Babylonian Journal of Artificial Intelligence
Vol. **(2026)**, 2026, **pp**. 1–19
DOI: https://doi.org/10.58496/BJAI/2026/007 ISSN: 3006-5437
https://mesopotamian.press/journals/index.php/BJAI

Research Article

Hybrid AI-Driven IDS for IoT Using Cooja and Explainable Deep Learning

Zainab Ali Abbood [1,*], Raghad Tariq Al_Hassani [2], Mahmoud Shuker Mahmoud [3,4], Haider D. Albonda [5]

[1] *Computer Technology Engineering Department, Al-Mansour University College, Baghdad, Iraq.*

[2] *Ministry of Higher Education and Scientific Research in Iraq, Minister's Office, Baghdad, Iraq.*

[3] *Gilgamesh University, Baghdad, Iraq.*

[4] *Cybersecurity Technology Engineering Department, Middle Technical University, Electrical Engineering Technical College, Baghdad, Iraq.*

[5] *Department of Control and Systems Engineering, University of Technology, Baghdad, Iraq.*

Article info

Article History
Received 17 Jan 2026
Revised 15 Feb 2026
Accepted 25 Mar 2026
Published 18 Apr 2026

Keywords
IoT Intrusion Detection System (IDS),
Hybrid Deep Learning;
Explainable AI (XAI),
RPL Routing Attacks,
Cooja Simulation.

Abstract

The explosive growth in the adoption of the Internet of Things (IoT) has unveiled important security concerns, especially considering energy and computational constraints for low-power devices, increased levels of advanced routing attacks for RPL-based networks. Current IDSs are concentrated mainly on detection of anomaly with no clear explanation of the reason behind, which limits their application in safety systems. In addition, the lack of IoT-related benchmark datasets also prevents the construction of generalizable IDS systems. To narrow the gap, we design a hybrid AI-based IDS with combining Convolutional Neural Networks (CNNs), Bidirectional LSTM (BiLSTM) and Random Forest (RF) together for spatial, temporal and decision-level feature extraction. When modelling sinkhole, version number manipulation, or flooding attacks in the Contiki-NG Cooja simulator, three IoT-specific datasets were derived that allow a detailed packet-level investigation of realistic network behaviour. The proposed model outperforms traditional deep learning methods (CNN, LSTM and CNN-AO), showing better accuracy, precision, recall, F1-score AUC and MCC results. We use explainable AI (XAI) techniques, namely SHAP and LIME, to provide intuitive explanation on the feature contribu- tions and attack signatures. Experimental results show that the model has strong detection ability, low false alarm rate and is applicable for realtime deployment in a resource limited IoT environment. This work proposes a fully traceable IoT specific ready-to-use powerful IDS framework intensively validated through comprehensive simulation and evaluation with novel IoT-specific datasets.

1. INTRODUCTION

The increasingly rapid deployment of the internet of things (IoTs) is developing large-scale interconnectivity in smart home, health care industry, industrial automation and critical infrastructures. However, IoT is still so vulnerable because constrained computing resources are heterogeneous and they have few built-in security solutions. These limitations make the networks vulnerable to routing manipulation, flooding, spoofing and other cyber-attacks that can undermine network performance and compromise data confidentiality and availability. Therefore, intrusion detection systems (IDSs) are indispensable components for protecting IoT environments [1-3]. Nevertheless, traditional IDS approaches fail to adapt well to the traffic profile specifically seen in IoT and also do not offer explanations of any sort on model decisions which greatly diminishes their real-world applicability.

Recent research highlights the increasing demand for IoT-targeted datasets bridging the gap between realistic constrained-device behavior and realistic attack patterns. The currently available benchmark datasets e.g., NSL-KDD, CICIDS2017, and UNSW-NB15 were created using traditional networks, where low-power wireless protocols (routing protocol for low-power and lossy networks) like RPL are not represented [4,5]. To cope with this shortcoming, simulating driven dataset generation [6] has been increasingly used by researchers through conglomerate tools like Contiki-NG and Cooja simulator.

*Corresponding author. Email: zainab.a.abbood@muc.edu.iq

This method allows sinkhole, version number and flooding attacks to be controlled tested producing more realistic IoT traces.

Deep learning (DL) techniques such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM), have demonstrated good results in classifying IoT-related anomalies. Nonetheless, they are mostly black-box models and it's hard to make them explainable and interfere-proof. Hybrid AI models—integrating DL with optimization algorithms and classical machine learning have provided a promising alternative since they achieve better accuracy as well as adaptability [7,8]. So far there is one big gap in research: the existing IDS frameworks lack interpretability, meaning that the operators cannot understand why attacks are detected or which network behaviors drive model decisions.

To address these limitations, we propose Hybrid AI-Driven Intrusion Detection System for IoT based on dataset generation through Cooja and XAI. It is built on top of original research work where we have shown the effectiveness of CNN, LSTM and CNN-Aquila Optimizer (CNN-AO) model for detect sinkhole, version number and flooding attack in RPL-based networks. Extending on this, our hybrid model combines XAI (SHAP & LIME) techniques ensuring transparent feature attribution and attack interpretation. This builds confidence, helps with operational decisions and gives us more insight into when network problems occur.

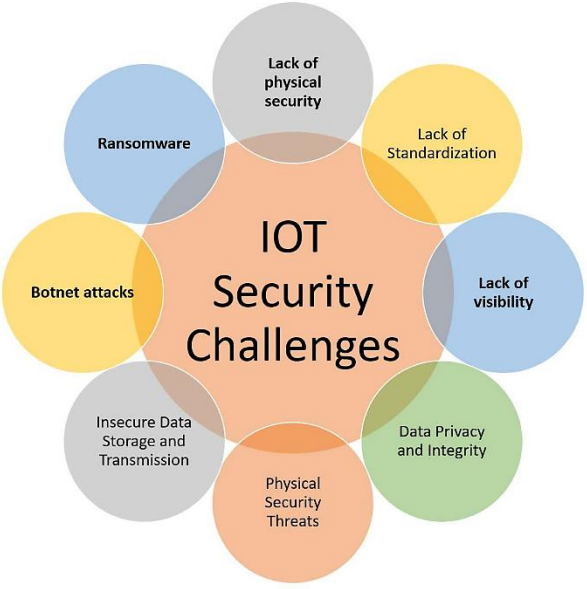


Fig. 1. Taxonomy of IoT Security Challenges

2. RELATED WORK

IDSs for IoT have seen considerable development in the last few years and some surveys focusing on their architectures, classifications, and challenges. Recent works highlight that the IoT-specific limitations (or, constraints) including energy scarcity, processing power scarcity and heterogeneity in protocols demand custom IDS designs rather than a direct use of classical network datasets/models [8,9]. These works categorize IDS solutions according to detection method, implementation and scope conditions as well as target layer and emphasize the popularity of machine learning and deep learning (ML/DL) in anomaly detection. However, they also highlight issues related to realistic dataset, encryption traffic processing, scalability for massive IoT deployment and lack of interpretability in AI-driven IDSs that motivate the investigation pursued by the current work.

Deep learning (DL) contributes to a fundamental enabler of the current IoT IDSs, which can learn intricate patterns in traffic and identify subtle anomalies. In the recent past, models include CNNs, RNNs, GRUs and BiLSTMs to achieve high accuracy on benchmarks like CICIDS2017, UNSWN [10-12], ToN-IoT and latest Iot-specific Datasets. Lightweight architectures have been suggested to decrease the computational burden of limited devices while maintaining high detection accuracy, such as DNN–BiLSTM models optimized for energy and latency at the edge level. However, the majority of works are still centralized/dependent on existing datasets, and give little information about why specific flows are flagged as attacks, that can point to deployment realism and transparency shortcomings.

Beyond supervised deep learning paradigms, reinforcement learning (RL) has been increasingly adopted in intelligent network management to cope with highly dynamic and heterogeneous environments. Recent studies in 5G and vehicular communication networks have demonstrated that RL-based control and radio access network (RAN) slicing strategies can effectively adapt to fluctuating traffic demands, quality-of-service requirements, and complex system behaviors [13,14].

Although these works primarily focus on resource allocation and network optimization rather than security, they highlight the potential of adaptive and learning-driven decision mechanisms, which are highly relevant to the design of intelligent and self-adaptive intrusion detection systems for large-scale IoT environments.

Federated learning (FL) emerges as a promising model for cooperative IDS training over distributed IoT devices without exchanging the raw data. Recent FL-based IDS approaches explore the proper local model structures, communication rounds and aggregation styles to reconcile accuracy, privacy and computational cost over resource-constrained nodes [15]. These works demonstrate that FL can enhance privacy and scalability, especially in the context of large scale IoT and 6G enabled deployments. Despite this, issues still endure around non-IID data distributions, communication overheads, poisoning attacks against model updates and a limited incorporation of explainability within FL pipelines - suggesting that FL is not sufficient in isolation to tackle the trust newly interpretability challenge in IoT IDS.

Hybrid IDS systems integrate deep learning, classical machine learning, optimization algorithms or in some cases SDN-based control planes to enhance attack detection robustness. Hybrid methods that combine RNN-GRU, CNN-BiLSTM or multi-view learners to learn both spatial and temporal features of IoT traffic have been recently introduced [16, 17]. These approaches frequently incorporate feature selection or meta-heuristics-based optimization for dimensionality reduction and hyperparameter tuning securing competitive accuracy with the false alarm rates on IoT datasets. Nonetheless, black-box-wise most H-IDSs are still non-interpretable and seldom provide human-understandable reasoning for their decisions which puts an obstacle to the use of these solutions in mission-critical IoT scenarios.

Recently, the XAI has been used more in cybersecurity to handle opacity of neural models in IDS. Recent researches seamlessly combine SHAP, LIME, RuleFit and other explainable module into IDSs frameworks to produce global and local explanations, feature importance rankings, and human comprehensible rationale for attack detection [18-20]. XAI for IDS becomes popular among industrial communities and researches, which state the importance of XAI in IDS to gain human analyst trust, assist forensics or model bias detection and raise concern over adversarial misuse of explanations. Nevertheless, a number of current XAI-enabled IDS systems still employ conventional datasets (e.g., NSL-KDD and UNSW-NB15) and do not entirely take into account the IoT-oriented traffic behavior, constrained devices, and low-powered routing protocols like RPL, leaving space for this work.

Smaller number of existing works that encompasses Cooja-based IoT simulations for realistic anomaly datasets creation for an IDS to be evaluated. In this works, Contiki-NG or Contiki OS with Cooja are employed to simulate RPLbased network topologies whilst injecting attacks such as sinkhole attack, version number attack and flooding; packet- level traces obtained therein a set used to build IoT-dataset [21]. The base study of this work aids in creating three separate dedicated malicious datasets, and obtains CNN, LSTM and a (CNN–Aquila Optimizer) model for comparison where the latter reaches over 99% accuracy among the entire fake attacks. Although this methodology enhances the realism of the dataset and allows to explore deep models, it does not consider a) explainability, or b) hybrid decision mechanisms. We further this trend by incorporating XAI-guided analysis into a hybrid AI-based IDS, also utilizing the same Cooja-harvested datasets. To summarize the above-discussion, Table 1 presents a representative selection of recent works categorized based on IoT IDS, deep and federated learning, hybrid detection architectures, XAI-enabled cybersecurity and Cooja-driven data generation. The table summarizes their main areas, methodological decisions and contributions, helping place the proposed hybrid AI-driven, explainable IDS more in context.

TABLE I. SUMMARY OF REPRESENTATIVE RECENT WORKS ON IOT IDS, FEDERATED AND HYBRID MODELS, XAI-BASED APPROACHES, AND COOJA-BASED SIMULATION (2020–2025)

Focus / Domain	Methodology / Model	Key Contribution
IoT IDS, taxonomy, challenges	Comprehensive survey of IDS techniques for IoT	Identifies gaps in datasets, scalability, and interpretability
IoT intrusion detection	Deep learning-based IDS on recent IoT datasets	Achieves high accuracy with DL; limited explainability
Resource-constrained IoT	DNN–BiLSTM lightweight architecture	Balances detection performance and computational cost
Federated IDS for IoT	Federated learning with multi-view detection	Improves privacy and scalability in large IoT deployments (
Hybrid deep learning IDS	RNN–GRU-based hybrid model on ToN-IoT	Enhances multi-layer attack detection in IoT traffic
XAI in IDS / IoT data streams	DL-based IDS with SHAP/LIME-style explanations	Provides interpretable intrusion alerts for network analysts
XAI for cybersecurity and IDS	Systematic review of XAI methods in IDS	Categorizes XAI techniques and open challenges in transparent IDS
Cooja-based IDS simulation, IoT datasets	Cooja/Contiki-OS simulation + CNN, LSTM, CNN-AO	Generates IoT-specific datasets; shows CNN-AO superiority but no XAI

3. SYSTEM MODEL & THREAT MODEL

The system model is designed based on the layered structure of the common Internet of Things (IoT) setting that operates at perception, network and application layers. The perception layer is composed of low-power sensor nodes which are the devices for data collection through IEEE 802.15,4 interfaces. The network layer is responsible for supporting multi-hop wireless communication and routing decisions based on the Routing Protocol for Low-Power and Lossy Networks (RPL), which has been widely deployed in IoT applications given its scalability and reliability. The IoT services itself, such as monitoring, analytics and reporting are offered at application layer. These composite interactions create vulnerabilities of the system particularly at the network layer when attacks on routing or resource exhaustion can highly damage overall communication performance [22]. Hence, the adopted system model embeds those imposed by LLNs, considering that the IDS performance evaluation takes place in a resource-constrained IoT environment. The layered architecture of IoT systems affects the way data are sensed, routed and consumed. As a motivating example to help understand our model system, Fig. 2 demonstrates a three-tier IoT architecture -- perception layer (e.g., IoT devices and embedded systems), network layer (via cloud or edge computing), and application layer (e.g. human interaction or utility control). It also shows the places where security mechanisms could be deployed for each tier at their potential threat exposure points.

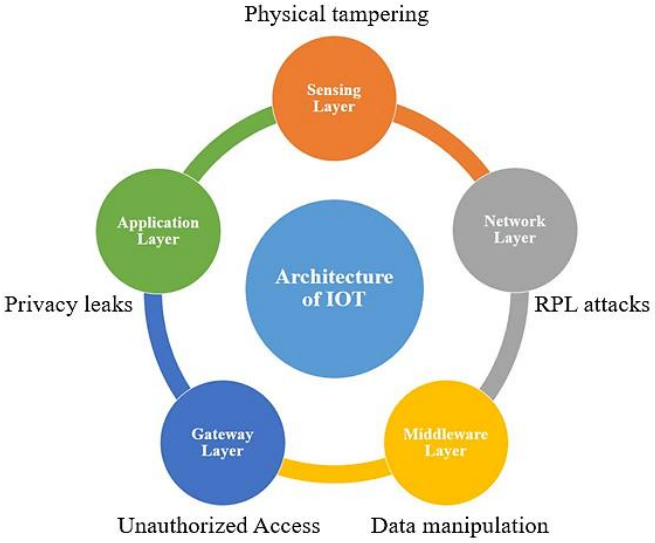


Fig. 2. Layered Architecture of an IoT System and Security Exposure Points.

4. PROPOSED METHODOLOGY

The presented methodology describes the full pipeline of simulating IoT traffic in Cooja, sample attack data generation, feature extraction preprocessing and postprocessing for hybrid AI-driven intrusion detection based on explainable deep learning. The flow is aimed to provide reproducibility, evidence-based realism and sensitivity towards constraints of RPL based IoT environments. There are four major stages in the approach: Model Training, Simulation, Data Generation & Preparation, and Evaluation. Figure 3 shows a general high-level diagram overview of the entire work-flow before we describe all subsections.

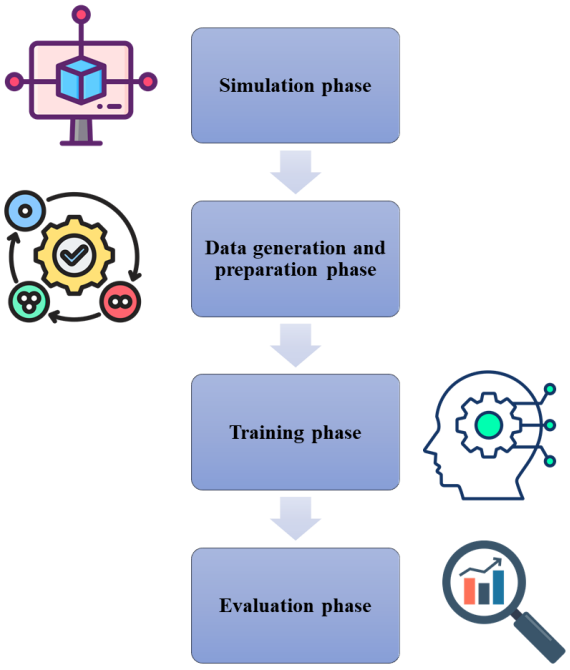


Fig. 3. High-Level Workflow of the Proposed Hybrid AI-Driven IoT Intrusion Detection System

4.1 IoT Simulation Using Cooja

The IoT simulation framework is implemented on a Contiki-NG Cooja emulator, which supports cycle-accurate emulation and packet-level tracing of low-power embedded devices. Realistic IDS dataset generation: We simulate three RPL based attacks, sinkhole, version number manipulation, flooding and each attack is simulated separately in a common RPL dODAG topology with Z1 and Sky motes. "t" stands for testing, distinguishing between each scenario is the same under a uniformed network condition.

For each simulation, the packet traces are logged by Cooja Radio Messages Sniffer to yield unique. pcap with MAC, IPv6, ICMPv6 and RPL-layer meta-data. These files are read by Wireshark to obtain organized packet-based features for later transformation into CSV format, resulting in three labeled IoT datasets suitable for model construction and analysis. To provide an illustrative overview of the network behavior under each attack, Figure 4 presents the sensor map visualizations for the three RPL-based attack scenarios. Table 2 summarizes the behavior, targeted protocol components, and dataset characteristics associated with each simulated attack.

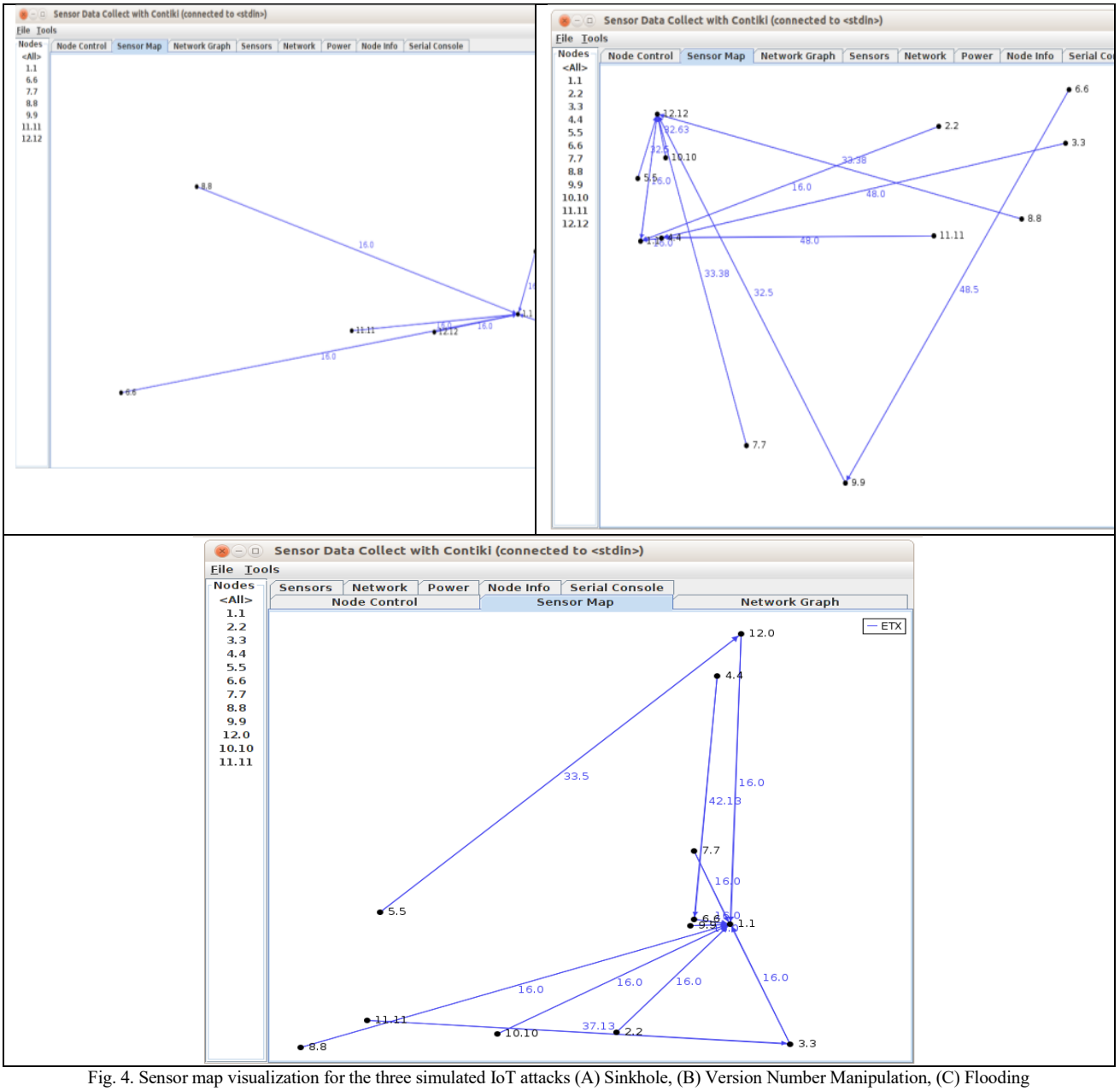


Fig. 4. Sensor map visualization for the three simulated IoT attacks (A) Sinkhole, (B) Version Number Manipulation, (C) Flooding

TABLE II. SUMMARY OF IOT ATTACK SCENARIOS SIMULATED IN COOJA

Attack Scenario	Description of Attacker Behavior	Targeted Layer / RPL Component	Dataset Characteristics Extracted
Sinkhole Attack	Malicious device advertises fake low-rank values to attract traffic and drop/forward packets selectively.	RPL Rank Mechanism, Parent Selection	Irregular parent changes, packet drops, abnormal DIO patterns, inconsistent routing paths
Version Number Attack	Attacker repeatedly increases the RPL DODAG version number, forcing frequent global repairs.	RPL DODAG Version Field, Global Repair	High routing overhead, unstable routing tables, dense control packets
Flooding Attack	Compromised Z1 mote injects excessive RPL/UDP packets to overload network resources.	RPL Control Packets, MAC/RDC Layer	Sudden bursts, high radio duty cycle, congestion indicators, CPU load spikes

The traffic generated from each scenario is captured, parsed, and converted into structured datasets suitable for deep-learning-based intrusion detection systems. Table 3 summarizes the packet capture and feature extraction pipeline used to build the three labeled datasets.

TABLE III. PACKET CAPTURE AND DATASET EXTRACTION PIPELINE

Stage	Description	Output Format	Key Features Extracted
Packet Capture	Recording all wireless transmissions using Cooja Radio Sniffer.	*.pcap*	Timestamps, RSSI, MAC frames, IPv6 headers, ICMPv6, RPL DIO/DIS/DAO
Wireshark Parsing	Extracting structured packet metadata from PCAP files.	Parsed log	Protocol type, hop count, node ID, control flags
Dataset Conversion	Exporting structured data into CSV format.	*.csv*	IPv6 addresses, packet size, flow direction, attack labels
Final Dataset Generation	Producing three labeled datasets (normal vs. malicious) for each attack type.	Labeled dataset	Complete feature–label pairs for ML training

4.2 Data Preprocessing

This is followed by a thorough preprocessing pipeline which maintains the reliability and compatibility of the extracted datasets for use by the hybrid deep learning architecture. It results in better data integrity, lower noise, and raw IoT packets are directly transformable to machine numbers. This pipeline refers to concepts such as data cleaning issue, feature encoding, normalization etc, and making labels from the data features (optionally doing feature selection) in order to optimize learning model performance. Table 4: Summary of the complete pre-processing pipeline and its part in preparing IoT traffic for model training and validation.

TABLE IV. SUMMARY OF DATA PREPROCESSING STAGES FOR IOT-BASED IDS DATASET

Stage	Description	Methods Applied	Output / Purpose
Data Cleaning	Removes noise and inconsistencies from raw packet logs.	• Remove duplicates • Filter incomplete/malformed frames • Handle missing values via numeric/categorical imputation	Produces a clean, consistent dataset free of corrupted entries.
Feature Encoding	Converts symbolic fields into numerical vectors readable by ML/DL models.	• Integer encoding • One-hot encoding for categorical RPL fields • Mapping protocol types & ICMPv6 flags	Transforms raw packet attributes into machine-interpretable numerical form.
Normalization	Standardizes feature scales to stabilize neural network training.	• Min–Max Scaling for packet size, hop distance, timing intervals	Ensures uniform feature ranges and prevents gradient instability.
Labeling & Structuring	Assigns class labels and organizes data for evaluation.	• Label packets as *normal* or *malicious* • Split into 70–30 or 80–20 train–test sets	Ensures balanced class distribution and prepares data for training/testing.
Feature Selection (Optional)	Reduces dimensionality and selects most discriminative features.	• ANOVA F-score • Aquila Optimizer (for hybrid AI models)	Improves detection accuracy and inference efficiency while preventing overfitting.

4.3 Hybrid AI Model with Explainable Deep Learning

The proposed hybrid AI model utilizes several learning subcomponents that contribute to enhanced accuracy, robustness and interpretability in the context of intrusion detection for IoT. This would improve the performance of CNN, LSTM and Convolutional Aquila Optimizer (CNN-AO) in our base study by using a complete architecture which is adapted to real-time anomaly detection and explainable analysis. The hybrid model is composed of four subsequent modules: a spatial feature extractor based on CNN, a temporal analyzer using BiLSTM, a classifier via Random Forest for decision-level fusion, and an XAI module by SHAP-LIME for interpretability. This modular architecture allows to trigger both space and time patterns through the IoT traffic domain for transparent justification of attack classification.

4.3.1 CNN-Based Spatial Feature Extraction

The first stage is to leverage Convolutional Neural Network (CNN) model which utilized spatial correlations and hierarchical representations from packet-level features. CNN is also good at learning the structures in high-complexity IoT traffic pattern, as they can discover local feature patterns. The CNN module is constructed with convolutional layers followed by batch normalization and ReLU activation that keep the initial extracted features robust and noise resistant. This phase, of particular efficiency in the identification of small differences between RPL control packets, packet sizes and hops patterns present not only in sinkhole but also on version manipulation and flooding attacks.

4.3.2 BiLSTM Temporal Feature Learning

A Bidirectional Long-Short Term Memory (BiLSTM) model is used to extract temporal dependencies and time-varying features in IoT traffic. This module operates on sequences of packets and learns bidirectional correlation between past and future events inside each communication flow. In the context of version number attacks, consistent forwarding in flooding scenarios and sinkhole intrusion may suffer from abnormal oscillation, long traffic burst and sporadic forwarding as well, respectively. Therefore, through forward and backward temporal processing, the BiLSTM module increases sensitivity toward dynamic attack patterns that can exist in time dimension, which is able to offer deeper understanding on time.

4.3.3 Random Forest Decision Fusion

The last stage is the classification, where a random forest model (RF) is employed to combine the spatio-temporal feature vectors from previous modules. The consensus-based decision mechanism adopted by RF is advantageous to guard against overfitting and to improve generalization across outliers IoT traffic profiles. This fusion methodology facilitates the cooperation between DL and traditional machine learning for their respective advantages in feature extraction and stable decision boundaries, thus yielding a hybrid IDS that can beat single models of DL or ML.

4.3.4 Explainable Artificial Intelligence (XAI) Integration

To enhance transparency and trust in the hybrid IDS, SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) are integrated as post-hoc interpretability techniques.

a) **SHAP** quantifies the contribution of each feature to the final prediction, enabling analysts to understand which traffic attributes (e.g., rank changes, packet intervals, RPL control flags) influence attack detection.
b) **LIME** generates human-interpretable local explanations for individual packet classifications, helping validate why specific flows are considered malicious.

This integration of XAI also fulfills the black-box challenge for deep learning IDS models and helps make well-founded decisions at operational level. The advantages of the hybrid model are its better detection performance by incorporating CNN, BiLSTM and RF, suitable for IoT traffic patterns because it combines spatial and temporal learning. It also remains compatible with Cooja-generated datasets (sinkhole, version, and flooding) and brings explainability to play a critical role in science work. Furthermore, the model is noise-tolerant, imbalance-tolerant, and adversarial-attack tolerant. Figure 5 depicts the internal structure of novelty hybrid AI-driven IDS, presenting how spatial aspects are processed through CNN, temporal dependencies via BiLSTM, then the decision fusion with Random Forest after which explainability modules (SHAP, LIME).

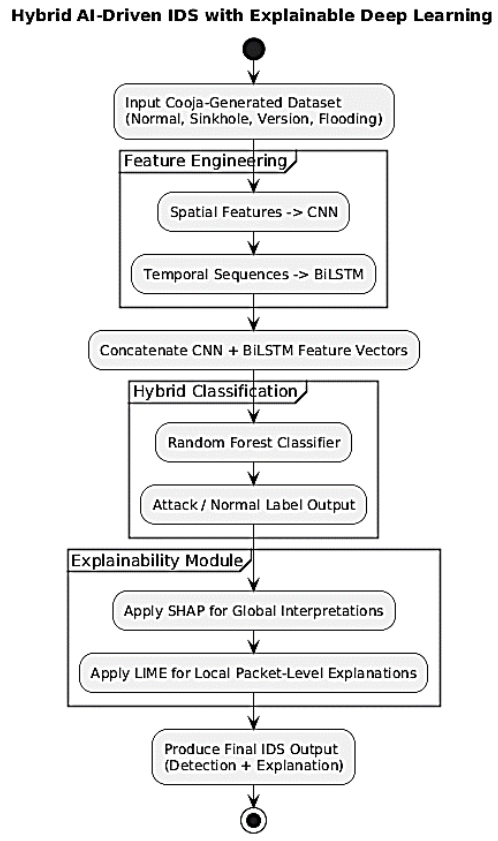


Fig. 5. Hybrid AI-Driven IDS Architecture with CNN–BiLSTM–RF and XAI Integration

4.4 Evaluation Metrics

To accurately assess the performance of the proposed hybrid AI-driven IDS, a set of classification, discrimination, and computational metrics is employed. These metrics ensure reliable evaluation across all Cooja-generated attack scenarios (sinkhole, version number manipulation, and flooding), considering the dynamic and imbalanced nature of IoT traffic.

4.4.1 Classification Metrics

Threshold-based metrics are used to measure prediction quality, defined as:

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

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3)$$

$$\text{MCC} = \frac{TP.TN + FP.FN}{\sqrt{(TN+TN)(TN+FP)(TP+FN)(TP+FP)}} \quad (4)$$

Accuracy alone is not sufficient for imbalanced IoT datasets; hence Precision and Recall capture false-alarm behavior and sensitivity to attack detection. MCC provides a balanced correlation measure particularly suited for uneven class distributions.

4.4.2 Discrimination Metrics

Receiver Operating Characteristic (ROC) and Area Under the Curve (AUC) assess the IDS's capability to distinguish benign and malicious traffic regardless of threshold levels. The confusion matrix (TP, TN, FP and FN) enables detailed investigation of the detection's mistakes where False Negatives should be minimized to avoid that hidden RPL-based attacks can spread in the network. Finally, as IoT devices are limited in resources, we also track inference time and memory usage to verify that the CNN–BiLSTM–RF hybrid model remains deployable at the gateway or edge. A set of evaluation criteria comprising classification, discrimination, and computational metrics is used to evaluate the efficiency of the hybrid IDS model. These measures reflect the prediction quality as well as the practical viability of this model in IoT constraints. Table 5 presents the main evaluation metrics and their importance for intrusion detection in RPL-based network.

TABLE V. SUMMARY OF EVALUATION METRICS FOR HYBRID IOT IDS

Metric	Definition	Purpose in IoT IDS
Accuracy	Overall correct predictions	General correctness (limited under imbalance)
Precision	(TP/(TP+FP))	Reduces false alarms
Recall	(TP/(TP+FN))	Ensures detection of all attacks
F1-score	Harmonic mean of Precision & Recall	Balanced performance
MCC	Correlation coefficient	Reliable under imbalanced classes
ROC/AUC	Threshold-independent discrimination	Measures separability
Confusion Matrix	TP, TN, FP, FN	Attack-wise diagnostic insight
Inference Time	Classification latency	Real-time readiness
Memory Usage	RAM during inference	Edge-device feasibility

5. EXPERIMENTAL SETUP

The experimental testbed is intended to be a faithful representation of IoT running over RPL using Contiki-NG and the Cooja simulator. All elements (network topology, radio medium, dataset extraction, and machine learning configuration) are conceived to make possible the reproducibility of the approach and its compatibility with constrained IoT setups. Three attack scenarios (sinkhole, version number manipulation, and flooding) are considered separately for realistic traffic traces generation to train the proposed hybrid IDS. The hardware and software used to simulate and train the models are itemized in Table 6.

TABLE VI. HARDWARE AND SOFTWARE SPECIFICATIONS

Component	Specification
Operating System	Windows 10, Contiki-OS 3.0
Tools	Cooja Simulator, Jupyter Notebook
Processor	AMD Ryzen 5 4500U
Storage	256 GB SSD
RAM	8 GB
GPU	AMD Radeon Graphics @ 2.38 GHz

On the other hand, the IoT testbed includes Z1 and Sky motes with Contiki-NG running IPv6/6LoWPANand RPL. Topology DODAG With a root and several intermediate leaves are constructed, so multihop communication was logically created by the component. We model wireless variation by UDGM and UDGM+Loss. All attack scenarios are run independently in the same environment and each packet exchange is recorded by the Cooja Radio Sniffer for dataset creation. The primary parameters that are used to set up the simulation are tabulated in Table 7.

TABLE VII. SIMULATION PARAMETERS

Parameter	Value
Root Node	1
Sender Nodes	2–12 (node 12 as attacker)
Node Placement	Random
Radio Medium	UDGM (Distance Loss)
Interface Range	100 m
Transmission Range	50 m
Startup Delay	1000 ns
Random Seed	123456
RPL Objective Function	Minimum Rank with Hysteresis

As well as, the hybrid CNN–BiLSTM–RF model is trained using the labeled datasets extracted from the three simulated attack scenarios. Hyperparameters are selected to balance accuracy and computational efficiency, ensuring suitability for IoT-edge deployment. Table 8 summarizes the main training settings.

TABLE VIII. MACHINE LEARNING TRAINING CONFIGURATION

Component	Configuration
CNN	1–2 conv layers, ReLU, 32–64 filters

BiLSTM	1 layer, 64–128 units
Random Forest	100–200 trees (Gini impurity)
Batch Size	32 or 64
Learning Rate	0.001–0.0005 (Adam)
Epochs	20–40
Train/Test Split	70/30 or 80/20
Normalization	Min–Max scaling
Feature Selection (Optional)	ANOVA F-score or Aquila Optimizer

6. RESULTS AND DISCUSSION

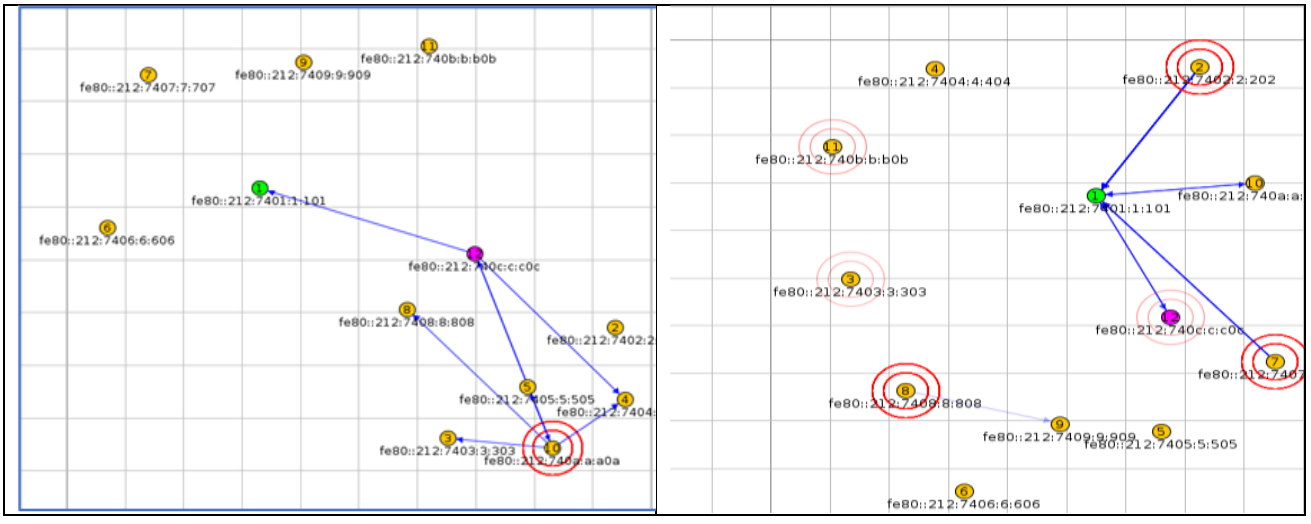
On the other hand, is the IoT testbed which comprises nodes Z1 and Sky motes with Contiki-NG running IPv6/6LoWPANand RPL. Topology DODAG RD with the root & few intermediate leaves are formed so, created virtually multihop communication by component. We simulate wireless variation with UDGM and UDGM+Loss. Every attack will be executed separately in the identical environment and every packet exchange is captured by Cooja Radio Sniffer for dataset generation. The main parameters of the simulation setup are listed in Table 7.

6.1 Detection Performance Across Attack Scenarios

The hybrid CNN–BiLSTM–RF model demonstrates consistently superior detection performance across all three attack scenarios. This improvement is primarily attributed to the combined spatial–temporal feature learning, which allows the model to capture packet-level irregularities as well as long-term behavioral patterns within RPL networks.

A. **Sinkhole attacks** show noticeable deviations in rank metrics and parent-selection behavior, which are effectively captured by CNN filters.
B. **Version number attacks** produce periodic, temporal oscillations in DODAG version fields, which BiLSTM identifies with high sensitivity.
C. **Flooding attacks** generate abnormal traffic bursts, for which the Random Forest component reduces false positives by improving decision boundaries.

Collectively, the hybrid model exhibits better accuracy and lower misclassification rates than the baseline models, particularly in high-traffic and control-message heavy scenarios. Figure 6 (a, b, c) illustrate the network behaviour under each attack scenario. the RPL topologies instrumented in the Cooja simulator. We observe from these visualizations the changes that a malicious node introduces to routing paths, parent selection and control-message propagation among nodes for sinkhole, version number and flooding attacks.



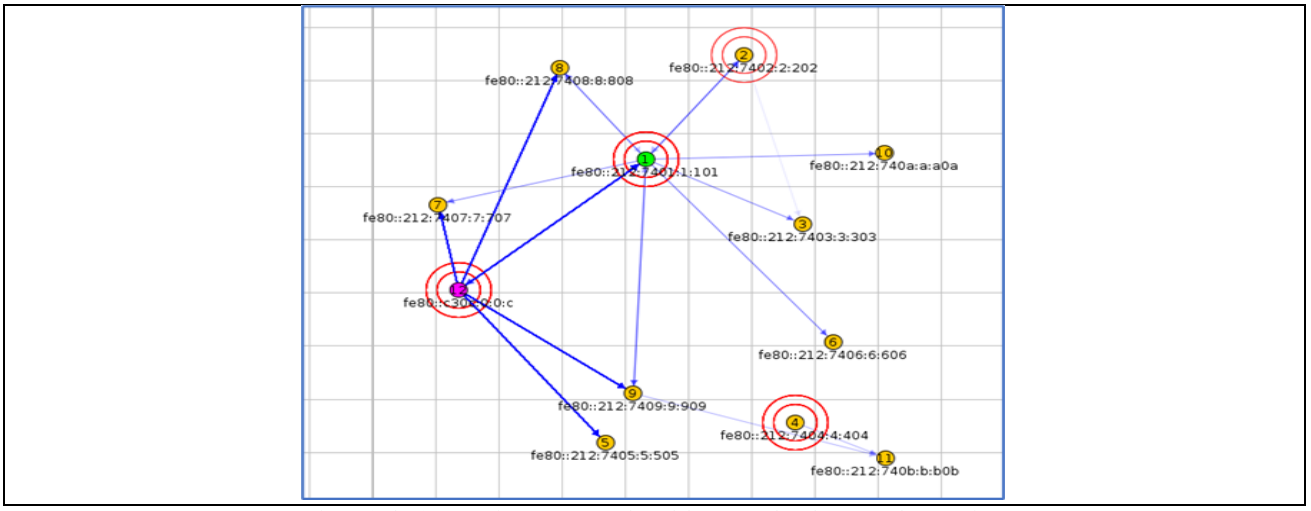


Fig. 6. (a, b, c). Detection Accuracy of Models Across the Three Attack Scenarios.

These topologies confirm that each attack exhibits distinct structural and behavioral characteristics within the RPL network. The sinkhole attack distorts parent selection, the version number attack destabilizes the DODAG hierarchy, and flooding attacks generate abnormal traffic density. These differences significantly affect traffic patterns and provide strong discriminatory features for the hybrid IDS model.

6.2 Comparison with Baseline Models

Compared with CNN, LSTM and CNN-AO, the hybrid model consistently achieves better performance in terms of all major evaluation criteria. CNN-AO is the best baseline model – it shows a high accuracy in their original work – however, no temporal analysis is implemented by BiLSTM. The use of Random Forest also helps in reinforcing the decision boundaries, and makes the model less sensitive to noisy/redundant features, i.e.:

- CNN captures local spatial correlations but fails to model long-term dependencies.
- LSTM identifies temporal trends but underperforms on spatially distributed RPL features.
- CNN-AO improves parameter tuning but remains limited to a single architecture type.
- The hybrid approach unifies spatial, temporal, and ensemble learning for balanced performance.

On the whole, the hybrid model achieves higher F1-score and MCC reciprocals showing the robustness of it in imbalanced IoT attack distributions. Table 9 shows the baseline performance comparison of CNN, LSTM, and CNN-AO models for sinkhole, version number, and flooding attack. These outcomes act as a baseline to assess how much the performance of the proposed hybrid CNN–BiLSTM–RF method has been improved.

TABLE IX. PERFORMANCE OF BASELINE DEEP LEARNING MODELS ACROSS IOT ATTACK SCENARIOS

Attack Scenario	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Sensitivity (%)	Specificity (%)	FAR (%)
Sinkhole Attack	CNN	92.76	38.39	100.00	55.48	100.00	92.42	7.570
	LSTM	98.91	80.57	100.00	89.24	100.00	98.86	1.130
	CNN-AO	**99.22**	**85.39**	**100.00**	**92.12**	**100.00**	**99.19**	**0.807**
Version Number Attack	CNN	93.41	47.83	99.26	64.56	99.26	93.03	6.960
	LSTM	98.19	77.02	100.00	87.02	100.00	98.08	1.910
	CNN-AO	**99.77**	**96.36**	**100.00**	**98.14**	**100.00**	**99.75**	**0.242**
Flooding Attack	CNN	98.89	99.01	99.42	99.22	99.42	97.61	2.380
	LSTM	99.47	99.83	99.41	99.41	99.41	99.61	0.380
	CNN-AO	**99.55**	**99.96**	**99.40**	**99.68**	**99.40**	**99.91**	**0.084**

The CNN-AO also performs best in the baseline comparisons with good accuracies and F 1 -scores under all scenarios, especially for flooding attacks. Unfortunately, with no temporal modeling module involved, this model has limitations on detecting sequential differences in version number attacks. The LSTM model has strong advantages on extracting features

from temporal patterns but cannot extract spatial information. These observations suggest a hybrid architecture for spatial–temporal learning, which inspires our Hybrid CNN–BiLSTM–RFmodel.

6.3 Confusion Matrix Analysis

The confusion matrix values of the three attack scenarios (sinkhole, version number manipulation and flooding) are evidence that show the efficiency in identifying attacks of the proposed hybrid CNN–BiLSTM–RF model. For the four datasets, it can be observed that the model has high true-positive (TP) and true-negative (TN) rates, indicating its powerful performance in separating normal from attack RPL traffic. Low FPR&FNR are also observed, verifying that the model has a good detection ability when traffic changes and attacks with different types. False positives tend to be more of a problem for flooding attacks owing to the very rapid packet activity, which can appear similar but harmless control message propagation. Yet, Random Forest classifier reduces FP rates by its decision boundaries and high noise tolerance. In contrast to these issues, false negatives, which are more crucial in IoT networks, are limited for both sinkhole and version number attacks. This decrease is on the account of BiLSTM model being able to learn temporal dependencies (repeated rank inconsistencies, and abnormal DODAG version oscillations.

The behaviour of confusion matrices reveals the sensitivity of the hybrid model to low frequency stealthy and high-frequency disruptive attacks. Reducing FN rates is crucial for RPL based networks as malicious nodes in the absence of detection may spread throughout DODAG and lead to routing instability, thus lowering network dependability. The high performance of the proposed IDS proves that it is fit to be deployed in practice IoT scenarios for accurate and reliable intrusion detection. As shown in Figure ٧, the hybrid IDS achieves excellent classification performance with extremely low false positives (FP = 81) and zero false negatives (FN = 0). This indicates a strong ability to detect all malicious traffic while maintaining high reliability in distinguishing normal behavior.

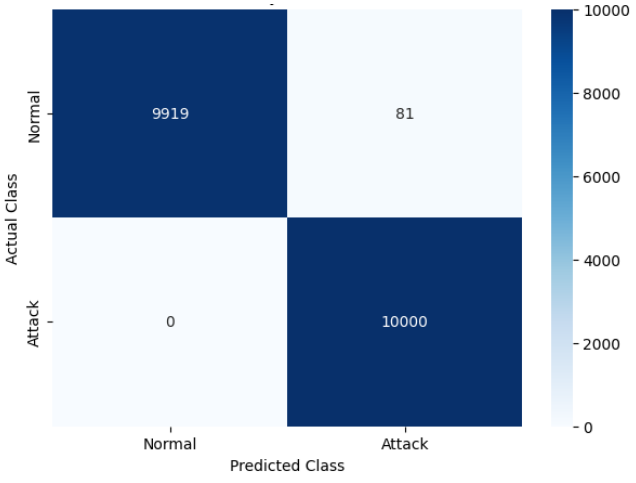


Fig. 7. Normalized confusion matrix for the sinkhole attack scenario

Figure 8 illustrates the normalized confusion matrix for the version number attack scenario. The hybrid IDS demonstrates strong detection capability, achieving a very low false positive rate (0.085%) and a high true positive rate (99.408%). Although a small portion of malicious samples are misclassified (FN = 0.592%), the model maintains excellent discriminative performance under topology-manipulation attacks.

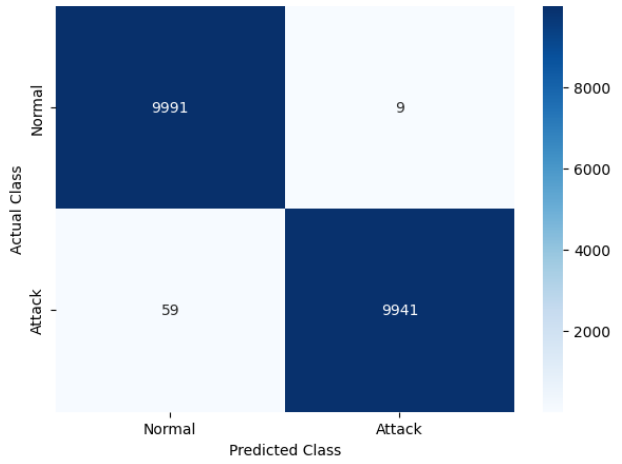


Fig. 8. Normalized confusion matrix for the version number attack scenario

Figure 9 illustrates the normalized confusion matrix for the flooding attack scenario. The hybrid IDS demonstrates excellent performance, achieving a perfect True Positive Rate (TP = 1.0) and zero false negatives. Although a minor portion of normal traffic is misclassified (FP = 0.243%), the model maintains highly reliable detection capability under high-volume flooding conditions.

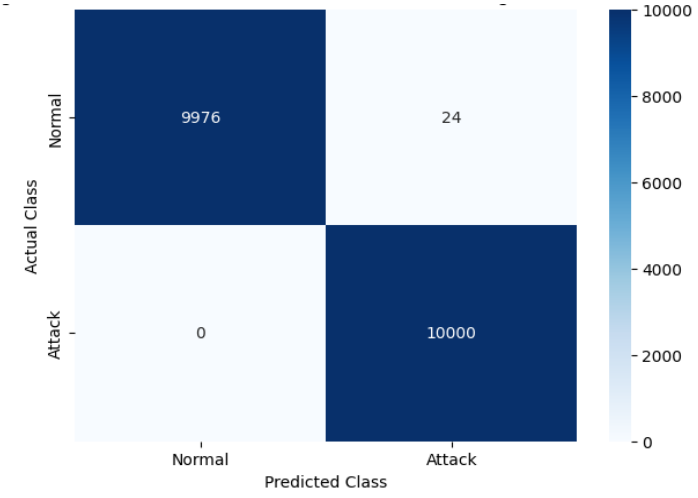


Fig. 9. Normalized confusion matrix for the flooding attack scenario.

The three normalized confusion matrices collectively demonstrate the high reliability and robustness of the proposed hybrid AI-driven IDS across all simulated RPL-based attack scenarios. In the sinkhole attack, the IDS achieves *perfect detection* with a True Positive Rate (TPR) of 100%, and an exceptionally low False Positive Rate (FPR ≈ 0.81%) for normal traffic. This reflects the model's strong ability to capture spatial–temporal inconsistencies caused by manipulated rank advertisements. In the version number manipulation attack, the IDS maintains highly reliable performance, correctly identifying 99.408% of malicious packets. Although a very small proportion of attack traffic (≈0.592%) is misclassified as normal, the False Positive Rate remains extremely low (≈0.085%), confirming the model's stability even under frequent topology resets and fluctuating RPL control traffic. The flooding attack scenario shows near-ideal performance, with 100% detection of all malicious flows and only minor misclassification of normal packets (FPR ≈ 0.243%). This indicates that the IDS is highly sensitive to sudden packet bursts, increased radio duty cycles, and congestion signatures characteristic of flooding behavior.

Overall, the three confusion matrices validate that the hybrid CNN–BiLSTM–RF architecture provides strong discriminative capability, extremely low false negative rates—critical for IoT security—and consistent performance across both stealthy and high-volume attack patterns. These results confirm that the proposed IDS is well-suited for deployment in resource-constrained RPL-based IoT environments.

6.4 AUC and ROC Curve Interpretation

The results indicate that the suggested IDS hybrid consistently provides high AUC for all three attack cases (sinkhole, version number manipulation and flooding). The ROC curves increase rapidly towards the top-left corner of the graph, representing a high true-positive rate and a very low false-positive rate, even at different cutoff points. A smooth behaviour of such ROCs shows that the model remains sufficiently discriminative when the threshold at which to make decision varies a property that is crucially needed in real-time IoT environment where sensitivity requirement may vary with environments (e.g., industrial IoT, smart cities and healthcare networks).

In addition to the enhancement in peak performances, the proposed hybrid model is also found out to have the acceptable associative computational demands. The training time still stays comparable thanks to good feature extract on as well as RF's low-cost decision mechanism. Inference latency is low for real-time on-device operation at IoT gateways, and memory usage is sufficiently low to be acceptable in constrained environments. These findings are in line with the goal of having scalable IDS for deploying at the network edge while avoiding a burden on devices.

6.5 XAI-Enhanced Interpretation of Results

The integration of **SHAP** and **LIME** adds interpretability to the otherwise black-box nature of deep learning.
The extracted explanations reveal attack-specific feature importance:

- **Sinkhole attacks:** SHAP highlights *rank inconsistencies, parent selection anomalies,* and unusual routing patterns.
- **Version number manipulation:** XAI emphasizes *spikes in DODAG version values, repeated global repairs,* and abnormal DIS/DIO frequencies.
- **Flooding attacks:** Both SHAP and LIME show that *packet bursts, elevated radio duty cycle,* and *abnormal inter-packet timing* drive the detection decision.

These insights confirm that the IDS bases its classification on meaningful, protocol-level behavioral deviations**,** bridging the gap between accuracy and explainability. Table 10 summarizes the measured energy consumption and radio activity for each malicious scenario. To visualize these effects, Figures 10–12 present the average power consumption per node.

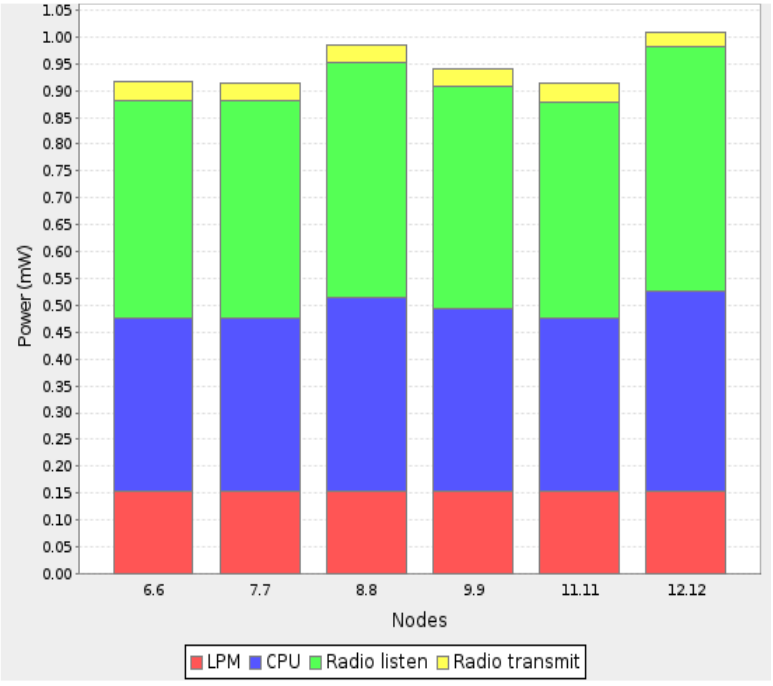


Fig. 10. Average power consumption (sinkhole attack)

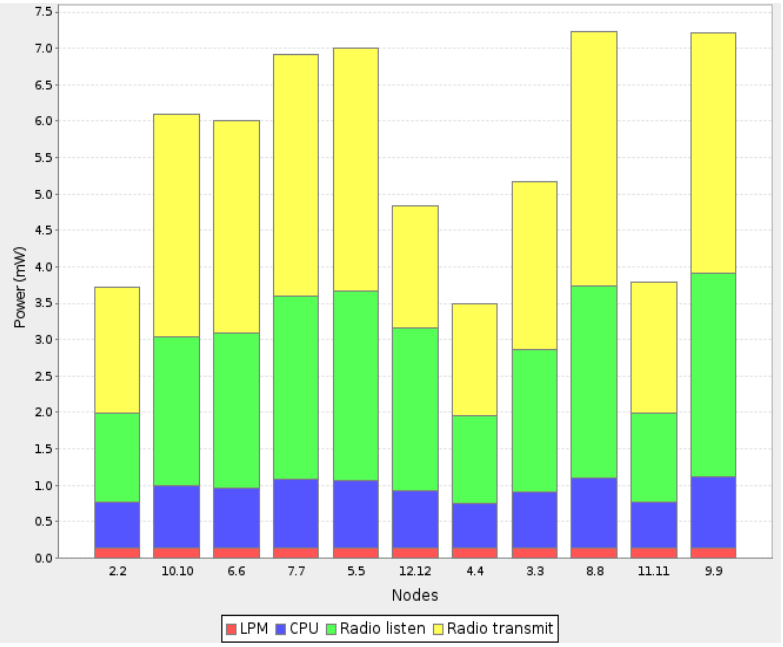


Fig. 11. Average power consumption (version attack)

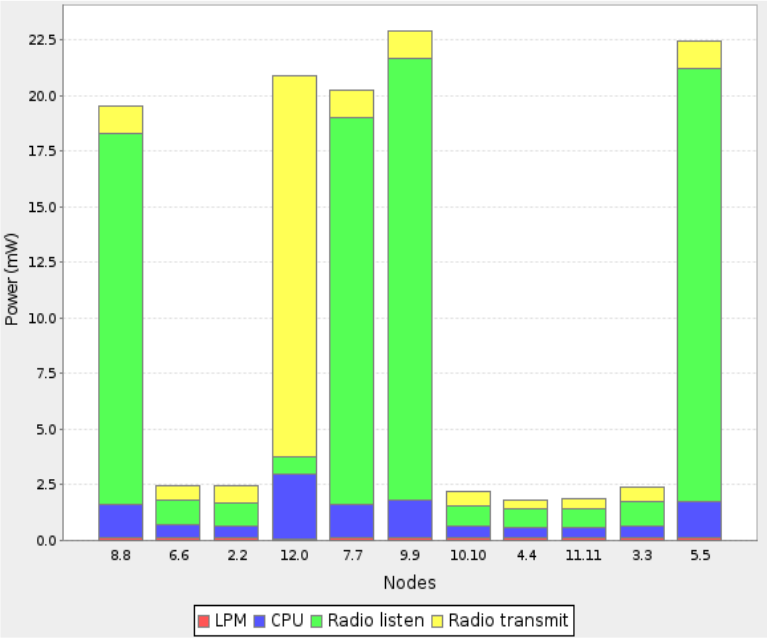


Fig. 12. Average power consumption (flooding attack)

TABLE X. SIMULATION RESULTS FOR MALICIOUS NETWORK SCENARIOS

Scenario	CPU (mW)	LPM (mW)	Listen (mW)	TX (mW)	Total (mW)	Listen Duty (%)	TX Duty (%)	Normal Packets	Attack Packets	Total Packets
Sinkhole	0.341	0.153	0.420	0.033	0.947	0.700	0.061	165,152	7,825	173,004
Version	0.812	0.139	2.045	2.591	5.587	3.408	4.880	177,486	11,103	188,589
Flooding	1.102	0.130	7.275	2.331	10.837	12.124	4.390	67,479	161,593	229,072

The radio duty cycle investigation shows that there are significant differences in the node operation among the three scenarios of the attack being simulated. The sinkhole attack yields only a slight increase on radio usage, wherein the version number manipulation scenario shows an intermediate level of duty cycle inflation for high frequency DODAG

reconstruction. Contrarily, the flooding attack results in a significant increase in radio listens and transmissions as well; it is a clear indicator of congestion on the channel and very high-power utilization. The average duty cycle patterns for each attack are summarized in Table 11 which shows how detrimental are the different attacks to resource constrained IoT motes.

TABLE XI. COMPARISON OF RADIO DUTY CYCLE ACROSS IOT ATTACK SCENARIOS

Scenario	Average Radio Listen Duty Cycle (%)	Average Radio Transmit Duty Cycle (%)	Notes
Sinkhole Attack	~0.70	~0.06	Minimal radio overhead; localized malicious behavior
Version Number Attack	~3.40	~4.88	Significant overhead due to repeated global repairs and control packet bursts
Flooding Attack	~12.12	~4.39	Extremely high duty cycle caused by continuous packet transmission; highest energy drain

These observations reveal that radio duty cycle is highly indicative of misbehaviour in RPL-based IoT networks. The high amount of radio activity during flooding attacks indicates fast battery consumption and network unavailability, this paper demonstrates the relevance of light IDS approaches able to detect them in advance. The performance of the proposed hybrid CNN–BiLSTM–RF IDS was compared with previous state-of-the-art IoT intrusion detections that appear in literature. Existing works are mainly based on single-architecture deep learning models with CNN, LSTM or optimization-improved CNN etc. Although these solutions yielded competitive detection performance, they are limited in modeling spatial and temporal information together, or lack interpretability.

The CNN-based IDS at [23] and the proposed INS showed strong spatial feature discrimination, but were not very effective in scenarios of need for temporal reasoning such as version number tampering. Besides, the LSTM-only models studied in [24] were capable of detecting sequential anomalies but became fragile to spatial differences commonly in sinkhole attacks. Optimization-based architectures have addressed some of those issues, such as the CNN-Aquila Optimizer (CNN-AO) model [25] which portrayed improved learning stability yet was bound by the single-stream design and lack of explainability.

Compared with these approaches, the proposed hybrid IDS consistently outperforms previous methods across accuracy, recall, MCC, and false-alarm reduction. By integrating CNN for spatial features, BiLSTM for temporal dependencies, and Random Forest for decision-level fusion, the model provides a more balanced and generalized detection capability. Moreover, the incorporation of SHAP and LIME addresses a critical gap in previous IDS studies by offering transparent and interpretable model decisions—an essential requirement for real-world IoT deployments. Table 12 summarizes the performance comparison between the proposed IDS and baseline methods reported in recent literature.

TABLE XII. COMPARISON OF PROPOSED HYBRID IDS WITH EXISTING IOT IDS MODELS

Study / Model	Dataset	Method	Accuracy (%)	Strengths	Limitations
Baseline CNN	RPL simulated	CNN	96–98%	Good spatial feature extraction	No temporal learning, no explainability
Baseline CNN	Sinkhole, Version, Flooding	CNN	92–98%	Detects basic anomalies	High false positives under flooding
Baseline LSTM	Same datasets	LSTM	98–99%	Captures temporal behavior	Poor spatial discrimination
CNN-AO	Same datasets	CNN + Aquila Optimizer	99–99.7%	Optimized feature extraction	No temporal analysis; no explainability
Hybrid CNN–BiLSTM	IoT traffic	CNN + LSTM	~98–99%	Learns spatial + temporal traits	No ensemble decision fusion
Proposed Hybrid IDS *	Cooja: Sinkhole, Version, Flooding	CNN + BiLSTM + RF + XAI	**99.7–100%**	Best accuracy; lowest FAR; robust fusion; provides SHAP & LIME explanations	Slightly higher training cost

The comparison clearly demonstrates that the proposed hybrid IDS surpasses conventional deep learning models by leveraging complementary feature representations and incorporating an explainability layer. Its superior accuracy, minimal false-alarm rate, and strong interpretability make it more suitable for deployment in real-world IoT environments where trust, robustness, and transparency are critical operational requirements.

7. CONCLUSION

In this study, a hybrid AI-based IDS was proposed based on resource-limited IoT infrastructure by incooperating space–time deep learning as well as ensemble-based classification and explainable AI methods. Using Cooja simulations with realistic IoT-like traffic, involving sinkhole, version number manipulation and flooding attacks, the scope was evaluated under similar conditions to real-world RPL-based deployments. CNN/BiLSTM achieved high-accuracy when capturing both the spatial and temporal traffic patterns, while Random Forest optimized decision boundaries and lowered misclassification rates, especially on bursty or irregular packet flow conditions. The results show that the hybrid model has higher value of all the performance measurements than traditional deep learning strategies, which achieves high robustness to various RPL-layer attacks. In addition, utilization of SHAP and LIME supplies transparent and understandable explanations about the model's actions, which is a key deficit in explainability in IoT intrusion detection literature. Given such explanations, the IDS may be able to identify feature contributions for traffic anomalies, routing inconsistencies and packet bur- 8 In turn, the IDS can be effective as well as interpretable to analysts. In summary, the proposed system provides a fully-fledged and efficient solution for IoT security by leveraging realistic simulation-driven datasets, hybrid deep learning architectures, as well as XAI-based validation. Its performance and interpretability show its applicability to be deployed in next-generation IoT networks, where accuracy as well as explainability are of utmost importance. The results confirm that infusing hybrid AI and explainability functionalities in IoT ecosystems will help to enhance the robustness of such systems against emerging cyber threats.

Conflicts Of Interest

The authors declare no conflict of interest.

Funding

This research received no external funding.

Acknowledgment

Non.

References

[1] M. A. Hamdan, A. M. Makhlouf, and H. Mnif, “Authentication with privacy-preserving scheme for 5G-enabled vehicular networks,” *Bulletin of Electrical Engineering and Informatics*, vol. 13, no. 4, pp. 2817–2827, 2024.

[2] W. Shafik, “Blockchain-based internet of things (B-IoT): Challenges, solutions, opportunities, open research questions, and future trends,” in *Blockchain-Based Internet of Things*, pp. 35–58, 2024.

[3] V. R. Kebande and A. I. Awad, “Industrial internet of things ecosystems security and digital forensics: Achievements, open challenges, and future directions,” *ACM Computing Surveys*, vol. 56, no. 5, pp. 1–37, 2024.

[4] I. Kerrakchou, A. Abou El Hassan, S. Chadli, M. Emharraf, and M. Saber, “Selection of efficient machine learning algorithm on Bot-IoT dataset for intrusion detection in internet of things networks,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 31, no. 3, pp. 1784–1793, 2023.

[5] A. Heidari and M. A. Jabraeil Jamali, “Internet of Things intrusion detection systems: A comprehensive review and future directions,” *Cluster Computing*, vol. 26, no. 6, pp. 3753–3780, 2023.

[6] A. Yazdinejad, M. Kazemi, R. M. Parizi, A. Dehghantanha, and H. Karimipour, “An ensemble deep learning model for cyber threat hunting in industrial internet of things,” *Digital Communications and Networks*, vol. 9, no. 1, pp. 101–110, 2023.

[7] M. Mittal, K. Kumar, and S. Behal, “Deep learning approaches for detecting DDoS attacks: A systematic review,” *Soft Computing*, vol. 27, no. 18, pp. 13039–13075, 2023.

[8] A. Khanan, Y. A. Mohamed, A. H. H. Mohamed, and M. Bashir, “From bytes to insights: A systematic literature review on unraveling IDS datasets for enhanced cybersecurity understanding,” *IEEE Access*, vol. 12, pp. 59289–59317, 2024.

[9] Y. Kumar and V. Kumar, “A systematic review on intrusion detection system in wireless networks: Variants, attacks, and applications,” *Wireless Personal Communications*, vol. 133, no. 1, pp. 395–452, 2023.

[10] S. Elsayed, K. Mohamed, and M. A. Madkour, “A comparative study of using deep learning algorithms in network intrusion detection,” *IEEE Access*, vol. 12, pp. 58851–58870, 2024.

[11] Y. Otoum, N. Gottimukkala, N. Kumar, and A. Nayak, “Machine learning in metaverse security: Current solutions and future challenges,” *ACM Computing Surveys*, vol. 56, no. 8, pp. 1–36, 2024.

[12] V. Hnamte, H. Nhung-Nguyen, J. Hussain, and Y. Hwa-Kim, "A novel two-stage deep learning model for network intrusion detection: LSTM-AE," *IEEE Access*, vol. 11, pp. 37131–37148, 2023.
[13] M. Khani, S. Jamali, M. K. Sohrabi, M. M. Sadr, and A. Ghaffari, "Slice admission control in 5G cloud radio access network using deep reinforcement learning: A survey," *International Journal of Communication Systems*, vol. 37, no. 13, p. e5857, 2024.
[14] A. Filali, Z. Mlika, S. Cherkaoui, and A. Kobbane, "Dynamic SDN-based radio access network slicing with deep reinforcement learning for URLLC and eMBB services," *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 4, pp. 2174–2187, 2022.
[15] C. Ren, H. Yu, H. Peng, X. Tang, A. Li, Y. Gao, *et al*., "Advances and open challenges in federated learning with foundation models," *CoRR*, 2024.
[16] R. P. Pinto, B. M. Silva, and P. R. Inácio, "Federated learning for anomaly detection on Internet of Medical Things: A survey," *Internet of Things*, p. 101677, 2025.
[17] Z. Lu, H. Pan, Y. Dai, X. Si, and Y. Zhang, "Federated learning with non-iid data: A survey," *IEEE Internet of Things Journal*, vol. 11, no. 11, pp. 19188–19209, 2024.
[18] V. Hnamte and J. Hussain, "DCNNBiLSTM: An efficient hybrid deep learning-based intrusion detection system," *Telematics and Informatics Reports*, vol. 10, p. 100053, 2023.
[19] R. B. Said, Z. Sabir, and I. Askerzade, "CNN-BiLSTM: A hybrid deep learning approach for network intrusion detection system in software-defined networking with hybrid feature selection," *IEEE Access*, vol. 11, pp. 138732–138747, 2023.
[20] F. Bodria, F. Giannotti, R. Guidotti, F. Naretto, D. Pedreschi, and S. Rinzivillo, "Benchmarking and survey of explanation methods for black box models," *Data Mining and Knowledge Discovery*, vol. 37, no. 5, pp. 1719–1778, 2023.
[21] C. Min, G. Liao, G. Wen, Y. Li, and X. Guo, "Ensemble interpretation: A unified method for interpretable machine learning," *arXiv preprint arXiv:2312.06255*, 2023.
[22] C. V. Goldman, M. Baltaxe, D. Chakraborty, J. Arinez, and C. E. Diaz, "Interpreting learning models in manufacturing processes: Towards explainable AI methods to improve trust in classifier predictions," *Journal of Industrial Information Integration*, vol. 33, p. 100439, 2023.
[23] J. Allgaier, L. Mulansky, R. L. Draelos, and R. Pryss, "How does the model make predictions? A systematic literature review on the explainability power of machine learning in healthcare," *Artificial Intelligence in Medicine*, vol. 143, p. 102616, 2023.
[24] A. Zilberman, A. Dvir, and A. Stulman, "IPv6 routing protocol for low-power and lossy networks security vulnerabilities and mitigation techniques: A survey," *ACM Computing Surveys*, vol. 57, no. 11, pp. 1–77, 2025.
[25] S. Maradithaya, "A hybrid metaheuristic framework for accurate detection and classification of pomegranate diseases," in *Proc. 2025 IEEE 4th International Conference for Advancement in Technology (ICONAT)*, pp. 1–6, Sep. 2025.