Journal of Computer Science

Research Article

LSTM-Based AI Model for Sinkhole Attack Detection With Legal Basis in an Ecuadorian Public Institution

Estefanía Alejandra Mora Parra[1], Rubén Nogales Portero[1], Moisés Toapanta T.[2], Estefanía Monge Martínez[2], Santiago Vayas Castro[2], Jeanette Elizabeth Jordán Buenaño[2], Juan Escobar Naranjo[1], Diego Gustavo Andrade Armas[3] and Rodrigo Del Pozo Durango[4]

[1]Carrera Ingeniería en Sistemas, Electrónica e Industrial, Universidad Técnica de Ambato, Av. Chasquis, Ambato, Ecuador
[2]Carrera de Derecho, Universidad Técnica de Ambato, Av. Chasquis, Ambato, Ecuador
[3]Subsistema de Posgrados, U Centro de Estudios de Seguridad (CESEG), Universidad de Santiago de Compostela (USC), Av. Doctor Ángel Echeverri s/n, Compostela, Spain
[4]Carrera de Software y Tecnologías de la Información, Universidad Estatal de Bolívar, Av. Ernesto Che Guevara s/n y Av. Gabriel Secaira, Guaranda, Ecuador

Article history
Received: 30-07-2025
Revised: 16-10-2025
Accepted: 25-11-2025

Corresponding Author:
Segundo Moisés Toapanta Toapanta
Carrera Ingeniería en Sistemas, Electrónica e Industrial, Universidad Técnica de Ambato, Av. Chasquis, Ambato, Ecuador
Email: sm.toapanta@uta.edu.ec

Abstract: Wireless Sensor Networks (WSN) are an essential component of the Internet of Things (IoT). However, their decentralized nature, data transmission over unencrypted channels, and the physical exposure of nodes make them especially vulnerable to attacks, among which the sinkhole attack stands out. This research aims to develop a machine learning–based model to detect sinkhole-type attacks in wireless sensor networks, with the purpose of strengthening the security and resilience of a public institution. The deductive method was used to analyze the legal framework and technical background, and the experimental method was used for the design and evaluation of the detection model. The results obtained include an LSTM model trained on data from a network simulation conducted in Contiki with the Cooja simulator. The model achieved 98 accuracy, 96 precision, 97 recall, and a 96% F1-score. It was concluded that this neural network based model offers a promising solution to enhance WSN security in IoT environments.

Keywords: Sinkhole Attack, Legal Basis, Machine Learning, WSN, Public Institution

Introduction

The Internet of Things (IoT) is recognized as a key technology impacting sectors such as health, environment, transportation, industry, and agriculture by enabling more efficient, interactive, and autonomous infrastructures. Within this context, Wireless Sensor Networks (WSN) constitute an essential component of the IoT by facilitating the effective collection and transmission of real-time data (Bakar et al., 2023). With the accelerated advancement of communication network technologies, public organizations including those dedicated to higher education are incorporating hybrid network architectures to improve connectivity, automate processes, and optimize institutional management (Guo and Xia, 2022). In urban infrastructures, WSNs also play a fundamental role in vehicular traffic monitoring, enabling the continuous collection of real-time data to inform users about traffic conditions, identify critical points in the road network, and estimate arrival times, thereby contributing to more efficient urban mobility (Choudhary, 2024). As these networks expand and integrate with more complex systems such as data centers, gateways, and multiple smart devices the communication architecture becomes increasingly diverse, interconnected, and sophisticated, posing new challenges in its management and operation (Huang et al., 2024).

Despite their benefits, WSNs present multiple vulnerabilities that make them attractive targets for malicious actors. Common weaknesses include the absence of centralized infrastructure, communication over unencrypted channels, and the physical exposure of nodes, which permits direct or remote manipulation (Oztoprak and Ozkan, 2025). A further source of vulnerability lies in the Routing Protocol for Low-Power and Lossy Networks (RPL), the predominant standard for routing in IoT-WSN environments. RPL organizes the network into a Directed Acyclic Graph (DODAG) based

SCIENCE Publications

on the metric of rank, which determines the relative distance of each node to the root. However, the protocol lacks robust authentication mechanisms and depends on the honest exchange of control messages (DIO, DAO), making it particularly susceptible to rank manipulation (Hachemi et al., 2020).

This weakness directly enables one of the most dangerous threats to WSNs: the sinkhole attack. In this attack, a malicious node advertises itself as the shortest or most reliable route to the base node, thus attracting most of the data traffic. Once in control of the traffic, the attacker can drop, modify, or selectively forward packets, severely affecting the availability and veracity of transmitted information (Sejaphala and Velempini, 2020). This attack exploits weaknesses in routing protocols and the lack of mutual authentication between nodes, and it can spread quickly if not detected in time. Effective detection of sinkhole attacks requires advanced mechanisms, as their effects can be silent but highly disruptive, particularly in critical institutional environments (Abdul-Bari Ahmed Mohammed et al., 2024).

In 2022, Ecuador obtained a score of 53.25 on the National Cyber Security Index (NCSI), ranking 67th worldwide, reflecting the country’s cybersecurity preparedness in terms of public policies, institutional capacities, and national strategies (Academy e-Governance, 2022). This position underscores the importance of reinforcing cybersecurity in national institutions, particularly in scenarios where WSNs are deployed for critical services.

The increasing sophistication of attacks targeting WSNs, such as sinkhole attacks, highlights the limitations of traditional security approaches based on static rules or manual detection mechanisms. While useful in certain circumstances, these methods prove insufficient in dynamic environments and against evolving threats (Zhou et al., 2021). In this context, Artificial Intelligence (AI) emerges as an effective and necessary alternative to strengthen WSN security, especially in institutional settings where data availability and accuracy are critical. AI models can detect anomalous patterns in real time, adapt to node behavior, and anticipate potential attacks using supervised and unsupervised learning techniques (Sivagaminathan et al., 2023). Furthermore, hybrid approaches that combine neural networks, decision trees, and optimization algorithms have demonstrated high performance in early attack detection, minimizing false positives, and enhancing network resilience (Nandhini et al., 2024).

How effective is an artificial intelligence model in detecting sinkhole attacks on wireless sensor networks in the context of a public institution in Ecuador?

Simulation results show that the proposed AI-based model demonstrates a highly effective detection capability against sinkhole attacks in WSNs, with performance equal to or better than models previously reported in the literature.

The objective in this research is develop a machine learning–based model to detect sinkhole-type attacks in Wireless Sensor Networks, with the aim of strengthening the security and resilience of a public higher education institution.

As a result, a model was obtained for detecting sinkhole attacks with 98% accuracy and 97% recall in a simulation environment with Cooja.

It is concluded that employing machine learning on network-layer metrics is an effective strategy for detecting sinkhole attacks in WSNs.

Materials

A distributed IoT attack detection framework based on Deep Learning (DL), deployed on fog nodes, was proposed. Six models were evaluated across five different datasets, with LSTM standing out by achieving a detection rate of 99.97and a precision of 99.96 in binary classification. In multiclass classification, precision reached 99.65% (Samy et al., 2020). A new type of internal attack in IoT networks, called the loophole attack, is presented, which exploits vulnerabilities in the RPL routing protocol. The attack was implemented on the Contiki operating system using the Cooja simulator, demonstrating a significant impact on network performance, such as increased energy consumption and packet loss. To detect it, a machine learning based model was proposed, evaluating multiple algorithms (such as Random Forest and XGBoost) with high accuracy (Chowdhury et al., 2021). AIEMLA, a neural network–based model for detecting RPL attacks such as hello-flood, decreased-rank, and increased-version was validated using hold-out and K-fold cross-validation, achieving 100% precision, recall, and F1-score across all scenarios (Sharma and Verma, 2021). ELNIDS, an ensemble learning based IDS for RPL attacks using the RPL-NIDDS17 dataset, achieved a best performance of 94.5% precision and 0.98 AUC with Boosted Trees, while Subspace Discriminant reached 77.8% precision and 0.87 AUC (Verma and Ranga, 2019). PSO-optimized classifiers for IoT 6LoWPAN cyberattack detection employing KNN, SVM, and Naive Bayes attained up to 99.6% precision with KNN (Maleh et al., 2021). An improved ACO-based technique incorporating a hash table reached 96% detection rate for sinkhole attacks in WSNs, reducing false alarm rates (Nwankwo et al., 2021). RUDRA, a unified classifier combining LSTM, RNN, and ELM, obtained up to 96.89% precision and 95.45% F1-score on blackhole attacks, outperforming previous methods by over 10% (Sridevi and Anandan, 2020). GBCRP, a secure routing protocol using blockchain and GANs, integrated an IDS with FDGAN to enhance military WSN security, energy efficiency, and performance (Rajasoundaran et al., 2021). An approach is proposed to trace attacks in 5G networks using the Dominance-based Rough Set Approach (DRSA) and Formal Concept Analysis (FCA). The system achieves 98.19% accuracy and 93.33% recall. DRSA reduces irrelevant attributes, and FCA identifies conceptual patterns (Acharjya and Ahmed, 2021). An advanced intrusion detection system for hierarchical WSNs is developed by combining Random Forest, XGBoost, and K-Means. The hybrid approach achieved 100% precision, recall, and F1-score in binary classification on the UNSW-NB15 and CICIDS2017 datasets (Gebremariam et al., 2023a). A hybrid approach based on optimized neural networks (EO-NN, PSO-NN, and SCO-LSTM) combined with threat intelligence was proposed for the detection of attacks in IoT-WSN, including false data injection, brute force, and hybrid brute force. The SCO-LSTM model achieved an accuracy of 99.89% on a custom dataset generated in a test-bed (Nandhini et al., 2024). A trust-based anomaly detection scheme for WSNs is proposed, combining behavior evaluation and trust metrics. The system uses dynamic update rules and distributed detection among neighboring nodes (Ahmadi and Javidan, 2024). A system for simultaneous localization and detection of multiple attacks in WSNs using a multilayer perceptron (MLP) neural network is developed. The model was trained and evaluated with three different datasets, achieving up to 100% accuracy. It effectively detects attacks such as sinkhole and wormhole (Gebremariam et al., 2023b). A model based on Convolutional Neural Networks (CNN) is proposed to detect routing attacks in IoT healthcare environments. The system was evaluated in Cooja. It effectively detects sinkhole and blackhole attacks (Kamel and Elhamayed, 2020). RFTRUST, a trust-aware security mechanism to detect sinkhole attacks in RPL-based IoT environments, is presented. It uses the Random Forest algorithm along with subjective logic to classify nodes based on metrics such as RSSI and delay. It was validated in the Cooja simulator with Contiki nodes (Prathapchandran and Janani, 2021). An intelligent intrusion detection system for RPL attacks in IoT based on machine learning is proposed. The model combines a genetic feature selection algorithm with a fuzzy k-NN classifier. It detects attacks such as sinkhole, blackhole, version, and selective forwarding (Raghavendra et al., 2022). A model for detecting anomalies induced by routing attacks in IoT networks using a hybrid RBM-LSTM architecture is presented. The system predicts the normal behavior of the network and detects deviations caused by attacks such as the version number attack. It was evaluated in Cooja (Sahay et al., 2024). A machine learning framework to classify network anomalies in IoT-based Cyber-Physical Systems (CPS) is proposed, differentiating between faults and attacks. The study focuses on an Energy-Aware Smart Home (EASH) environment, analyzing the behavior of communication channels. The system was evaluated in simulated and real-time test environments. Algorithms such as J48, Naive Bayes, MLP, and multinomial logistic regression were used (Tertytchny et al., 2020). A framework for improving QoS and security for RPL in IoT networks through machine learning is proposed. The system integrates Random Forest for intrusion detection and reinforcement learning for adaptive decision-making. A detection rate above 95% was achieved, with improvements in throughput, latency, and energy efficiency (Wakili et al., 2024). A secure localization technique in WSNs against routing attacks is developed using hybrid machine learning models. Algorithms such as K-Means and Random Forest are combined to identify malicious nodes and estimate their location. The system achieved 100% detection accuracy. It was validated with benchmark datasets in attack scenarios such as sinkhole and wormhole (Gebremariam et al., 2023c).

Methods

In this research, the IMRYD research methodology was applied, the deductive method for the analysis of the legal framework and the technical background, and the experimental method for the design and evaluation of the detection model. From this, a three-phase methodology was defined that structures the necessary activities to achieve the stated objective and can be replicated in similar scenarios.

Phase 1: Review of the Institutional Legal and Regulatory Framework

The first phase consisted of identifying technical, legal, and administrative guidelines relevant to the comprehensive management of cybersecurity in a public organization. For this purpose, national and international standards, institutional policies, and widely recognized best practices in the Ecuadorian public sector were reviewed. Table 1 presents a structured summary of the key elements associated with each of these three approaches, including their respective reference sources, following CEDIA’s proposal on legal protection and institutional cybersecurity (Corporación Ecuatoriana para el Desarrollo de la Investigación y la Academia [CEDIA], 2024).

Table 1: Comprehensive Cybersecurity Approach in a Public Organization

Aspect	Description	Key Elements	Reference
Technical	Implementation of controls and tools to protect the IT infrastructure and data	Firewalls; IDS/IPS; data encryption; network segmentation; multi-factor authentication; vulnerability management; backups; network monitoring	ISO/IEC 27002:2022; NIST SP 800-53 Rev. 5
Legal	Regulatory compliance and adoption of national and international legal frameworks on data protection and cybercrime	Constitution of the Republic of Ecuador; Organic Law on Personal Data Protection; COIP; Organic Telecommunications Law; Digital Transformation Law	National Assembly (2021); Constitution (2008)
Administrative	Policies, procedures, and governance structures to manage information security in the public sector	Information Security Policy; business continuity plans; staff training; audits; security committees; EGSI	ISO/IEC 27001; EGSI (Ministerial Agreement); National Cybersecurity Strategy 2022–2025

Phase 2: Systematic Review of Technical Background

This phase consisted of a systematic review of scientific literature related to attack detection in Wireless Sensor Networks (WSN), with emphasis on routing attacks such as the sinkhole. The objective was to identify methodological approaches, artificial intelligence algorithms applied, datasets used, evaluation metrics, and limitations reported in previous research.

To this end, twenty academic articles published between 2020 and 2024 were selected, prioritizing those that presented experimental proposals, clear quantitative metrics, and application of machine learning or deep learning techniques in the context of cybersecurity for IoT or WSN environments. The selection was made from indexed databases such as IEEE Xplore, Taylor and Francis SpringerLink, ScienceDirect, ACM Digital Library, and Wiley Online Library.

As a result of the documentary analysis, a comparative synthesis of the most relevant technical background was elaborated, considering methodological approach, applied algorithm, validation environment, performance metrics, and limitations. This information is presented in Table 2, which allows identifying trends, gaps, and best practices in attack detection in WSN and IoT networks, serving as the basis for the experimental design developed in the next phase. The comparative analysis presented in Table 2 highlights several key trends in the detection of attacks within WSN/IoT networks. A predominance of supervised learning algorithms such as Random Forest, XGBoost, and ANN can be observed, consistently achieving detection rates above 95% in simulated environments. In recent years, however, deep learning models, including LSTM, CNN, and hybrid RBM-LSTM architectures, have gained increasing prominence, reporting near-perfect performance in binary classification tasks, albeit at the expense of higher computational costs and more demanding training requirements. With respect to validation settings, most studies continue to rely on simulations with Contiki/Cooja or NS-3, whereas real-hardware deployments remain scarce, underscoring the gap between experimental validation and practical applicability in resource-constrained networks. Furthermore, the absence of public and standardized datasets for RPL attacks persists, limiting replicability and hindering rigorous comparison across different approaches. Finally, a growing trend toward hybrid solutions can be discerned, integrating machine learning with metaheuristic optimization, blockchain, or reinforcement learning, with the aim of enhancing resilience and reducing false positives; nevertheless, these approaches continue to face significant challenges related to scalability and implementation costs in real-world environments.

Phase 3: Experimental Design of the Detection Model

Simulation

The experimental phase was designed to evaluate the performance of the detection model under diverse network contexts, simulated in Contiki (v3.0) using the Cooja emulator. The scenarios incorporated variations in node density, traffic load, number of attackers, and attacker mobility.

Scenario 1: Baseline topology of 34 nodes*:* A network composed of 34 Sky motes was configured, consisting of one RPL root node, 30 normal nodes, and 3 malicious nodes executing the sinkhole attack through rank manipulation. This scenario served as the baseline reference for subsequent experiments.

Scenario 2: Stress test with 100 nodes: To assess scalability, a topology of 100 nodes was implemented with a single sinkhole, one root node, and 50 normal nodes generating intensive traffic. This scenario enabled the observation of the impact of node density and heavy traffic conditions on the detection performance.

Table 2: Comparative summary of machine learning and artificial intelligence–based approaches applied to intrusion detection in IoT/WSN networks

#	Reference	Technique	Algorithm	Dataset / Simulation	Metrics	Limitations
1	(Samy et al., 2020)	Deep Learning	LSTM	UNSW-NB15, CICIDS-2017, RPLNIDS-2017, N_BaIoT-2018, NSL-KDD	Accuracy: 99.85% (N_BaIoT), 98.82% (UNSW-NB15), ~99% (others)	Edge-layer labeling issues
2	(Chowdhury et al., 2021)	Supervised ML	XGBoost	Simulation (Contiki, Cooja)	Accuracy: 93.8%, Precision: 98.2%, Recall: 89.1%, F1: 93.4%	No specific limitations specified
3	(Sharma and Verma, 2021)	Supervised ML	ANN	Simulation (Contiki, Cooja)	Accuracy: 100%	Cross-validation requires more resources
4	(Maleh et al., 2021)	Supervised ML + optimization (PSO)	RF-PSO	6LoWPAN simulation (Contiki, Cooja)	Hello Flood: Precision: 99.8%, Recall: 99.5%, Accuracy: 99.8%; Wormhole: Precision, Recall, and Accuracy: 100%; Sinkhole: Precision: 99.6%, Recall: 99.4%, Accuracy: 99.7%.	Lack of public and standardized datasets in IoT-6LoWPAN
5	(Nwankwo et al., 2021)	Supervised ML + optimization (ACO)	EACO	Simulation of 50 malicious nodes (NS-3.30.1)	DR: 96%, FPR: 1%, PDR: 0.9, Latency: 153 ms, Throughput: 8.72 kbps	Evaluation is limited to simulation, not validated in real environments
6	(Sridevi and Anandan, 2020)	Hybrid Deep Learning	LSTM + ELM	Proprietary dataset generated in a real environment (Arduino/NodeMCU + TI CC2540)	Accuracy: 99.5%, Sensitivity and Specificity evaluated	It is recommended to evaluate more attack types to verify model generalization
7	(Rajasoundaran et al., 2021)	Deep Learning + Blockchain	FD-GAN, GBCRP	NS-3 simulation, KDD'99 dataset	Blockchain reconstruction: 97.8%, Data loss: 1.32%–4.56%, False positive rate: 0.97%	The system does not yet incorporate an inference engine or knowledge base, which limits the intelligent decision-making capability of the IDS
8	(Acharjya and Ahmed, 2021)	Hybrid ML	DRS + FCA	Real network traffic (sFlow version 5)	Accuracy: 98.19%, Precision: 93.33%, Recall: 93.17%	It is acknowledged that it can be improved with more advanced methods such as SVM, neural networks, etc.
9	(Gebremariam et al., 2023a)	Hybrid ML	RF-XGB, CLK-M	NSL-KDD, UNSW-NB15, CICIDS2017	Accuracy up to 100%, Precision: 99.89%, Recall: 99.93%, F1-Score: 99.91%	Limited scalability in large networks; complexity and computational overhead
10	(Nandhini et al., 2024)	Hybrid Deep Learning + Threat Intelligence	EO-NN, PSO-NN, SCO-LSTM	Custom dataset generated in IoT-WSN test-bed	Accuracy: 99.89%, Precision and Recall > 99%	Validation restricted to a controlled test-bed, without considering RPL-specific attacks
11	(Verma and Ranga, 2019)	Supervised ML	Boosted Trees	RPL-NIDDS17	Accuracy: 94.5%, AUC: 0.98 (30% hold-out)	The model has not yet been implemented on physical nodes, only in simulation.
12	(Ahmadi and Javidan, 2024)	Hybrid Deep Learning	stacked LSTM + Seq2Seq	Simulation (Contiki, Cooja)	MAPE, Attack frequency, Temporal analysis	Continuous real-time training of the model in windows is costly for constrained IoT devices
13	(Gebremariam et al., 2023b)	Hybrid Deep Learning	MLPANN	UNSW-NB15, WSN-DS, NSL-KDD, CICIDS2018	- Accuracy: up to 100% (WSN-DS), 99.83% (CICIDS2018), 99.65% (UNSW-NB15), 98.95% (NSL-KDD); - Localization Accuracy: up to 99.12% with 160 beacon nodes	It is recommended to extend tests to other attack classes and use other tools
14	(Kamel and Elhamayed, 2020)	Supervised Deep Learning	CNN	Simulation (Contiki, Cooja, TMote Sky)	Accuracy: 98.57%, Precision: 99.47%, Recall: 99.55%, F1-score: 99.47%, ROC: 0.986	Training is costly due to the complexity and size of the dataset and limitations in the scale and diversity of simulated scenarios
15	(Prathapchandran and Janani, 2021)	Supervised ML	Random Forest (RF) +	Simulation (Contiki, Cooja, TMote Sky)	Accuracy: 85%, FPR: 1.4%, FNR: 1.8%, PDR: 72%,	The RFTrust model does not continuously evaluate node

			Subjective Logic (SL)		Consumption: 115.8 J; Average Throughput: High (exact value not specified)	behavior
16	(Raghavendra et al., 2022)	Supervised ML	Fuzzy k-NN	Simulation (Contiki, Cooja)	Accuracy, Precision, Recall, F1-Score	No specific values are given; the current system does not efficiently detect unknown attacks
17	(Sahay et al., 2024)	Deep Learning	RBM + LSTM	Simulation (Contiki, Cooja)	Evaluation based on visual and tabular comparison between actual vs. predicted values; Trend consistency is used as a criterion to validate the prediction	The current model does not apply hyperparameter optimization and does not detect new patterns or unknown attacks well
18	(Tertytchny et al., 2020)	Supervised ML	J48, NB, MLP, MLG	Simulation and Testbed with Wireshark and ZigBee	Testbed: Accuracy: 98.18%, Kappa: 0.97, Precision: 0.979, Recall: 0.978 (case 1: Normal vs. Faults); Simulation: Accuracy: 91.67%, Kappa: 0.89, Precision: 0.935, Recall: 0.917	Similarity in network effects between attacks and faults impedes precise differentiation
19	(Wakili et al., 2024)	Hybrid: – Supervised (ML); – Reinforced (RL)	Random Forest	Simulation (Contiki-NG, Wireshark)	Accuracy: 0.99, Precision: 0.99, Recall: 1.00, F1-score: 0.99	High overhead; not yet tested in real environments (simulations only)
20	(Gebremariam et al., 2023c)	Hybrid ML	RF + K-means	CICIDS2017, NSL-KDD, UNSW-NB15	CICIDS2017 and UNSW-NB15: 100% in Accuracy, Precision, Recall, and F1-score; NSL-KDD: 99.80% in Accuracy, Precision, Recall, F1-score, with ROC of 99.90%	High deployment and maintenance costs due to specialized hardware or software

Scenario 3: Two uncoordinated sinkholes: A network of 50 normal nodes and one root node was deployed, incorporating two independent sinkhole nodes operating without coordination. The objective was to analyze the model's performance in situations where multiple, dispersed attack points were present within the network.

Scenario 4: Attacker mobility: Finally, a topology with 50 normal nodes, one root node, one mobile sinkhole, and one static sinkhole was simulated. The malicious mobile node was programmed to change its position every 30 simulated minutes using control scripts. This scenario was intended to evaluate the model's resilience against dynamic adversaries capable of altering their location.

Figure 1 visually illustrates the resulting topology of the Baseline topology of 34 nodes in Cooja. In the visual representation, the nodes are distinguished by color according to their function in the network:

- The green node represents the RPL root node, responsible for coordinating the topology
- The yellow nodes correspond to normal nodes operating as UDP clients within the network
- The purple nodes are the malicious nodes, configured to execute the Sinkhole attack by advertising artificially low ranks

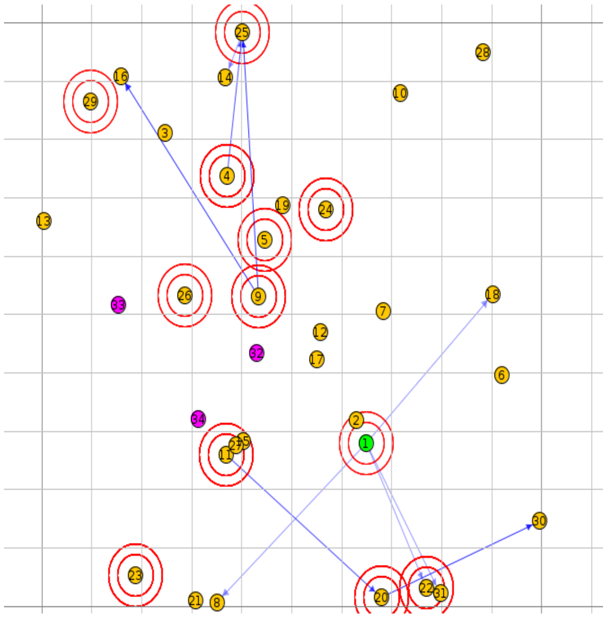

Fig. 1: Simulation environment in Cooja, where the RPL network structure is visualized

Threat Model and Attack Implementation

The adversary compromises nodes to execute a sinkhole, advertising an artificially low rank and acting as a "false root" to attract traffic toward itself. This strategy allows packets to be dropped either completely or selectively, degrading network performance and simulating a realistic attack within the test scenarios. The implementation in Contiki/C, through the creation of a fake DAG with rank = 1 and prefix advertisements, ensures that malicious behavior is consistent and measurable, providing a controlled basis for evaluating the detection model's effectiveness under varying levels of adversary activity and topology conditions.

Data Generation and Preparation

The simulation data were collected from two sources: (1) the console logs of each mote, which provided detailed information about node behavior and internal RPL metrics; and (2) .pcap files generated by the simulator's radio-message tool, which captured all network traffic for subsequent analysis.

The .pcap files were analyzed with Wireshark and converted to .csv format using tshark for further processing in Python. In parallel, relevant data were extracted from the motes' console logs, including information on RPL ranks, network hops, and node states. Both data sources were integrated to construct a single structured dataset for each scenario.

The dataset underwent several preprocessing steps: first, the initial 208 rows corresponding to startup events or system artifacts were removed; next, uninformative or highly redundant attributes were discarded. Missing values were imputed using the median, and numerical variables received a Yeo Johnson power transformation to stabilize their variance and approximate a Gaussian distribution, thus facilitating classifier training.

Model Training

For the automatic detection of anomalous behaviors, a Long Short-Term Memory (LSTM) neural network model was trained, suitable for time-series analysis due to its ability to capture long-term dependencies.

Prior to training, the dataset was preprocessed as described above, including normalization, imputation of missing values, and variance-stabilizing transformations. Input sequences for the LSTM were constructed using a sliding window approach: each sequence contained `ts` consecutive packets, capturing temporal patterns across multiple features. Overlapping sequences were generated to maximize temporal information while preserving the order of packets.

The dataset was split using a stratified strategy, ensuring that both training and validation sets maintained the original class distribution. The LSTM model consisted of two layers (128 and 64 units) with dropout in between, followed by a fully connected dense layer with sigmoid activation for binary classification. Training was performed with the Adam optimizer and binary cross-entropy loss over multiple epochs using batch learning. Class imbalance in the training sequences was addressed via weighted loss to improve the detection of the minority class.

Evaluation

Once training was complete, the model's performance was evaluated on the test set using the following metrics:

- Accuracy: Overall percentage of correct predictions
- Precision: Proportion of attacks correctly identified among all positive predictions
- Sensitivity (Recall): Model's ability to identify all present attacks
- F1-score: Harmonic mean of precision and recall

Figure 2 provides a structured summary of the methodological flow followed to build the sinkhole attack detection model. The process begins with data collection from .pcap files and console logs generated during the Cooja simulation. These data undergo preprocessing including cleaning, normalization, and encoding followed by feature extraction. The resulting dataset is then split into training and test subsets. An LSTM neural network is trained on the training data and evaluated on the test data. Finally, the resulting model is consolidated as an anomaly detection tool applicable to institutional IoT environments.

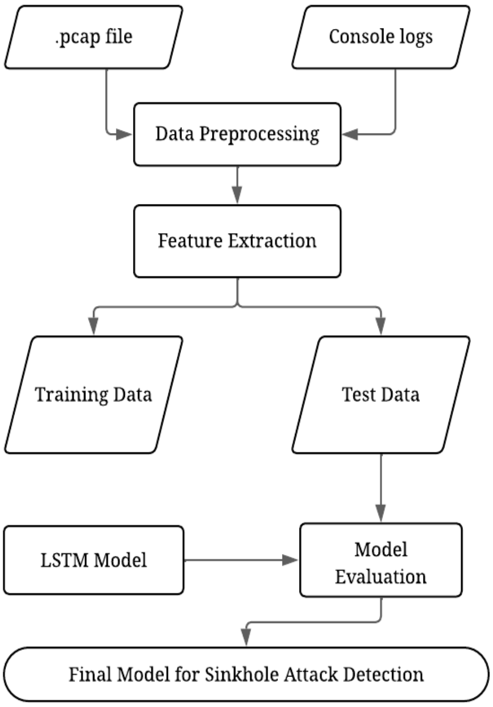

Fig. 2: Methodological diagram for the construction of the LSTM-based sinkhole attack detection model

Results

The evaluation of the Long Short-Term Memory (LSTM) model across the four experimental scenarios provides a comprehensive view of its detection capability under progressively more complex network conditions. The analysis considers both aggregated metrics accuracy, precision, recall, F1-score, and ROC curves. This approach not only quantifies performance but also highlights the strengths and limitations of the model when exposed to variations in network density, traffic load, number of attackers, and adversarial mobility. The following sections describe the results obtained in each scenario.

Scenario 1: Baseline (34 Nodes, 3 Sinkholes)

In the baseline scenario, the model achieved an overall accuracy of 0.98. For normal traffic (Class 0), it reached a precision of 0.99, a recall of 0.98, and an F1-score of 0.98. For attack traffic (Class 1), precision was 0.96, recall was 0.97, and F1-score was 0.96. These results demonstrate that the classifier maintained an excellent balance between sensitivity and specificity, ensuring reliable detection of malicious behavior while preserving high accuracy on legitimate traffic. The ROC curve (Figure 3) further supports this conclusion, with an AUC of 0.99, confirming the strong discriminative power of the model in this baseline configuration.

Scenario 2: Stress Test (100 Nodes, 1 Sinkhole, High Traffic)

Under intensive traffic with 100 nodes, the model achieved an overall accuracy of 0.956. For normal traffic, it obtained a precision of 0.987, a recall of 0.966, and an F1-score of 0.977. For attack traffic, precision was 0.592, recall 0.795, and F1-score 0.678. While the model detects most attacks (high recall), the lower precision indicates a notable rate of false positives, which reduces its reliability under heavy traffic. The ROC curve (Figure 4) reflects this trade-off, with AUC=0.959 on test (0.965 on validation) at a decision threshold of 0.319, showing greater overlap between legitimate and malicious traffic compared with the baseline. stress scenario (100 nodes).

Scenario 3: Two Uncoordinated Sinkholes (50 Nodes)

When two independent sinkholes were introduced, the model achieved an overall accuracy of 0.986. For normal traffic, it reached a precision of 0.996, a recall of 0.996, and an F1-score of 0.996. For attack traffic, precision was 0.991, recall was 0.974, and F1-score was 0.982. These results confirm the model's ability to generalize and sustain high performance even with multiple simultaneous attackers. The ROC curve (Figure 5) illustrates this stability, with an AUC of 0.999 on both validation and test sets at a threshold of 0.298, showing an almost perfect separation between legitimate and malicious traffic.

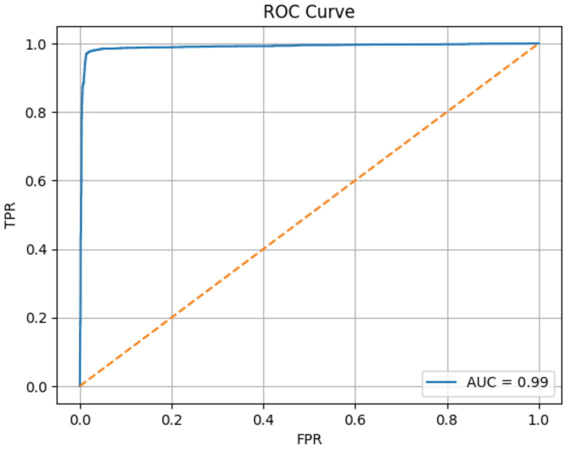

Fig. 3: ROC curve for the baseline scenario (34 nodes)

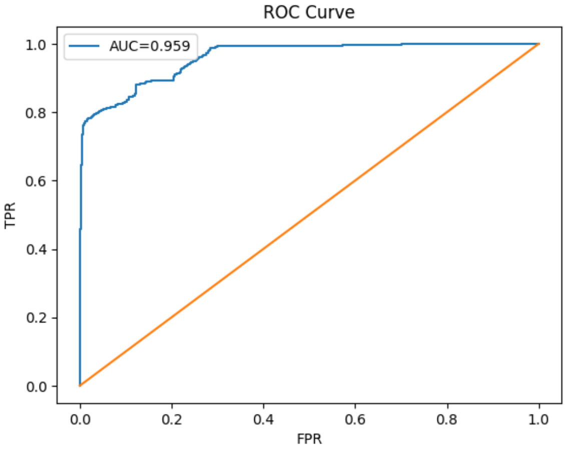

Fig. 4: ROC curve for the stress scenario (100 nodes)

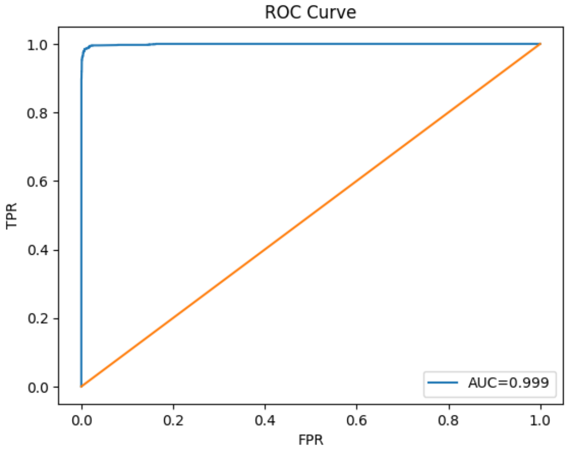

Fig. 5: ROC curve for the two sinkhole scenario (50 nodes)

Scenario 4: Attacker Mobility (50 Nodes, 1 Fixed and 1 Mobile Sinkhole)

In the mobility scenario, the model achieved the highest performance across all experiments, with an overall accuracy of 0.997. For normal traffic, it obtained a precision of 0.998, a recall of 0.998, and an F1-score of 0.998. For attack traffic, precision was 0.997, recall was 0.997, and F1-score was 0.997. These results confirm that the model not only adapts to dynamic topologies but also leverages the temporal patterns introduced by attacker relocation. The ROC curve (Figure 6) remained nearly aligned with the upper-left boundary, with an AUC of 0.999 on both validation and test sets at a threshold of 0.558, evidencing the model's ability to sustain extremely high sensitivity and specificity under adversarial mobility

In addition, the primary LSTM-based model was compared against three representative baselines: Random Forest (RF) as a robust ensemble, Logistic Regression (LR) as a simple linear classifier, and Support Vector Machine (SVM) as a non-linear classifier. Table 3 summarizes the metrics obtained in terms of AUC, accuracy, recall, and F1-score.

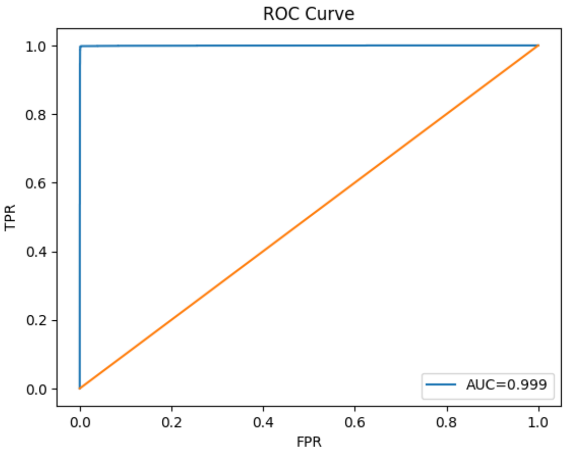

Fig. 7: ROC curve for the mobility scenario (50 nodes, one fixed and one mobile attacker)

Table 3: Baseline model performance (AUC, precision, recall, F1) across scenarios

Scenario	Model	AUC	Precision	Recall	F1
1	RandomForest	0.9999	0.994	1.000	0.997
	LogisticRegression	0.992	0.946	0.989	0.967
	SVM	0.924	0.760	0.950	0.844
2	RandomForest	1.000	1.000	0.998	0.999
	LogisticRegression	0.895	0.464	0.083	0.140
	SVM	0.892	0.854	0.041	0.078
3	RandomForest	1.000	1.000	1.000	1.000
	LogisticRegression	0.965	0.915	0.912	0.913
	SVM	0.978	0.836	0.913	0.873
4	RandomForest	1.000	1.000	1.000	1.000
	LogisticRegression	0.984	0.995	0.954	0.974
	SVM	1.000	0.999	0.999	0.999

Furthermore, ablation experiments were conducted on the LSTM to assess the impact of key design elements. When the temporal window was removed (“LSTM_no_window”), performance remained high, suggesting a certain robustness of the model. However, omitting class rebalancing (“LSTM_no_class_weight”) caused the model to fail in detecting attacks, highlighting the critical importance of this adjustment. With shorter temporal windows (“LSTM_short_window"), the model exhibited a notable drop in AUC and F1, confirming that sequence length is crucial for capturing significant temporal patterns

Discussion

The results show that Random Forest constitutes a surprisingly strong baseline, achieving near-perfect metrics in most scenarios. This suggests that packet-derived features provide sufficient discriminative information for tree-based classifiers without requiring complex temporal modeling. However, this performance should be interpreted cautiously, as it may reflect overfitting to simulated datasets and may not generalize to real-world environments with higher noise and variability.

In contrast, Logistic Regression and SVM demonstrate more realistic limitations, struggling to generalize under highly imbalanced conditions or with multiple attackers, underscoring the need for more expressive models to capture dynamic traffic patterns. LSTM ablation analyses highlight the critical role of class weighting in mitigating majority-class bias and the importance of an adequate temporal window for capturing attack evolution. Without these adjustments, LSTM performance drops sharply, justifying the selection of a temporally-sequenced, class-balanced architecture as the central model.

The proposed LSTM achieves outstanding performance, with AUC and F1 scores approaching 1 in balanced scenarios, and remains robust under adversarial dynamics, including multiple attackers and adversary mobility. Compared to prior works such as RFTrust (Prathapchandran and Janani, 2021) and AIEMLA (Sharma and Verma, 2021), it demonstrates significant improvements under complex and dynamic conditions. Some false negatives persist, particularly in baseline or stress scenarios, suggesting that certain malicious nodes can mimic normal traffic. Incorporating additional temporal features such as rank variations, hop counts, or DIO/DAO message irregularities could help capture subtler anomalies.

In stress scenarios with large, congested networks, false positives increase, reducing precision while recall remains high. This indicates that legitimate traffic can occasionally resemble attacks, emphasizing the need for threshold calibration or cost-sensitive learning. Conversely, scenarios with multiple attackers and adversary mobility confirm the advantage of sequential modeling: temporal irregularities amplify anomaly signals, enabling near-perfect classification and demonstrating that dynamic behaviors can enhance separability when properly captured.

Overall, these results reinforce the suitability of deep learning for IoT-WSN intrusion detection, while highlighting the necessity of validation on real hardware, assessment of energy and latency costs, and extension to compound attacks combining sinkholes with other routing disturbances.

Conclusion and Future Work

For future work, the methodology will be extended to cover other relevant attacks, such as wormhole, selective forwarding, or jamming, while implementing the classifier on real hardware to experimentally evaluate its impact on energy consumption and latency. Additionally, distributed or federated learning techniques will be explored to enable lightweight versions of the model to run directly on edge nodes, and the system will be validated in WSN scenarios with dynamic topologies and diverse traffic patterns, moving closer to real IoT deployment environments.

This study demonstrated that the extraction of representative indicators of node activity at the network layer, combined with deep learning algorithms, enables effective detection of sinkhole attacks in RPL-based WSNs. The proposed LSTM model achieved an overall accuracy of 98% in the baseline scenario, with 96% precision in identifying attack traffic, confirming its capacity to balance sensitivity and specificity under realistic conditions. The validation across four experimental scenarios, including stress tests, multiple attackers, and adversarial mobility, further highlighted the robustness of the approach, as well as its adaptability to dynamic and complex topologies.

It is concluded that the methodology presented here provides a replicable workflow, from simulation and data extraction to preprocessing, model training, and evaluation, offering a practical path for strengthening WSN security against routing attacks. Moreover, the results establish a solid foundation for real-world adaptation, showing that sequence-based models such as LSTM are particularly suited to capturing temporal patterns that characterize adversarial behaviors.

Acknowledgment

The authors would like to thank the Dirección de Investigación y Desarrollo (DIDE), Facultad de Jurisprudencia y Ciencias Sociales, Carrera de Derecho de la Universidad Técnica de Ambato (UTA), Centro de Estudios de Seguridad (CESEG), Universidad de Santiago de Compostela (USC) España, Universidad Estatal de Bolívar (UEB), and RNI-Senescyt.

Funding Information

The authors have not received any financial support or funding to report.

Author's Contributions

Estefanía Alejandra Mora Parra: Research Identification, objective, synthesis, and bibliographic study. Stake in the writing of the manuscript and the interpretation of the results.

Rubén Nogales Portero: Participation in writing the manuscript. The development of the questionnaire and its distribution, as well as data collection.

Moisés Toapanta T.: Participation in writing the manuscript. The development of the questionnaire and its distribution, as well as data collection.

Estefanía Monge Martínez: Participation in data analysis and reference management.

Santiago Vayas Castro: Participation in the interpretation of results. Investigate previous research similar to the study.

Jeanette Elizabeth Jordán Buenaño: The contribution to the correction of the English language. Workflow monitoring and general monitoring.

Juan Escobar Naranjo: Participation in data analysis and reference management.

Diego Andrade Armas: Participation in writing the article. Investigate previous research similar to the study, analyze data, and analyze results.

Rodrigo Del Pozo Durango: Participation in the interpretation of results. Investigate previous research similar to the study.

Ethics

Authors should address any ethical issues that may arise after the publication of this manuscript.

References

Abdul-Bari Ahmed Mohammed, F., Mekky, N. E., Soliman, H., & Hikal, N. A. (2024). Sinkhole Attack Detection by Enhanced Reputation-Based Intrusion Detection System. *IEEE Access*, *12*, 86985–86996. https://doi.org/10.1109/access.2024.3416270

Academy, e-Governance. (2022). *NCSI: Ecuador – National Cyber Security Index*. https://ncsi.ega.ee/country/ec_2022/

Acharjya, D. P., & Ahmed, N. S. S. (2021). Tracing of online assaults in 5G networks using dominance based rough set and formal concept analysis. *Peer-to-Peer Networking and Applications*, *14*(1), 349–374. https://doi.org/10.1007/s12083-020-00983-6

Ahmadi, K., & Javidan, R. (2024). A Trust Based Anomaly Detection Scheme Using a Hybrid Deep Learning Model for IoT Routing Attacks Mitigation. *IET Information Security*, *2024*(1). https://doi.org/10.1049/2024/4449798

Bakar, K. B. A., Zuhra, F. T., Isyaku, B., & Sulaiman, S. B. (2023). A Review on the Immediate Advancement of the Internet of Things in Wireless Telecommunications. *IEEE Access*, *11*, 21020–21048. https://doi.org/10.1109/access.2023.3250466

Choudhary, A. (2024). Internet of Things: a comprehensive overview, architectures, applications, simulation tools, challenges and future directions. *Discover Internet of Things*, *4*(1), 31. https://doi.org/10.1007/s43926-024-00084-3

Chowdhury, M., Ray, B., Chowdhury, S., & Rajasegarar, S. (2021). A Novel Insider Attack and Machine Learning Based Detection for the Internet of Things. *ACM Transactions on Internet of Things*, *2*(4), 1–23. https://doi.org/10.1145/3466721

Corporación Ecuatoriana para el Desarrollo de la Investigación y la Academia [CEDIA]. (2024). *Cybersecurity and Legal Protection: A Comprehensive Approach for Institutions - Connect by CEDIA*. https://connect.cedia.edu.ec/en/cybersecurity-and-legal-protection-a-comprehensive-approach-for-institutions/.

Gebremariam, G. G., Panda, J., & Indu, S. (2023a). Design of advanced intrusion detection systems based on hybrid machine learning techniques in hierarchically wireless sensor networks. *Connection Science*, *35*(1), 2246703. https://doi.org/10.1080/09540091.2023.2246703

Gebremariam, G. G., Panda, J., & Indu, S. (2023b). Localization and Detection of Multiple Attacks in Wireless Sensor Networks Using Artificial Neural Network. *Wireless Communications and Mobile Computing*, *2023*(1), 2744706. https://doi.org/10.1155/2023/2744706

Gebremariam, G. G., Panda, J., & Indu, S. (2023c). Secure localization techniques in wireless sensor networks against routing attacks based on hybrid machine learning models. *Alexandria Engineering Journal*, *82*, 82–100. https://doi.org/10.1016/j.aej.2023.09.064

Guo, D., & Xia, L. (2022). WSNs-Based Data Transmission Bandwidth Allocation Method for Smart Campus Communication Network in Colleges and Universities. *Mobile Information Systems*, 1–11. https://doi.org/10.1155/2022/4393305

Hachemi, F.-E., Mana, M., & Bensaber, B. A. (2020). *Study of the Impact of Sinkhole Attack in IoT Using Shewhart Control Charts*. 1–5. https://doi.org/10.1109/globecom42002.2020.9322603

Huang, H., Chen, W., Fang, W., Li, Z., Chen, W., Ip, W.-H., & Yung, K.-L. (2024). State-of-the-art Review on Intelligent Computing-based Privacy Preservation Technologies for Power Internet of Things. *Proceedings of the 2024 6th International Conference on Big Data Engineering*. BDE 2024: 2024 6th International Conference on Big Data Engineering, Xining China. https://doi.org/10.1145/3688574.3688582

Kamel, M. S. O., & Elhamayed, S. A. (2020). Mitigating the Impact of IoT Routing Attacks on Power Consumption in IoT Healthcare Environment using Convolutional Neural Network. *International Journal of Computer Network and Information Security*, *12*(4), 11–29. https://doi.org/10.5815/ijcnis.2020.04.02

Maleh, Y., Sahid, A., & Belaissaoui, M. (2021). Optimized Machine Learning Techniques for IoT 6LoWPAN Cyber Attacks Detection. *Proceedings of the 12th International Conference on Soft Computing and Pattern Recognition*, *1383*, 669–677. https://doi.org/10.1007/978-3-030-73689-7_64

Nandhini, S., Rajeswari, A., & Shanker, N. R. (2024). Cyber attack detection in IOT-WSN devices with threat intelligence using hidden and connected layer based architectures. *Journal of Cloud Computing*, *13*(1), 1–21. https://doi.org/10.1186/s13677-024-00722-9

Nwankwo, K. E., Abdulhamid, S. M., Ojeniyi, J. A., Misra, S., Oluranti, J., & Ahuja, R. (2021). A Panacea to Soft Computing Approach for Sinkhole Attack Classification in a Wireless Sensor Networks Environment. *Futuristic Trends in Network and Communication Technologies*, *1395*, 78–87. https://doi.org/10.1007/978-981-16-1480-4_7

Oztoprak, A., Hassanpour, R., Ozkan, A., & Oztoprak, K. (2025). Security Challenges, Mitigation Strategies, and Future Trends in Wireless Sensor Networks: A Review. *ACM Computing Surveys*, *57*(4), 1–29. https://doi.org/10.1145/3706583

Prathapchandran, K., & Janani, T. (2021). A trust aware security mechanism to detect sinkhole attack in RPL-based IoT environment using random forest – RFTRUST. *Computer Networks*, *198*, 108413. https://doi.org/10.1016/j.comnet.2021.108413

Raghavendra, T., Anand, M., Selvi, M., Thangaramya, K., Santhosh Kumar, S., & Kannan, A. (2022). An Intelligent RPL attack detection using Machine Learning-Based Intrusion Detection System for Internet of Things. *Procedia Computer Science*, *215*, 61–70. https://doi.org/10.1016/j.procs.2022.12.007

Rajasoundaran, S., Kumar, S. V. N. S., Selvi, M., Ganapathy, S., Rakesh, R., & Kannan, A. (2021). Machine learning based volatile block chain construction for secure routing in decentralized military sensor networks. *Wireless Networks*, *27*(7), 4513–4534. https://doi.org/10.1007/s11276-021-02748-2

Sahay, R., Nayyar, A., Shrivastava, R. K., Bilal, M., Singh, S. P., & Pack, S. (2024). Routing attack induced anomaly detection in IoT network using RBM-LSTM. *ICT Express*, *10*(3), 459–464. https://doi.org/10.1016/j.icte.2024.04.012

Samy, A., Yu, H., & Zhang, H. (2020). Fog-Based Attack Detection Framework for Internet of Things Using Deep Learning. *IEEE Access*, *8*, 74571–74585. https://doi.org/10.1109/access.2020.2988854

Sejaphala, L. C., & Velempini, M. (2020). The Design of a Defense Mechanism to Mitigate Sinkhole Attack in Software Defined Wireless Sensor Cognitive Radio Networks. *Wireless Personal Communications*, *113*(2), 977–993. https://doi.org/10.1007/s11277-020-07263-9

Sharma, S., & Verma, V. K. (2021). AIEMLA: artificial intelligence enabled machine learning approach for routing attacks on internet of things. *The Journal of Supercomputing*, *77*(12), 13757–13787. https://doi.org/10.1007/s11227-021-03833-1

Sivagaminathan, V., Sharma, M., & Henge, S. K. (2023). Intrusion detection systems for wireless sensor networks using computational intelligence techniques. *Cybersecurity*, *6*(1), 1–15. https://doi.org/10.1186/s42400-023-00161-0

Sridevi, S., & Anandan, R. (2020). RUDRA—A Novel Re-concurrent Unified Classifier for the Detection of Different Attacks in Wireless Sensor Networks. *Intelligent Computing in Engineering*, *1125*, 251–259. https://doi.org/10.1007/978-981-15-2780-7_29

Tertytchny, G., Nicolaou, N., & Michael, M. K. (2020). Classifying network abnormalities into faults and attacks in IoT-based cyber physical systems using machine learning. *Microprocessors and Microsystems*, *77*, 103121. https://doi.org/10.1016/j.micpro.2020.103121

Verma, A., & Ranga, V. (2019). ELNIDS: Ensemble Learning based Network Intrusion Detection System for RPL based Internet of Things. *Proceedings of the 2019 International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU 2019*, 1–6. https://doi.org/10.1109/iot-siu.2019.8777504

Wakili, A., Bakkali, S., & Alaoui, A. E. H. (2024). Machine learning for QoS and security enhancement of RPL in IoT-Enabled wireless sensors. *Sensors International*, *5*, 100289. https://doi.org/10.1016/j.sintl.2024.100289

Zhou, I., Makhdoom, I., Shariati, N., Raza, M. A., Keshavarz, R., Lipman, J., Abolhasan, M., & Jamalipour, A. (2021). Internet of Things 2.0: Concepts, Applications, and Future Directions. *IEEE Access*, *9*, 70961–71012. https://doi.org/10.1109/access.2021.3078549