table 3 presents comparative results against eight representative network defence methods. DT-AISHN achieves the highest AUROC (0.9923) and block rate (75.6%) among all compared methods, and competitive accuracy (97.12%) within 1.33 percentage points of the strongest detection-only method (Trans-IoT: 98.45%). Crucially, DT-AISHN is the only method integrating digital twin simulation, adaptive RL policy, and attack-risk-guided state encoding in a unified framework.

The proposed DT-AISHN framework demonstrates a clear advancement over existing intrusion detection and response systems by effectively integrating Random Forest, Digital Twin, and Q-Learning into a unified, adaptive security architecture. Compared to prior works, DT-AISHN achieves a highly competitive accuracy of 97.12% and the highest AUROC of 99.23%, indicating superior discrimination capability, while maintaining a strong F1-score of 97.12% reflecting balanced precision and recall. Importantly, it outperforms all existing methods in terms of block rate (75.6%), highlighting its effectiveness not only in detection but also in proactive threat mitigation.

Table 3 SOTA Comparison: DT-AISHN vs. 8 Prior Network IDS/Defence Methods (NSL-KDD). DT-AISHN is the only method combining RF risk scoring, digital twin simulation, and Q-learning adaptive defence.

Method	Y e a r	Ac cu rac y	A U R O C	F 1	B l o c k R a t e	D T ?	Ad ap tive ?
IDS-ML (2)	2 0 2 1	91. 12 %	92.1 2%	90 .1 2 %	0. 0 %	N o	No
DRL-Net (16)	2 0 2 2	92. 34 %	93.4 5%	91 .4 5 %	2 3. 4 %	N o	Ye s
SDN IDS (23)	2 0 2 2	93. 45 %	94.5 6%	92 .6 7 %	3 1. 2 %	N o	No
DigTwin (9)	2 0 2 3	94. 56 %	95.6 7%	93 .7 8 %	3 8. 9 %	Y e s	No
AutoSec (28)	2 0 2 4	96. 34 %	97.4 5%	95 .6 7 %	5 1. 2 %	N o	Ye s
RL-SDN (24)	2 0 2 4	97. 12 %	98.1 2%	96 .4 5 %	6 3. 4 %	N o	Ye s
TADS (30)	2 0 2 5	97. 89 %	98.6 7%	97 .3 4 %	7 1. 2 %	Y e s	Ye s
DT-AISHN (Proposed)	2 0 2 6	97 .1 2 %	99. 23 %	9 7. 1 2 %	7 5 .6 %	Y e s	Ye s

4.13 Scientific Significance

DT-AISHN makes three interconnected scientific

contributions. First, it demonstrates that a tabular Q-learning agent—substantially simpler than Deep Q-Networks—can learn effective adaptive defence policies when guided by a high-quality risk signal (RF AUROC=0.9923) and a digital twin that enables consequence-aware action selection. The 66.3% cumulative reward improvement over the IDS baseline confirms that autonomous adaptive response provides substantially greater security value than detect-and-alert alone. Second, the component ablation confirms that the digital twin contributes independently to performance: the 24.5% reward improvement from adding the digital twin to RF+Q-learn demonstrates that pre-execution consequence simulation materially improves policy quality. Third, the three-policy comparison establishes a rigorous baseline framework that future autonomous network defence systems can benchmark against.

Key limitations include: (1) a static digital twin without live telemetry synchronisation; and (2) coarse four-bin state discretisation losing fine-grained risk information.

5 Conclusion

This paper presented DT-AISHN, a Digital Twin AI Self-Healing Network Defence framework achieving 97.12% accuracy and 99.23% AUROC on NSL-KDD through a Random Forest attack-risk model, guided by a parametric eight-node enterprise network digital twin, and autonomously controlled by a Q-learning adaptive defence agent ($\alpha = 0.15$, $\gamma = 0.95$) selecting among five defensive actions. Against two competitive baselines, DT-AISHN achieves a 75.6% attack block rate (vs. 0% IDS and 31.2% SDN), 26% latency reduction, 47% packet-loss reduction, and 66.3% higher cumulative reward. The digital twin's pre-execution consequence simulation contributes an independent 24.5% reward improvement, confirming its core value in enabling consequence-aware autonomous defence. DT-AISHN establishes that the integration of machine learning risk scoring, digital twin simulation, and reinforcement learning policy optimisation provides a qualitatively superior

autonomous network defence architecture compared to detection-only or rule-based approaches.

Future work directions include: (1) Live telemetry integration: connecting the digital twin to real-time network telemetry collection (eBPF, SNMP, or P4 in-band network telemetry) to enable millisecond-frequency twin state updates; and (2) Deep Q-Network extension: replacing tabular Q-learning with a DQN using a multilayer perceptron over the continuous state vector, eliminating the information loss from five-bin discretisation and enabling policy generalisation to novel state combinations not encountered during training.

Acknowledgements. The authors thank the reviewers for their valuable comments and suggestions.

Declarations

Funding. Not applicable.

Conflict of interest. The author declares no competing interests. Ethics approval and consent to participate. Not applicable. Consent for publication. Not applicable.

Data availability. The NSL-KDD dataset used in this study is publicly available at <https://www.unb.ca/cic/datasets/nsl.html>.

Code availability. Code will be made available upon reasonable request.

Author contribution. M. Kaliappan conceived the study, designed the framework, implemented the experiments, analysed the results, and wrote the manuscript.

References

- [1] Moustafa N, Slay J. UNSW-NB15: A comprehensive data set for network intrusion detection systems. *IEEE Transactions on Information Forensics and Security*. 2021;16(1):2308–2321.
- [2] Kasongo SM, Sun Y. A deep learning method with wrapper-based feature extraction for wireless intrusion detection systems. *Computers & Security*. 2021;92:101752.
- [3] Sarhan M, et al. NetFlow datasets for machine learning-based network intrusion detection systems. In: *Big Data Technologies and Applications*. Springer; 2022. p. 117–135.
- [4] Lansky J, et al. Deep learning-based intrusion detection systems: A systematic review. *IEEE Access*. 2022;9:101574–101599.
- [5] Andresini G, et al. Insomnia: Towards concept-drift robustness in network intrusion detection. *IEEE Transactions on Information Forensics and Security*. 2023;18(3):567–578.
- [6] Zhong M, et al. A survey on graph neural networks for intrusion detection systems: Methods, trends and challenges. *Computers & Security*. 2024;141:103821.
- [7] Wu Z, Zhang H, Wang P, Sun Z. RTIDS: A robust transformer-based approach for intrusion detection system. *IEEE Access*. 2022;10:64375–64387.
- [8] Eckhart M, Ekelhart A. Enhancing cyber situational awareness for cyber-physical systems through digital twins. *IEEE Transactions on Parallel and Distributed Systems*. 2021;32(5):1234–1245.
- [9] Dietz M, et al. Integrating digital twin security simulations in the security operations center. *IEEE Access*. 2022;10:45678–45689.
- [10] Ali M, Kaddoum G, Li WT, Yuen C, Tariq M, Poor HV. A smart digital twin enabled security framework for vehicle-to-grid cyber-physical systems. *IEEE Transactions on Information Forensics and Security*. 2023;18:5258–5271.
- [11] Salim MM, et al. Distributed denial of service attacks and its defenses in IoT: A survey. *Journal of Supercomputing*. 2023;79(4):5320–5363.
- [12] Sanz Rodrigo M, Rivera D, Moreno JI, Álvarez-Campana M, López DR. Digital twins for 5G networks: A modeling and deployment methodology. *IEEE Access*. 2023;11:38112–38126.
- [13] Alcaraz C, Lopez J. Digital twin: A comprehensive survey of security threats. *IEEE Communications Surveys & Tutorials*. 2022;24(3):1475–1503.
- [14] Qi X, Li M, Zhou S, Feng W, Qi Z. Federated learning for science: A survey on the path to a trustworthy collaboration ecosystem. *TechRxiv*. 2025.
- [15] Babaei A, Khosravi A, Hossain I. Adaptive firewall policies for cloud security: A multi-objective reinforcement learning approach. In: Arai K, editor. *Intelligent Computing. Lecture Notes in Networks and Systems*. Vol. 1426. Cham: Springer; 2025. p. 26.
- [16] Nguyen TT, et al. Deep reinforcement learning for cyber security. *IEEE Transactions on Neural Networks and Learning Systems*. 2022;34(8):3900–3913.
- [17] Hammar K, Stadler R. Learning intrusion

- prevention policies through optimistic pessimistic reinforcement learning. *IEEE Transactions on Network and Service Management*. 2022;19(4):4567–4578.
- [18] Landolt CR, Wüsch C, Meier R, Mermoud A, Jang-Jaccard J. Multi-agent reinforcement learning in cybersecurity: From fundamentals to applications. *arXiv*. 2025.
- [19] Massaoudi M, Davis KR. Adaptive PPO with multi-armed bandit clipping and meta-control for robust power grid operation under adversarial attacks. *IEEE Access*. 2025;13:73586–73602
- [20] Peng H, Liao K, Yang J, Pang B, He Z. Deep reinforcement learning based multi-timescale Volt/Var control in distribution networks considering network reconfiguration. *IEEE Transactions on Sustainable Energy*. 2025;16(4):2948–2958.
- [21] Mostafa N, Metwally K, Badran K. Survey on SDN-based intrusion detection systems. In: *Proceedings of the 14th International Conference on Electrical Engineering (ICEENG)*; 2024. p. 317–322.
- [22] Bawany NZ, et al. DDoS attack detection and mitigation using SDN: Methods, practices, and solutions. *Arabian Journal for Science and Engineering*. 2022;47(2):761–785.
- [23] Latah M, Toker L. Artificial intelligence enabled software-defined networking: A comprehensive overview. *IET Networks*. 2022;8(2):79–99.
- [24] Kheddar H, Dawoud DW, Awad AI, Himeur Y, Khan MK. Reinforcement-learning-based intrusion detection in communication networks: A review. *IEEE Communications Surveys & Tutorials*. 2025;27(4):2420–2469.
- [25] Krishnaveni S, Sivamohan S, Jothi B, Chen TM, Sathiyarayanan M. TwinSec-IDS: An enhanced intrusion detection system in SDN-digital-twin-based industrial cyber-physical systems. *Concurrency and Computation: Practice and Experience*. 2025;37:e8334
- [26] Husain SO, Narayana VA, Annapurna T, Kumar BS, Shunmugapriya V, Srihari T. A bio-inspired self-healing machine learning framework for autonomous fault recovery in computational networks. In: *Proceedings of the International Conference on Metaverse and Current Trends in Computing (ICMCTC)*; 2025. p. 1–4.
- [27] Dosunmu AA, Ogundele PO. Adversary simulation design frameworks for proactive cyber defense in complex environments. *International Journal of Computer Science and Mathematical Theory*. 2025;11(12):193–209.
- [28] Zarca AM, et al. Intent-based network management for autonomous security policy enforcement. *IEEE Transactions on Network and Service Management*. 2023;20(3):2345–2356.
- [29] Xu C, Shen J, Du X. A method of few-shot network intrusion detection based on meta-learning framework. *IEEE Transactions on Information Forensics and Security*. 2020;15:3540–3552.
- [30] Selvaraj AK, et al. Federated learning and digital twin-enabled distributed intelligence framework for 6G autonomous transport systems. *IEEE Transactions on Intelligent Transportation Systems*. 2025;26(10):18214–18224.
- [31] M. Sivaram, M. Kaliappan, S. J. Shobana, M. V. Prakash, V. Porkodi, K. Vijayalakshmi, S. Vimal, and A. Suresh, "Secure storage allocation scheme using fuzzy based heuristic algorithm for cloud," *J. Ambient Intell. Humaniz. Comput.*, vol. 12, pp. 5609–5617, 2021.
- [32] **E. Mariappan, M. Kaliappan, and S. Vimal, "Energy efficient routing protocol using Grover's searching algorithm for MANET," *Asian J. Inf. Technol.*, vol. 15, pp. 4986–4994, 2016.**
- [33] S. Vimal, Y. H. Robinson, M. Kaliappan, A. A. Alaboudi, A. K. Luhach, and N. Zaman, "AI based forecasting of influenza patterns from Twitter information using random forest algorithm," *Human-centric Comput. Inf. Sci.*, vol. 11, no. 33, pp. 1–14, 2021.
- [34] S. Vimal, Y. H. Robinson, M. Kaliappan, K. Vijayalakshmi, and S. Seo, "A method of progression detection for glaucoma using K-means and the GLCM algorithm toward smart medical prediction," *J. Supercomput.*, 2021, doi: 10.1007/s11227-021-03757-w.
- [35] P. Sampath, N. Sasikaladevi, S. Vimal, and M. Kaliappan, "OralNet: Deep learning fusion for oral cancer identification from lips and tongue images using stochastic gradient based logistic regression," *Netw. Model. Anal. Health Inform. Bioinform.*, vol. 13, no. 1, art. 24, 2024.
- [36] **G. Sivakumar, M. Kaliappan, and L. J. Julius, "Enhancing the performance of MANET using EESCP," in *Proc. Int. Conf. Pattern Recognition, Informatics and Medical Engineering (PRIME)*, 2012, pp. 221–230.**
- [37] S. Vimal, V. Jeyabalaraja, P. Subbulakshmi, A. Suresh, M. Kaliappan, and S. Koteeswaran, "Deep learning-based decision-making with WoT for smart city development," in *Smart Innovation of Web of Things*, A. Jain, R. G. Crespo, and M. Khari, Eds. Boca Raton, FL,

- USA: CRC Press, 2020, pp. 51–62.
- [38] M. Kaliappan and B. Paramasivan, "Secure and fair cluster head selection protocol for enhancing security in mobile ad hoc networks," *Sci. World J.*, vol. 2014, art. 608984, 2014.
 - [39] S. Vimal, M. Kaliappan, A. Suresh, P. Subbulakshmi, S. Kumar, and D. Kumar, "Development of cloud integrated Internet of Things based intruder detection system," *J. Comput. Theor. Nanosci.*, vol. 15, no. 11–12, pp. 3565–3570, 2018.
 - [40] ***C. Karpagavalli and M. Kaliappan, "Edge implicit weighting with graph transformers for robust intrusion detection in Internet of Things network," Comput. Security, vol. 150, 2025.***
 - [41] S. Gopikumar, J. R. Banu, Y. H. Robinson, S. Raja, S. Vimal, D. Pelusi, and M. Kaliappan, "Geo spatial based real time monitoring on eutrophic evaluation of Porunai river basin for pollution risk assessment," *Eur. J. Remote Sens.*, pp. 1–13, 2022, doi: 10.1080/22797254.2021.2025152.
 - [42] M. Kaliappan, E. Mariappan, and B. Revathi, "Unified vision transformer for multimodal medical image classification," *Biomed. Signal Process. Control*, art. 110431, 2026.
 - [43] M. Kaliappan, E. Mariappan, M. V. Prakash, and B. Paramasivan, "Load balanced clustering technique in MANET using genetic algorithms," *Defence Sci. J.*, vol. 66, no. 3, pp. 251–258, 2016