The Indonesian Journal of Computer Science
www.ijcs.net
Volume 15, Issue 5, October 2026
https://doi.org/10.33022/ijcs.v15i5.5228

Confidence-Based Hierarchical Framework for Known and Zero-Day Attack Detection in SDN-IIoT Networks

Reorapetse Molose[1], Bassey Isong[2]
reorapetse.molose@nwu.ac.za[1], bassey.isong@nwu.ac.za[2]
[1,2]Computer Science Department, North-West University, Mafikeng, South Africa

Article Information

Received : 12 Aug 2026
Revised : 25 Sep 2026
Accepted : 23 Oct 2026

Keywords

SDN-IIoT security, zero-day attack detection, NID, feature selection optimisation, anomaly detection, hybrid ensemble learning

Abstract

SDN-IIoT networks face increasing security risks due to centralised control logic, open southbound interfaces, and continuous inter-device signalling. These factors expand the attack surface and allow failures to cascade across interconnected components. Existing intrusion detection solutions focus heavily on closed-set evaluation, a single learning paradigm, or the isolation of zero-day threats, leaving a void for approaches that consistently detect both known and unknown attacks. This paper proposes a hierarchical anomaly detection framework combining SHAP-guided Improved Binary Sea Wasp Optimisation (IBSWO) feature selection, a Random Forest binary gate for traffic filtering, and a confidence-based routing mechanism that coordinates an OpenSet ExtraTrees classifier, a contrastive representation-based detector, and a Local Outlier Factor (LOF) detector to classify suspicious samples as known or zero-day attacks. The framework was evaluated through simulation on three benchmark datasets (UNSW-NB15, Edge-IIoTset, and TON-IoT) under two protocols: attack-class holdout and outlier-based novelty detection. Under held-out evaluation, it achieves Macro-F1 scores of 0.73-0.87 and stable Novelty-F1 scores of 0.74-0.76. Under outlier-based evaluation, the hierarchical density-based and one-class variants both reach Macro-F1 of at least 0.90, and the fully integrated detector eliminates false positives (FPR=0.00) while retaining perfect attack recall on Edge-IIoTset. Against recent zero-day IDS approaches, the framework matches or exceeds reported unknown-attack recall and novelty-F1 scores while holding up more evenly across the three datasets than baselines validated on fewer benchmarks. Confidence-based detector routing, rather than a single model or static ensemble voting, offers an efficient path toward reliable known- and zero-day attack detection in SDN-IIoT settings.

Attribution-ShareAlike 4.0 International License

A. Introduction

The fast growth of Industry 4.0-driven Industrial Internet of Things (IIoT) technology has led to extensive digitisation in the manufacturing industry [1]. This shift has exposed operational technology to greater cyber threats, making security a major concern in IoT environments. A well-coordinated cyberattack on IIoT systems can cause serious economic damage and may violate safety and environmental regulations [2], [3]. This makes secure and reliable communication necessary, supporting both horizontal integration across different fields and vertical integration across the whole IIoT system within organisations. Cybercriminals now have the knowledge and tools to launch sophisticated attacks on IIoT systems [4]. Software-defined networking (SDN) is an architectural approach for managing and orchestrating networks and their dynamic applications [5]. Originally developed as an ICT networking solution for data centres, SDN is now gaining attention in IIoT, where individual devices embody Industry 4.0 principles and integrate control elements under a central SDN controller. In the multilevel structure of IIoT environments, this facilitates autonomous interaction between control systems [5], [6]. SDN provides a flexible method for managing various industrial setups across cloud and on-premises infrastructure as IIoT settings grow into vast, linked networks [7].

The integration of SDN and IIoT has also brought about challenges with inefficient resource management and increasing system complexity [8]. The development of SDN-enabled IIoT architectures has been driven by SDN's ability to offer centralised, programmable network administration, which solves several problems. Despite these benefits, SDN-IIoT networks continue to face persistent security and privacy challenges that threaten network reliability and scalability [8], [9]. In particular, the separation of the control and data planes, reliance on open southbound interfaces, and continuous exchange of control messages across interconnected industrial devices broaden the attack surface and create multiple pathways for threats to propagate [10]. Because of these architectural characteristics, SDN-IIoT settings are especially susceptible to network attacks, such as flooding and traffic saturation attacks, where a breach in one area of the network can have a domino effect on other related components [8], [9] and [10].

Despite progress in recent NID approaches for IoT, IIoT and SDN-based environments, several limitations remain. Many existing solutions are still evaluated under closed-set conditions [11], [12], [13], rely on deep learning models [12] ,[13], [14], or address zero-day attacks in isolation [14], [15]. Consequently, existing approaches do not sufficientl provide lightweight and integrated mechanism for classifying known attacks while also detecting unseen attacks across heteregeneous IoT/IIoT datasets. This constitutes the specific research gap addressed in this study.

This points to a need for a lightweight, adaptive framework that can detect both known and unseen attacks while supporting practical mitigation within realistic SDN-IIoT environments. Addressing these evolving threats requires solutions that go beyond single-model approaches. One promising direction combines multiple techniques into a hybrid framework that draws on different learning paradigms. This paper proposes a hybrid hierarchical ensemble for SDN-IIoT anomaly detection that integrates supervised and unsupervised learning to

address known and zero-day attacks. Specifically, this work focuses on designing and implementing a hybrid anomaly detection construct that identifies both known and zero-day attacks. The implemented construct focuses on detection and classification rather than automated mitigation. Its detection outputs may, however, support future implementation adaptaptive SDN-based mitigation strategies [16]. The main contributions of this paper are summarised as follows:

- A novel hierarchical zero-day anomaly detection framework that integrates supervised and novelty-based learning for detecting known and unknown attacks in IoT/IIoT networks.
- Introduction of a lightweight SHAP-guided LightGBM-IBSWO feature optimisation method for efficient and discriminative feature selection.
- Development of a confidence-based adaptive routing mechanism that coordinates complementary detectors for robust zero-day attack detection.
- A comprehensive dual-evaluation strategy is established using attack class holdout and novelty detection across multiple heterogeneous IoT/IIoT benchmark datasets.

This study is structured as follows: section B presents related work. Section C, discusses the method and process applied. Section D presents the findings of the study. Section E, discusses the main findings and Section F concludes the paper presenting future works.

B. Related Work

Recent studies have focused on machine learning, deep learning and hybrid strategies to detect anomalies or cyberattacks in SDN, IIoT and SDN-IIoT networks. This section reviews and categorises these studies by technique and by their ability to detect known and zero-day attacks. Table I summarises the studies.

B.1. Machine Learning

Several studies have proposed machine learning-based solutions for anomaly detection in SDN, IIoT and IoT environments, mainly to improve feature selection and detection accuracy. Kumar and Ansari [17] proposed a lightweight explainable intrusion detection system for SD-IoT, combining nature-inspired optimisation with LASSO to reduce complexity while maintaining performance. Punia et al. [18] enhanced anomaly detection in IoT networks using a swarm-based whale optimisation algorithm combined with feature selection techniques including ANOVA, reporting close-to-perfect accuracy. Jain et al. [15] focused on zero-day attacks in IoT networks using Zd-UML, an unsupervised model paired with an autoencoder-based modified clustering algorithm, achieving 98% accuracy on the IoT-23 and CICIoT2023 datasets. Okfie and Mishra [19] combined isolation forest (ML) with a lightweight BC (bitcoin transaction network) solution to detect IIoT anomalies. Yıldız and Balta [20] compared different ML models in SDN-IoT for smart grid applications, with random forest performing best

B.2. Deep Learning

Deep learning studies have focused on using neural networks to capture complex, high-dimensional patterns in IIoT and IoT data, particularly where traditional solutions struggle. Chen et al. [21] combined an optimised LSTM model

with XGBoost-based feature selection to handle imbalanced IIoT datasets, achieving high AUC and low false alarm rates. Rhachi et al. [22] used autoencoders with ANOVA-based feature selection to improve anomaly detection in IoT networks, achieving 85% accuracy, an improvement over existing models. Fan et al. [23] proposed a spatio-temporal graph attention network that captures spatial and temporal dependencies among IIoT sensors, outperforming baseline models. Ammar and Ishak [24] combined an Elman RNN with a meta-heuristic pelican optimisation algorithm to tune parameters for zero-day anomaly detection, achieving accuracy above 99%.

B.3. Hybrid/Ensembles

Hybrid and ensemble-based solutions combine techniques such as deep learning models and optimisation methods to improve detection accuracy and adaptability in IoT and IIoT environments. Konatham et al. [25] combined deep learning algorithms, gated recurrent units (GRU) and CNN for anomaly detection, achieving 95.94% accuracy in IIoT. Jablaoui et al. [26] also used a GRU-CNN combination, additionally incorporating differential privacy and SHAP-based explainable AI; the solution was benchmarked on three datasets with accuracy above 98% and a false alarm rate of 98.6% for CIC-IoT-2023, 95.8% on ToN-IoT and 93.4% on Edge-IIoTset. Deep learning and blockchain were integrated by Sathaybama and Katiravan [27] to enhance anomaly detection while maintaining data integrity, resulting in a high detection rate with fewer false positives. In order to detect zero-day attacks in SDN-IoT and enable automated detection and mitigation, Alotaibi [28] used GNN, transformers, and reinforcement learning. To achieve excellent accuracy across several datasets, Fraihat et al. [29] integrated an LSTM-transformer with Cuckoo search feature optimisation and federated learning for scalability and privacy preservation. Ranpara et al. [30] provided context-aware, real-time threat detection for IoT networks by combining semantic reasoning with knowledge graphs. Molose and Bassey [36] reviewed 59 studies on mulitlayer IoT/IIoT security using SDN, blockchain and machine learning identifying dataset dependence, computational overhead and limited adaptability to emeerging attacks as continuing challenges. More specifically, recent zero-day investigations have employed unsupervised learning, isolation forest, attention-based models and zero-shot learning. Üstebay [39] combined supervised tree models with isolation forest and evaluated unseen attacks through attack-class exclusion across three IoT datasets. Jain et al. [15] employed unsupervised learning and autoencoder-based clustering while Nitrat et al. [40] integrated a residual vision transformer with zero-shot learning for unseen IoT malware detection. Nevertheless, incosistent zero-day evaluation protocols and limited cross-dataset validation remain unresolved.

Research methods can be supplemented by tables, graphs (pictures), and/or charts. The table does not contain vertical (upright) lines. Horizontal (flat) lines in the table are only found at the beginning and end of the table. Example of table format: Recent research shows a clear evolution in anomaly detection for IoT, IIoT, and SDN-IIoT networks, as summarised by theme in Table 1. The selected works are not exhaustive, but they reflect relevant contributions across major approaches and research fields, laying the groundwork for this work's perspective.

Early efforts relied heavily on machine learning techniques to detect abnormal attack patterns. This led to the development of deep learning systems that can capture complicated behaviours and detect both known and zero-day assaults. Hybrid and ensemble solutions, which incorporate several models and technologies, have recently been developed to improve detection accuracy, adaptability, and scalability. Despite these advancements, there are still some limits. Machine learning-based techniques continue to struggle to generalise to previously unknown attacks, while deep learning models sometimes have high computational costs, limiting adoption in resource-constrained IIoT scenarios. Hybrid methods have shown promising results, although the majority rely on static ensemble structures or traditional voting schemes, with little emphasis on adaptive detector selection for zero-day attack detection. Many prior studies similarly claim good performance under closed-set evaluation, but with insufficient proof of effectiveness against previously discovered threats in true IIoT systems.

Table 1. Summary of related SDN-based anomaly detection studies

Paper		Domain			Category			Type of attack	
		IoT	IIoT	SDN-IIoT	ML	DL	Hybrid	KA	Zd-A
2024	[17]			✓	✓			✓	
	[19]		✓		✓			✓	
	[23]		✓			✓		✓	✓
	[21]		✓			✓		✓	
2025	[18]	✓			✓			✓	
	[15]	✓			✓				✓
	[20]			✓	✓			✓	
	[22]	✓				✓		✓	✓
	[26]	✓					✓	✓	
	[27]	✓					✓	✓	✓
	[30]	✓					✓	✓	✓
	[31]			✓			✓	✓	✓
2026	[24]		✓			✓			✓
	[29]		✓				✓	✓	

IoT = internet of things; IIoT =Industrial internet of things; SDN-IIoT = Software-Defined Industrial internet of things; ML = Machine learning; DL = Deep learning; KA = known attacks; Zd-A = Zero-day

C. Results Methodology

This section introduces the conceptual framework that guides the proposed approach and describes the experimental methods and setup used to implement and evaluate its performance.

C.1. Conceptual Framework Design

A hierarchical anomaly detection framework is proposed by integrating the best-performing detection techniques into a unified architecture, described through Algorithm 1-3 (Table 2-4) and evaluated using quantitative performance benchmarks. Algorithm 1 in Table 2 preprocesses the input dataset and selects the most informative features using the LightGBM-SHAP-IBSWO feature selection method. Table 3 (Algorithm 2) then applies Random Forest traffic filtering to separate normal from suspicious traffic, while Table 4 (Algorithm 3) applies a

confidence-based routing mechanism to classify suspicious samples as either known or unseen (zero-day) attacks. The IBSWO is adopted for its feature selection efficiency from Shtayat et al. [32] however, it is further enhanced using a machine learning model and explainable AI for a better selection efficiency.

Table 2. Data preprocessing and feature selection (Phase 1)

Algorithm 1: SHAP-Guided metaheuristic feature selection

Input: Training dataset ($D_{training}$), original feature set (F), IBSWO population size (P), maximum iterations (T), minimum selected features (m)
Output: Optimal feature subset (F^*)
Steps:

1. Load $D_{training}$ and D_{test}
2. Remove irrelevant attributes and clean missing values
3. Assign X independent features and Y target variable
4. Encode categorical attributes
5. Normalise features using *Eq. 1*
6. Train LightGBM model
7. Compute SHAP importance using *Eq. 2*
8. Normalise SHAP scores using *Eq. 3*
9. Initialise IBSWO population using normalised SHAP scores
10. ***for*** t=1 to T ***do*****:**
    for each candidate solution z_p ***do***:
        Extract selected feature subset F_p
        if $|F_p|$ < m ***then***:
            Assign maximum fitness penalty
        else
            Evaluate candidate subset
            Compute fitness using Eq. 4
        end if
    end for
    Update IBSWO population
    Preserve the best solution
11. ***end for***
12. ***return*** Optimal feature subset F^*

Equations (1)-(4) provide the feature preprocessing and selection phase used in Algorithm 1. Equation (1) uses Min-Max scaling to normalise input features and arrange them on a common scale before training. Equation (2) calculates each feature's average contribution using its absolute SHAP value, and Equation (3) translates these contributions to normalised significance weights. Equation (4) defines the IBSWO fitness function, which maximises detection performance while minimising the number of selected features, producing a compact and informative feature subset.

$$x'_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)} \quad (1)$$

$$I_j = \frac{1}{n} \sum_{i=1}^{n} |\phi_{ij}| \quad (2)$$

$$w_j = \frac{w_j}{\sum_{k=1}^{d} I_k} \quad (3)$$

$$Fitness(z_p) = \alpha(1 - F1_p) + \beta\left(\frac{|F_p|}{d}\right) \quad (4)$$

Where:

- $F1_p = classification\ F1-score\ obtained\ using\ feature\ subset\ F_p$
- $|F_p| = number\ of\ selected\ features$
- $d = total\ number\ of\ original\ feautres$
- $\alpha = 0.97$
- $\beta = 0.03$

Table 3 (Algorithm 2) uses a Random Forest binary gate as the initial detection stage to distinguish between regular and suspicious communication. After training the classifier on the chosen feature subset, each test sample is classed as normal or suspicious. Normal samples are allocated their final labels directly, while suspect samples are routed to the confidence-based routing stage. Eq. (5) formalises this binary decision by mapping the Random Forest prediction to either the normal or attack classes.

Table 3. Traffic filtering (Phase 2)

Algorithm 2: Random Forest Binary Gate

Input: Selected training features (X^*_{train}), Binary labels (y_{train}), Selected testing features (X^*_{test}), Number of trees *N*

Output: Binary prediction y_{gate}, Suspicious sample set X_{sus}

Steps:

1. Train Random Forest classifier RF using X^*_{train} and y_{train}
2. ***for*** each sample *sample* $x_i \in X^*_{test}$ ***do***:
   Predict binary class:
   $y_{gate_i} \leftarrow RF(x_i)$
   if y_{gate_i} = Normal ***then***:
   Assign maximum x_i as Normal
   else
   Add x_i to suspicious sample set X_{sus}
   end if
3. ***end for***
4. ***return*** y_{gate} and X_{sus}

Table 4 (Algorithm 3) integrates three unsupervised learning approaches into a confidence-based routing architecture. Each model uses an optimised confidence threshold to evaluate if a suspicious sample can be safely classed as a known attack, an unknown assault, or one that requires additional investigation by the next detector. Eq. (6) through Eq. (11) define the confidence measures and routing decisions: Eq. (6) provides estimates of OpenSet confidence, and Eq. (7) uses predefined criteria to identify known attacks, unknown attacks, and uncertain samples. For uncertain samples, Eq. (8) computes the contrastive distance to the nearest known attack prototype, and Eq. (9) determines whether the sample belongs to a known or unknown class. Eq. (10) measures the confidence of this decision relative to the contrastive threshold, and Eq. (11) applies the LOF anomaly detector to resolve any remaining uncertain samples.

Table 4. Attack recognition ensemble (Phase 3)

Algorithm 3: Confidence-based routing detection

Input: Suspicious sample set X_{sus}, Trained OpenSet ExtraTrees model M_{ET}, Trained LOF detector M_{LOF}, Trained Contrastive encoder M_{CON}, LOF threshold τ_{LOF}, Contrastive threshold τ_{CON}, High OpenSet confidence threshold τ_{high}, Low OpenSet confidence threshold τ_{low}

Output: Final attack label $y_i \in (Know\ Attack, Unkown\ Attack)$

Steps:

1. ***for*** each suspicious $sample\ x_i \in X_{sus}$ ***do*:**

   Compute OpenSet class probabilities:

   $P_{open} \leftarrow M_{ET}(x_i)$

   Determine maximum OpenSet confidence:

   $C_{open} \leftarrow \max(P_{open})$

   if $C_{open} \geq \tau_{high}$ ***then***:

   $y_i \leftarrow$known Attack

   else if $C_{open} \leq \tau_{low}$ ***then***:

   $y_i \leftarrow$ Unknown Attack

   else

   Generate contrastive embedding:

   $z_i \leftarrow M_{CON}(x_i)$

   Compute contrastive distance:

   $d_{CON} \leftarrow Distance(z_i, nearest\ class\ prototype)$

   if d_{CON} is confidently above or below τ_{CON} ***then:***

   if $d_{CON} > \tau_{CON}$ ***then:***

   $y_i \leftarrow$ Unknown Attack

   else

   $y_i \leftarrow$ known Attack

   end if

   else

   Compute LOF anomaly score:

   $s_{LOF} \leftarrow M_{LOF}(x_i)$

   if $s_{LOF} > \tau_{LOF}$ ***then:***

   $y_i \leftarrow$ Unknown Attack

   else

   $y_i \leftarrow$ known Attack

   end if

   end if

   end if

2. ***end for***
3. ***return*** y_i

$$C_{open}(x_i) = \max_{k \in K} P(y = k | x_i) \quad (6)$$

Where:

- K *is the set of known attack classes*
- $P(y = k|x_i)$ *is the probability assigned to known classes* k.

$$R_{open}(x_i) = \begin{cases} Known,\ C_{open}(x_i) \geq \tau_{high} \\ Unknown,\ C_{open}(x_i) \leq \tau_{low} \\ Uncertain,\ \tau_{low} < C_{open}(x_i) \leq \tau_{high} \end{cases} \quad (7)$$

$$d_{con}(x_i) = 1 - \max_{k \in K} P(z_i \cdot \mu_k) \quad (8)$$

Where:

- z_i *is the normalized embedding of sample* x_i
- μ_k *is the prototype embedding of known attack class* k
- $z_i \cdot \mu_k$ *is cosine similarity*

$$R_{con}(x_i) = \begin{cases} Unknown,\ d_{con}(x_i) > \tau_{con} \\ Known,\ d_{con}(x_i) \leq \tau_{con} \end{cases} \quad (9)$$

$$M_{con}(x_i) = \frac{|d_{con}(x_i) - \tau_{con}|}{|\tau_{con}| + \epsilon} \quad (10)$$

$$R_{lof}(x_i) = \begin{cases} Unknown,\ s_{lof}(x_i) > \tau_{lof} \\ Known,\ s_{lof}(x_i) \leq \tau_{lof} \end{cases} \quad (11)$$

Where:

- $s_{lof}(x_i)$ *is the LOF anomaly score*
- τ_{lof} *is the tuned LOF threshold*

This study proposes a unified anomaly detection framework for identifying both known and zero-day attacks, addressing the challenge of combining supervised and unsupervised learning. As shown in Figure 1, the framework has three phases. The first phase preprocesses the dataset and performs SHAP-guided IBSWO feature selection to obtain a compact set of informative features. The second phase applies Random Forest traffic filtering to separate normal from suspicious traffic. The random forest routing confidence threshold was fixed to 0.80 to ensure theat only high-confidence predictions were accepted directly, while samples with confidence below 0.80 were routed for further analysis.

The third phase analyses suspicious samples using a confidence-based routing mechanism across three unsupervised detectors. Detector-specific thresholds were calculated exclusively from normal validation data and set at 95th percentile of each detectors validation scores which corresponds to a target false-positive rate of 5%. Because the detectors produce scores on different scales, the deviation from each threshold was normalized using interquartile range of their corresponding validation scores. The detector with the highest normalized threshold deviation was selected to produce the final classification and its ability to generalize to unseen attack types was therefore evaluated using held-out attacks strategy.

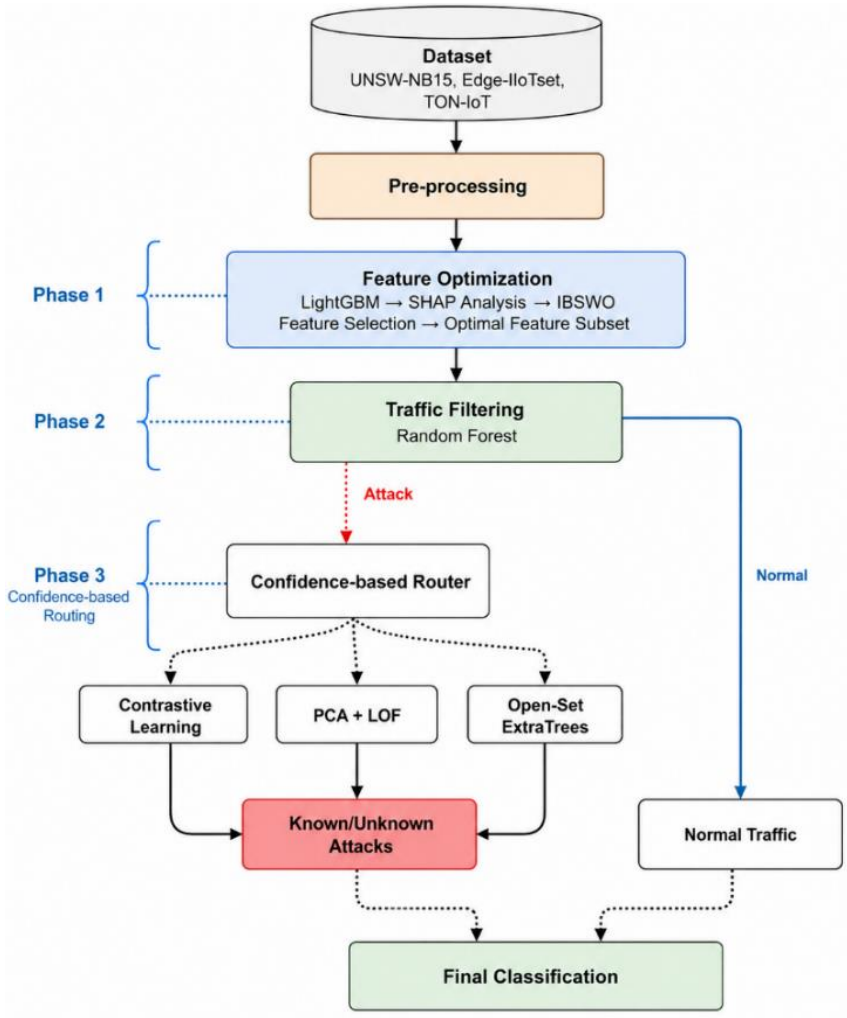

Figure 1. Proposed hybrid anomaly detection technique

C.2. Experimental Setup and Evaluation

This section introduces the conceptual framework that guides the proposed approach and describes the experimental methods used to implement and evaluate its performance.

a. *Experimental Setup*

The experiments, summarised in Table 5, comprise two stages: an empirical comparison of baseline machine learning models, and an evaluation of the proposed hierarchical anomaly detection framework. All experiments were run on an MSI Cyborg 15 A12VF with Windows 11 (64-bit), a 12th Gen Intel Core i7-12650H processor (2.30 GHz), 477 GB of storage, 16 GB RAM, and an NVIDIA GeForce RTX 4060 GPU (8 GB). The baseline model comparison used Python on Google Colab with Pandas, NumPy, Matplotlib and Scikit-learn. Five supervised and three unsupervised machine learning algorithms were evaluated on the Edge-IoTset dataset. Execution times ranged from 0 to 15 minutes depending on model complexity; performance was measured using accuracy, precision, recall, F1-score, ROC-AUC, and recall at a 1% false positive rate (FPR).

The proposed framework was evaluated on UNSW-NB15, Edge-IIoTset and TON-IoT using a held-out attack strategy to assess zero-day detection

performance. Data were preprocessed using Min-Max normalisation and the LightGBM-SHAP-IBSWO feature selection method. The framework combines Random Forest (the best-performing supervised model) for traffic filtering with a combination of unsupervised models: PCA-enhanced LOF (the best-performing unsupervised model), OpenSet ExtraTrees, and a contrastive learning detector. This stage was implemented in PowerShell on VS Code using Pandas, NumPy, Matplotlib, Scikit-learn, PyTorch and SciPy. Experiments were completed within 0-5 minutes and were evaluated using Macro-F1, Novelty-F1, and correct detection rates for known and unknown attacks.

To support repeatability, a fixed random seed of 42 was applied to the experiments. The training and testing ratio used were divided using a stratified 70:30 split. Preprocessing and feature selection were fitted only on the training data. RF and ExtraTrees used both normal and attack samples whereas PCA-LOF and Contrastive were trained on normal samples only. From these normal samples, 20% were reserved for threshold calibration at a 5% target FPR. The contrastive model was trained for up to 25 epochs with a batch size of 256, a learning rate of 0.001 and an early-stopping patience of five epochs.

Table 5. Evaluation parameters

Domains	Value
Dataset	UNSW-NB15, Edge-IIoTset, TON-IoT
ML models	Supervised learning models(LR, SVM, KNN, RF, NN(MLP)); Unsupervised learning models (LOF, GMN, Isolation Forest)
Evaluation strategy	Held-out attack categories
Data processing	Min-Max Normalisation
Feature selection	Ranking method: LightGBM + SHAP XAI
	Optimisation: IBSWO
	Fitness weights: $\alpha = 0.97, \beta = 0.03$
Known Detection	Random Forest
Zero-day Detection	Open-set classifier: ExtraTrees
	Density-based classifier: PCA+LOF
	Representation-Based detector: Contrastive learning
Performance metrics	Macro-F1, Novelty-F1, Recall (known and unknown correct detection rate), Accuracy, Precision, ROC-AUC, PR-AUC, Recall (1%) FPR
Simulation	Python programming language - Google Colab, VS Code, PowerShell, Pandas, NumPy, Matplotlib, Scikit-learn, PyTorch, Scipy
Simulation time	0-5 Minutes, 0-15 minutes

b. *Benchmark Dataset*

Various datasets summarised in Table 6 capture different types of network attacks; this study focuses on datasets representing SDN-enabled Industrial IoT and IIoT environments, supporting evaluation of both baseline models and the proposed framework. The comparative analysis uses Edge-IIoTset [33] for its realistic, diverse IoT/IIoT scenarios. To assess the framework's ability to detect unseen attacks, additional experiments used UNSW-NB15 [34] and TON-IoT [35] under a held-out attack evaluation strategy. Together, these datasets provide a

benchmark for evaluating known and unknown attack detection across different network environments [33-35].

Table 6. Summary of evaluation datasets

Dataset	Environment focus	Data source	Attack coverage	Strengths	Limitations
UNSW-NB15 [34]	Modern network intrusion detection	Network traffic with PCAP-derived features	Nine attack families	Evaluates contemporary network attacks using forty-nine features	Not tailored for IoT/IIoT telemetry or federated learning
TON-IoT [35]	IoT/IIoT intrusion detection	Telemetry data, OS logs, and network traffic	Nine attack types (DoS, DDoS, ransomware, backdoor, injection, XSS, password cracking, MITM)	Captures heterogeneous IoT/IIoT data across Edge, Fog, and Cloud; suitable for data-driven IDS evaluation	Less aligned with FL compared to Edge-IIoTset
Edge-IIoTset [33]	IoT/IIoT, edge, SDN, blockchain, FL	Network traffic, alerts, logs, system resources, IoT/IIoT device data	Fourteen attacks grouped into five categories (DoS/DDoS, information gathering, MITM, injection, malware)	Supports centralised ML and decentralised FL within a realistic seven-layer IoT/IIoT architecture	Complex preprocessing due to heterogeneity and high-dimensional features

D. Experimental Results

This section presents the experimental evaluation of the baseline machine learning models and the proposed intrusion detection framework for SDN-IIoT networks. First, supervised and unsupervised algorithms are compared to determine which detectors are more effective for known and unknown assaults. The chosen models are then integrated into the proposed framework and tested against a variety of benchmark datasets for detecting known and zero-day attacks.

D.1. Classical learner'NID comparison

Anomaly detection models from various machine learning families are contrasted in Figure 2. Random Forest (RF), with an accuracy of 0.964, recall of 0.958, F1-score of 0.980, and ROC-AUC of 0.991, outperforms KNN. Its high precision results in fewer false alarms, making it dependable for identifying serious threats. Given its performance, efficiency and stability without overfitting, RF was selected as the best model for the stacking ensemble. Figure 2-B compares three unsupervised learning models across six metrics for detecting unknown attacks. LOF outperforms Isolation Forest and GMM across all six metrics. GMM and Isolation Forest achieved perfect recall (1.0) but a Recall at 1% FPR of 0, meaning neither could separate attacks from normal traffic once false positives were tightly constrained. LOF, by contrast, delivered near-perfect results across all six metrics while maintaining a recall of 1.0 at a very low FPR, indicating it separates attack patterns from normal behaviour more effectively in this dataset.

LOF was therefore selected as the zero-day anomaly detector for the stacking ensemble.

D.2. Evaluation 1: Held-out Attack Evaluation

This evaluation is presented through three quantitative analyses of the proposed framework's performance.

a. *Cross-dataset performance*

As shown in Table 7 and Figure 3, the proposed hybrid anomaly detector performed consistently across all three benchmark datasets. Macro-F1 improved from 0.73 on UNSW-NB15 to 0.87 on TON-IoT, while Novelty-F1 remained stable (0.74-0.76). The known attack detection rate increased while the unknown attack detection rate decreased slightly, reflecting the trade-off between known-attack classification and novel-attack detection. The framework generalised across datasets while maintaining zero-day detection performance. Table 7 shows that each unsupervised detector contributes differently across datasets. The OpenSet detector was the most consistent, with Novelty-F1 scores between 0.75 and 0.79. PCA-enhanced LOF achieved the highest Novelty-F1 on Edge-IIoTset (0.81) but varied considerably on UNSW-NB15 and TON-IoT. The contrastive detector stayed stable across all three datasets. Combining the three complementary detectors gives the framework more consistent zero-day detection than any single detector alone.

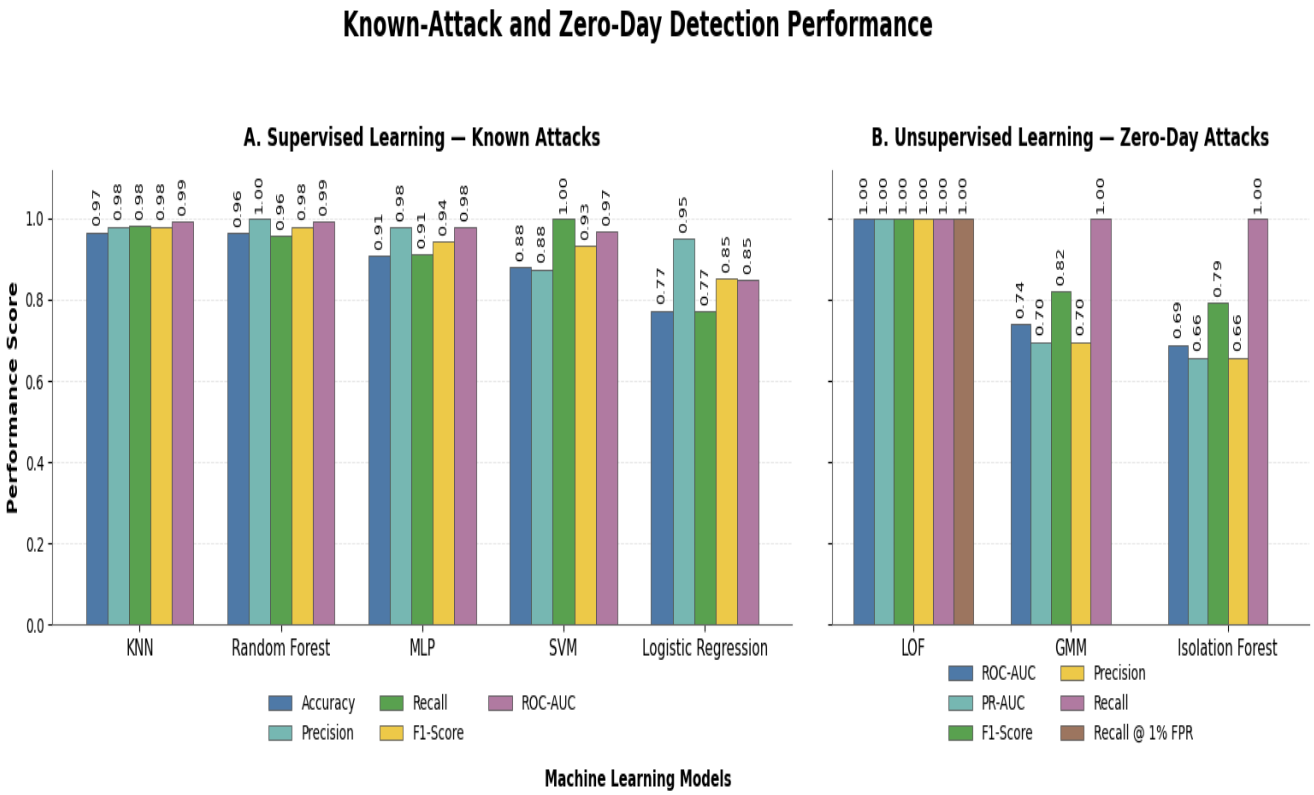

Figure 2. Comparison of machine learning models' results for known and unknown attacks

Table 7. Framework Performance across three datasets

Dataset	Overall framework performance				Confidence-based routing performance		
	Three_Class_Macro _ F1	Novelty _ F1	Known Attack Correct Rate	Unknown Attack Detection Rate	OpenSet (Novelty F1)	LOF (Novelty F1)	Contrastive (Novelty F1)
UNSW-NB15	0.73	0.75	0.57	0.86	0.79	0.08	0.66
EDGE-IIoT	0.82	0.74	0.69	0.78	0.75	0.81	0.73
TON-IoT	0.87	0.76	0.95	0.71	0.79	0.52	0.67

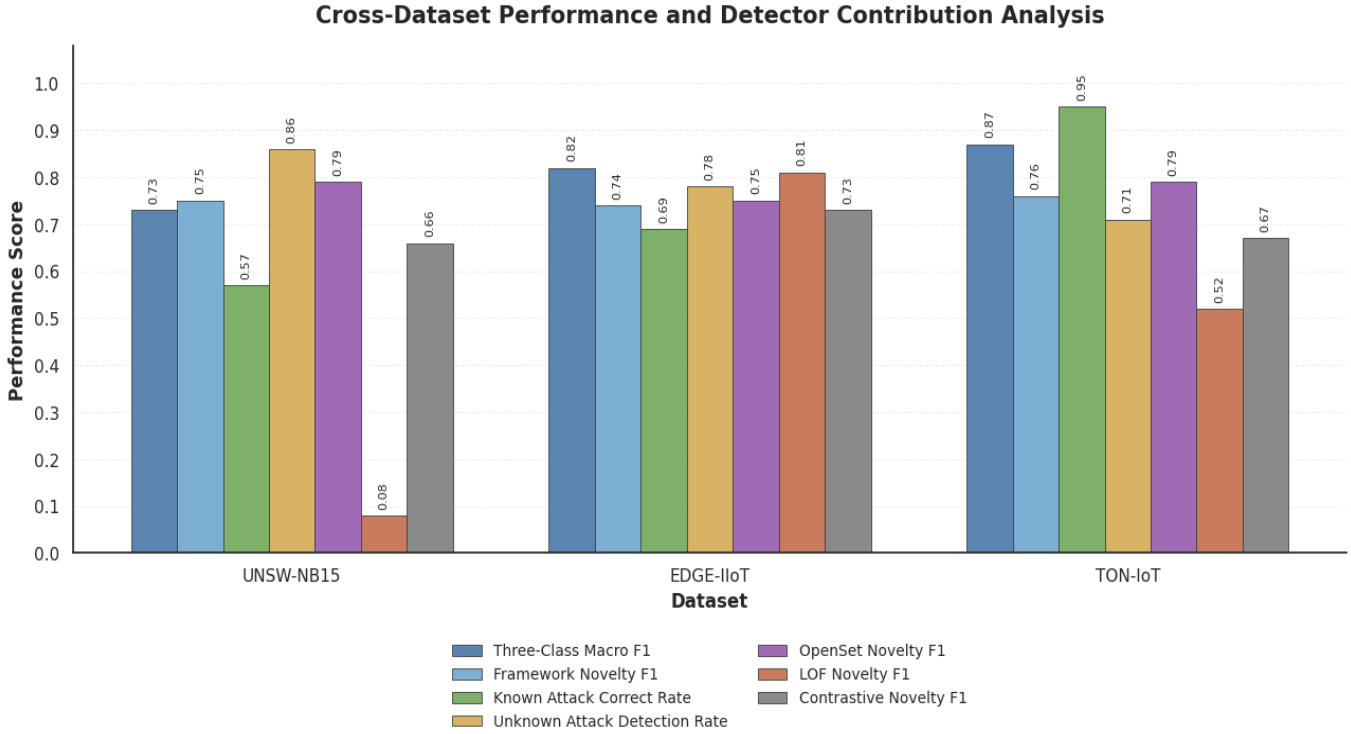

Figure 3. Overall framework performance across three datasets

b. *Attack-level Analysis*

Fig. 4 shows attack-level detection performance of the framework across the three datasets which ranged from 0.70 to 1.00, with several attacks detected perfectly. Detection rates for UNSW-NB15 Worms, shellcode and backdoor attacks may be attributed to their stealthy behaviou, limited traffic footprints and similarity to legitatimate system activity. Edg-IioTset MITM traffic also produced moderate scores becase of intercepted or modified communication may retain flow characteristics similar to normal exchanges. The weekest result is observed in Edge-IIoTset inject attack which recorded lower scores including of 0.00. Although its known-attack rate was 0.90, its unknown-attack detection rate was 1.00, its unkown-attack detection rate was 0.00, indicating that the framework failed to recognize this attack when treated as an unseen attack. The reason behind this lies in the nature of the injection attacks which manipulates application-layer inputs or

payload content while retaining traffic-flow patterns similar to legitimate communication. Consequently, the selected network-flow features may not contain sufficient payload or seemantic information to distinguish an unseen injection attack features from normal traffic. For the remaining attacks, performance varied across the metrics with several TON-IoT attacks achieving scores of 1.00. These findings demonstrate strong overall performance while highlighting the need for payload-aware or application-layer features to improve the detection of previously unseen injection attacks.

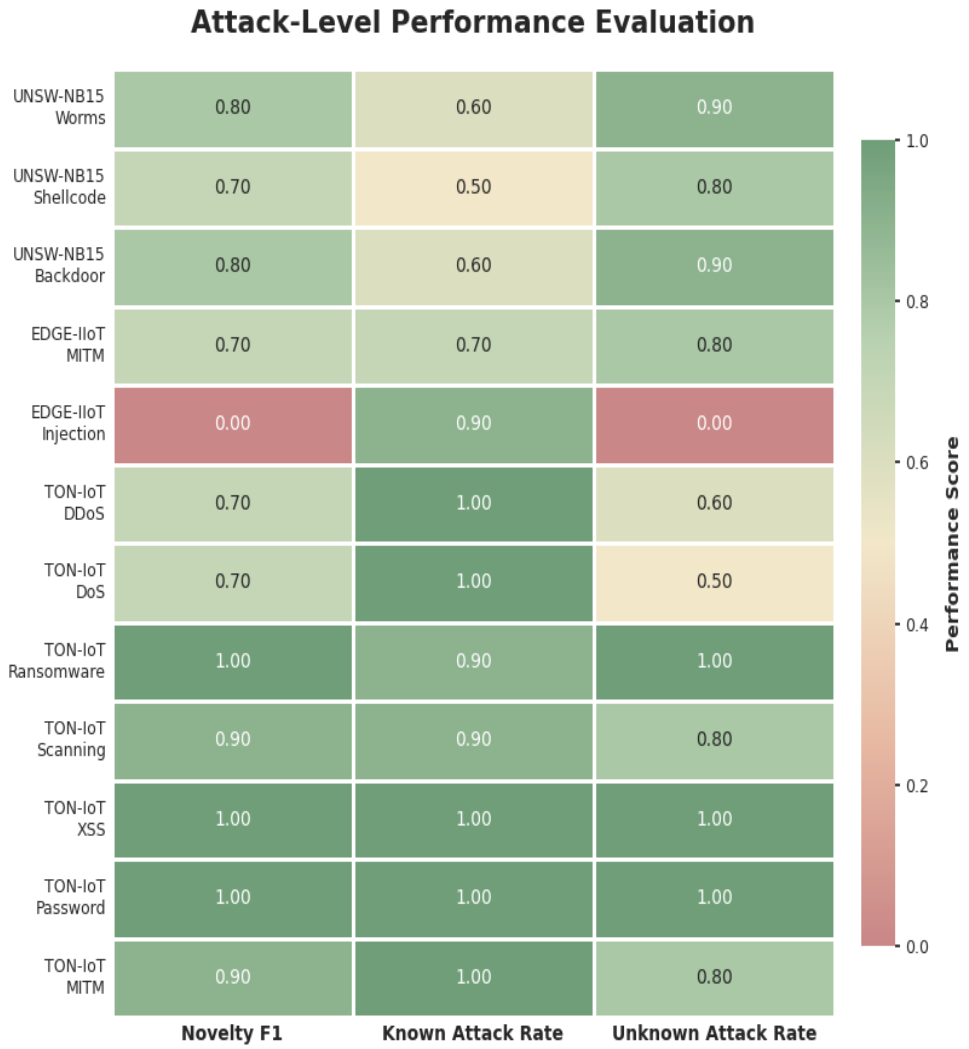

Figure 4. Detection rate performance for held-out attacks in each dataset

D.3. Evaluation 2: Outlier Detection Evaluation

This experiment extends the held-out evaluation by modelling normal network behaviour and identifying malicious traffic as unseen anomalies.

a. *Density-based NIDS*

The proposed detector was evaluated under outlier detection to test its zero-day detectors' ability to identify previously unseen malicious traffic, shown in Fig.

5: A-C (UNSW-NB15, TON-IoT and Edge-IIoTset). These results improved considerably over the stricter held-out evaluation, showing performance under both the stricter and more liberal evaluation settings.

b. *One-class NIDS*

The proposed hybrid anomaly detector was also adapted into a one-class variant that retains LOF but replaces the confidence- and representation-based detectors with Autoencoders and Deep SVDD to identify malicious deviations from normal network behaviour. Fig. 6 show the one-class framework's performance on UNSW-NB15, Edge-IIoTset and TON-IoT under outlier detection evaluation.

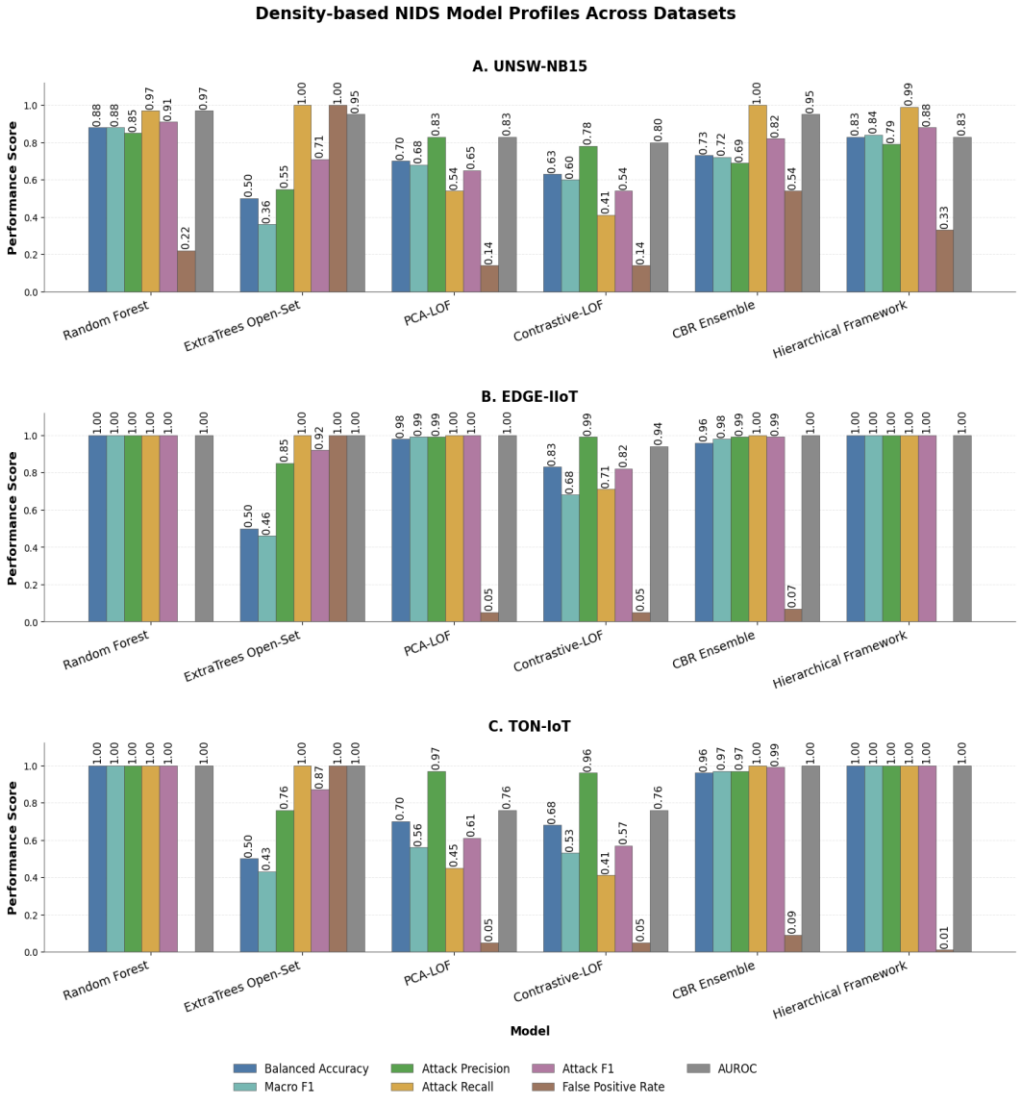

Figure 5. Density-based anomaly detector performance on using Outlier Detection Evaluation

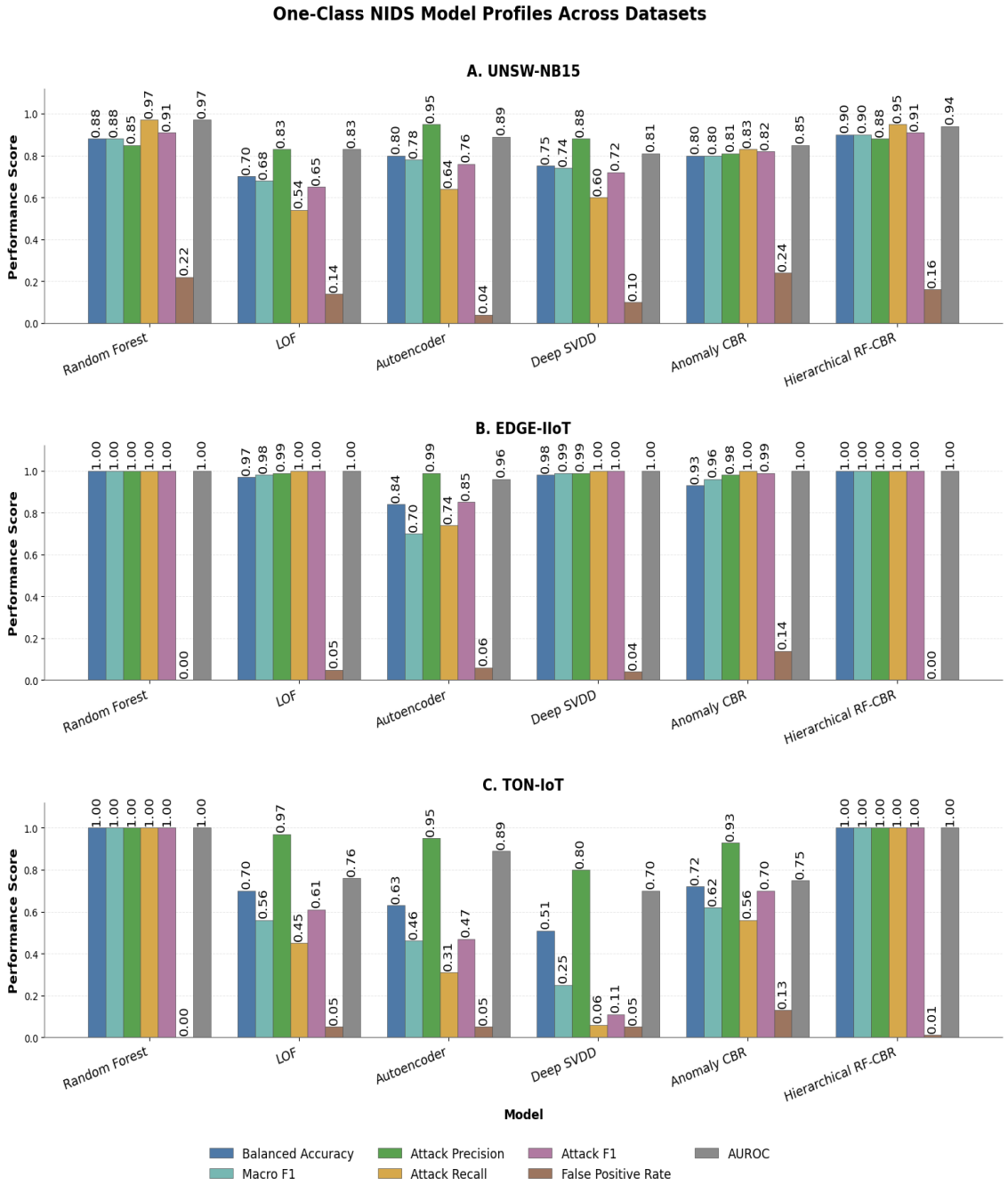

Figure 6. One-Class anomaly detector performance on using Outlier Detection Evaluation evaluation

E. Discussion

This section discusses the experimental findings from two perspectives: framework performance based on the conducted experiments, and comparison against recent zero-day intrusion detection approaches.

E.1. Proposed NIDS findings

The findings presented in this paper showed that no single learning paradigm proved sufficient for reliable intrusion detection in SDN-IIoT environments. Supervised models achieved high accuracy on previously observed attacks but showed limited ability to identify novel threats. Unsupervised models were better at identifying zero-day attacks, but they were less accurate at classifying known attacks. This trade-off motivated the proposed confidence-based hierarchical framework. Rather than relying on a single detector or conventional ensemble voting, the framework uses a staged process: a Random Forest filter removes normal traffic before complementary novelty detectors analyse suspicious samples, each operating where it performs best. Across UNSW-NB15, Edge-IIoTset and TON-IoT, the framework achieved strong Macro-F1 scores, correctly identified several complex and stealthy attack types. It generalised across the three datasets, offering a practical balance between known-attack classification, zero-day detection, and a plausible baseline for an adversarial-aware hierarchical federated learning approach which will is proposed in [36].

Modelling normal network behaviour under the outlier detection strategy proved more effective for zero-day detection than the stricter held-out evaluation: both proposed variants reached accuracy and Macro-F1 of at least 0.90, with an attack F1-score of 0.91 on UNSW-NB15 and exceeded 0.98 on Edge-IIoTset and TON-IoT with low false-positive rates. This confirms that combining complementary novelty detectors improves detection of unseen malicious traffic across IoT/IIoT environments.

E.2. Empirical Analysis of recent NIDS framework

Recent zero-day detection studies summarised in Table 8 generally improve one part of the detection pipeline. Yang et al. [37] strengthened adaptability through VAE-based open-set recognition with dynamic model updating; Lian and Guerra-Manzanares [38] used incremental learning to expand knowledge of emerging attacks. Krishnan et al. applied attention-based representation learning, combining multi-head self-attention, BiLSTM and explainable AI to interpret zero-day decisions. The proposed framework instead integrates SHAP-IBSWO feature optimisation, hierarchical traffic filtering and confidence-based detector routing, letting open-set, density-based and representation-based detectors jointly identify unseen attacks. Unlike Lian and Guerra-Manzanares [38] and Krishnan et al. [14], which were each validated on a single benchmark, the proposed framework was validated across UNSW-NB15, Edge-IIoTset and TON-IoT using attack-class holdout and novelty-based evaluation, giving a broader assessment of zero-day detection capability across datasets.

Under attack-class holdout evaluation (Table 9, the proposed framework reached unknown-attack recall of 0.86 and 0.71 and stable novelty-F1 scores of 0.74-0.76 across the three datasets. Lian and Guerra-Manzanares [38] reported a higher unknown recall (0.882 ± 0.80) and Yang et al. [37] reported higher peak unknown F1-scores (0.775-0.923) on selected datasets, but under different experimental settings and dataset configurations. The proposed framework instead held its performance steady across three heterogeneous datasets rather than peaking on one.

Moreover, the novelty-based evaluation (Table 10) shows a clearer advantage. The MI^{2}DAS-GMM detector [38] reached perfect attack recall at an FPR of 0.095; the proposed PCA+LOF and Deep SVDD detectors matched that recall while cutting false alarms to 0.05 and 0.04. The full hierarchical detector removed false positives entirely (FPR = 0.00) while keeping 1.00 attack recall and Macro-F1, showing that confidence-based routing lets each detector work where it performs best rather than relying on one detection paradigm. Rather than optimising a single detection mechanism, the proposed framework combines complementary zero-day detection strategies within one hierarchical architecture. Cross-dataset validation under two evaluation strategies shows the framework matches recent approaches in zero-day detection while holding up more evenly across IoT and IIoT environments.

Table 8. Comparative analysis of recent Zero-day IDS architectures

Author	Method	Benchmark dataset	Zero-day protocol	Adaptive mechanism
Krishnan et al. [14]	Explainable attention-fusion model: multi-head self-attention (short-term), BiLSTM (long-term), dense layers (classification). Interpretability via LIME/SHAP. Unknown attacks handled with centroid-based embedding proximity.	Edge-IIoTset	Held-out attack grouping	Attention fusion/embedding
Yang et al. [37]	Three-stage lightweight dynamic open-set IDS: (i) known attack classification, (ii) unknown attack recognition via reconstruction error distribution, (iii) dynamic update detection. Combines VAE + EVT with lightweight model updating.	CICIDS2017, NSL-KDD, Gas Pipeline, Water Storage	Open-set recognition	Dynamic updating
Lian and Guerra-Manzanares [38]	Multi-layer IIoT IDS: Layer 1 anomaly detection (normal vs attack); Layer 2 separation of known vs unknown attacks and classification of known attacks. Models include GMM, LOF, OC-SVM, RF, self-training, label spreading/propagation, active learning.	Edge-IIoTset	Open-set recognition	Incremental learning
Proposed	SHAP-IBSWO feature selection; RF binary gate for normal vs malicious traffic; supervised/open-set classification for known attacks; suspicious traffic routed to zero-day ensemble detectors (open-set ExtraTrees, PCA+LOF, contrastive representation-based detection).	Edge-IIoTset, UNSW-NB15, TON-IoT	Attack-class holdout, outlier detection	Confidence-based detector routing

Table 9. Performance comparison under Attack-Class Holdout (Open-set) Evaluation

Study	Unknown Recall	Unknown/Novelty F1-score
Krishnan et al. [14]	0.711	0.744
Yang et al. [37]	0.42 – 0.82	0.775 – 0.923
Lian and Guerra-Manzanares [38]	0.882 ± 0.80	-
Proposed framework (UNWS-NB15)	0.86	0.75
Proposed framework (Edge-IIoTset)	0.78	0.74
Proposed framework (TON-IoT)	0.71	0.76

Table 10. Comparative Performance under novelty-based Zero-day Detection

Study	Dataset	Training setting	Attack recall	FPR	Main reported results
MI^{2}DAS-GMM [38]	Edge-IIoTset	Normal-Only novelty detection	1.00	0.095	Accuracy 0.953
Proposed LOF (PCA)	Edge-IIoTset	Normal-Only novelty branch	1.00	0.05	Attack F1 0.98 -1.00
Proposed Deep SVDD	Edge-IIoTset	Normal-Only novelty branch	1.00	0.04	Attack F1 1.00
Proposed hierarchical density-based NIDS	Edge-IIoTset	Hybrid supervised and novelty stages	1.00	0.00	Macro-F1 1.00
Proposed hierarchical one-class NIDS	Edge-IIoTset	Hybrid supervised and one-class stages	1.00	0.00	Macro-F1 1.00

F. Conclusion and Future work

This paper presented a hierarchical anomaly detection framework for SDN-IIoT networks designed to address a persistent limitation in existing NID research: the tendency to evaluate detection performance under closed-set conditions or to treat known-attack classification and zero-day detection as separate problems. By combining SHAP-guided IBSWO feature selection, a Random Forest traffic-filtering gate, and a confidence-based routing mechanism across three complementary unsupervised detectors, the proposed framework allows each component to operate where it is most effective rather than relying on a single model or conventional ensemble voting. Across three heterogeneous benchmark datasets, the framework demonstrated consistent Macro-F1 and Novelty-F1 performance under attack-class holdout evaluation, and near-perfect detection with substantially reduced false-alarm rates under outlier-based novelty evaluation, while remaining computationally lightweight enough for near real-time deployment. The detector contribution analysis further showed that no individual unsupervised model performed reliably across all datasets, reinforcing the value of the routing mechanism over any single detector in isolation. Comparison against recent zero-day IDS approaches indicated that the proposed framework performs comparably to recent approaches in unknown-attack detection, while holding up more evenly across heterogeneous IoT/IIoT environments than several of the compared approaches, which were validated on a single benchmark dataset.

Despite these results, several limitations remain. First, the framework's routing thresholds were tuned empirically rather than through a formalised optimisation procedure, and their sensitivity to dataset distribution shifts has not been systematically characterised. Second, while the framework accurately classifies known and unknown traffic, it does not yet incorporate an active mitigation component; classification output is not translated into concrete response actions such as traffic isolation, policy enforcement, or SDN controller reconfiguration. Third, evaluation was conducted on static, pre-collected benchmark datasets, and the framework's behaviour under streaming, concept-drifting, or adversarially manipulated traffic has not been assessed. Future work will address these limitations in three directions. First, an adaptive mitigation layer will be developed to translate routing decisions into automated SDN-level response actions, directly linking detection to network defence. Second, the confidence thresholds governing the routing mechanism will be optimised jointly with the IBSWO feature selection stage, rather than tuned independently, to improve consistency across datasets with differing attack distributions. Third, the framework will be extended and evaluated under streaming and continual-learning conditions to assess its resilience to concept drift and its viability as a baseline detector within a broader adversarial-aware, hierarchical federated learning architecture for SDN-IIoT environments.

G. Acknowledgment

This research was supported by the Unit for Data Science and Computing (UDSC) and the Department of Computer Science at the North-West University Mafikeng campus, South Africa.

H. References

[1] O. Peter, A. Pradhan, and C. Mbohwa, "Industrial internet of things (IIoT): opportunities, challenges, and requirements in manufacturing businesses in emerging economies," *Procedia Computer Science,* vol. 217, pp. 856-865, 2023.

[2] O. A. Okunlola, J. Olaoye, O. O. Samuel, A. O. Okunlola, and O. Alao, "Cybersecurity Strategies for Integrating Industrial IoT and Edge Computing: Challenges, Risks, and Future Perspectives," *Int. J. Future Eng. Innov,* vol. 2, pp. 87-95, 2025.

[3] A. M. Alnajim, S. Habib, M. Islam, S. M. Thwin, and F. Alotaibi, "A comprehensive survey of cybersecurity threats, attacks, and effective countermeasures in industrial internet of things," *Technologies,* vol. 11, no. 6, p. 161, 2023.

[4] T. Zhukabayeva, L. Zholshiyeva, N. Karabayev, S. Khan, and N. Alnazzawi, "Cybersecurity solutions for industrial internet of things–edge computing integration: Challenges, threats, and future directions," *Sensors,* vol. 25, no. 1, p. 213, 2025.

[5] D. Tu, L. Liu, and B. Wang, "Optimizing Network Management through Software-Defined Networking in Computer Communications," in *2025 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2025: IEEE, pp. 1-5.

[6] M. Alam, N. Ahmed, R. Matam, M. Mukherjee, and F. A. Barbhuiya, "Sdn-based reconfigurable edge network architecture for industrial internet of things," *IEEE Internet of Things Journal,* vol. 10, no. 18, pp. 16494-16503, 2023.
[7] O. David, P. Thornley, and M. Bagheri, "Software defined networking (SDN) for campus networks, WAN, and datacenter," in *2023 International Conference on Smart Applications, Communications and Networking (SmartNets)*, 2023: IEEE, pp. 1-8.
[8] A. Rahdari *et al.*, "Security and Privacy Challenges in SDN-Enabled IoT Systems: Causes, Proposed Solutions, and Future Directions," *Computers, Materials & Continua,* vol. 80, no. 2, 2024.
[9] M. S. Farooq, S. Riaz, and A. Alvi, "Security and privacy issues in software-defined networking (SDN): A systematic literature review," *Electronics,* vol. 12, no. 14, p. 3077, 2023.
[10] J. Li, Z. Zhao, R. Li, and H. Zhang, "Ai-based two-stage intrusion detection for software defined iot networks," *IEEE internet of Things Journal,* vol. 6, no. 2, pp. 2093-2102, 2018.
[11] M. Cui, J. Chen, X. Qiu, W. Lv, H. Qin, and X. Zhang, "Multi-class intrusion detection system in SDN based on hybrid BiLSTM model," *Cluster Computing,* vol. 27, no. 7, pp. 9937–9956, 2024, doi: 10.1007/s10586-024-04477-5.
[12] F. Wahab, A. Shah, I. Khan, B. Ali, and M. Adnan, "An SDN-based Hybrid-DL-driven cognitive intrusion detection system for IoT ecosystem," *Computers and Electrical Engineering,* vol. 119, p. 109545, 2024/10/01/ 2024, doi: https://doi.org/10.1016/j.compeleceng.2024.109545.
[13] K. P. Ghosh, M. Hasan, M. T. I. Robin, M. A. Hossain, and M. S. Islam, "A novel deep learning framework with temporal attention convolutional networks for intrusion detection in IoT and IIoT networks," *Scientific Reports,* vol. 15, no. 1, p. 44624, 2025/12/13 2025, doi: 10.1038/s41598-025-32697-1.
[14] D. Krishnan, S. Singh, and V. Sugumaran, "Explainable AI for zero-day attack detection in IoT networks using attention fusion model," *Discover Internet of Things,* vol. 5, no. 1, p. 83, 2025/07/28 2025, doi: 10.1007/s43926-025-00184-8.
[15] A. Jain, R. Bagoria, and P. Arora, "An intelligent zero-day attack detection system using unsupervised machine learning for enhancing cyber security," *Knowledge-Based Systems,* vol. 324, p. 113833, 2025.
[16] R. Molose and B. Isong, "A Survey of Multi-Layer IoT Security Using SDN, Blockchain, and Machine Learning," *Electronics,* vol. 15, no. 3, p. 494, 2026.
[17] C. Kumar and M. S. A. Ansari, "An explainable nature-inspired cyber attack detection system in Software-Defined IoT applications," *Expert Systems with Applications,* vol. 250, p. 123853, 2024.
[18] A. Punia, M. Tiwari, and S. S. Verma, "A machine learning-based efficient anomaly detection system for enhanced security in compromised and maligned IoT Networks," *Results in Engineering,* vol. 26, p. 105562, 2025.
[19] M. I. H. Okfie and S. Mishra, "Anomaly detection in IIoT transactions using machine learning: A lightweight blockchain-based approach," *Engineering, Technology & Applied Science Research,* vol. 14, no. 3, pp. 14645-14653, 2024.
[20] H. Yıldız and M. Balta, "Comparison of Machine Learning Based Anomaly Detection for Energy Consumption Values in SDN-IoT Based Home Area

Networks," *Sakarya University Journal of Computer and Information Sciences,* vol. 8, no. 3, pp. 518-535, 2025.
[21] Z. Chen, Z. Li, J. Huang, S. Liu, and H. Long, "An effective method for anomaly detection in industrial Internet of Things using XGBoost and LSTM," *Scientific Reports,* vol. 14, no. 1, p. 23969, 2024/10/14 2024, doi: 10.1038/s41598-024-74822-6.
[22] H. Rhachi, Y. Balboul, and A. Bouayad, "Enhanced anomaly detection in IoT networks using deep autoencoders with feature selection techniques," *Sensors,* vol. 25, no. 10, p. 3150, 2025.
[23] Y. Fan, T. Fu, N. I. Listopad, P. Liu, S. Garg, and M. M. Hassan, "Utilizing correlation in space and time: Anomaly detection for Industrial Internet of Things (IIoT) via spatiotemporal gated graph attention network," *Alexandria Engineering Journal,* vol. 106, pp. 560-570, 2024.
[24] K. Ammar and M. K. Ishak, "Detecting Zero-Day Attacks Using Deep Learning with Pelican Optimization Algorithm in IIoT Environments," *Engineering, Technology & Applied Science Research,* vol. 16, no. 1, pp. 30703-30709, 2026.
[25] B. Konatham, T. Simra, F. Amsaad, M. I. Ibrahem, and N. Z. Jhanjhi, "A secure hybrid deep learning technique for anomaly detection in iiot edge computing," *Authorea Preprints,* 2024.
[26] R. Jablaoui, O. Cheikhrouhou, M. Hamdi, and N. Liouane, "Deep learning enabled intrusion detection system for IoT security," *EURASIP Journal on Wireless Communications and Networking,* vol. 2025, no. 1, p. 66, 2025/08/27 2025, doi: 10.1186/s13638-025-02477-6.
[27] S. A R and J. Katiravan, "Enhancing anomaly detection and prevention in Internet of Things (IoT) using deep neural networks and blockchain based cyber security," *Scientific Reports,* vol. 15, no. 1, p. 22369, 2025/07/01 2025, doi: 10.1038/s41598-025-04164-4.
[28] J. Alotaibi, "A hybrid software-defined networking approach for enhancing IoT cybersecurity with deep learning and blockchain in smart cities," *Peer-to-Peer Networking and Applications,* vol. 18, no. 3, p. 123, 2025.
[29] S. Fraihat, Q. Yaseen, Y. Sanjalawe, A. Abu-Errub, S. N. Makhadmeh, and M. A. Al-Betar, "Intrusion detection in industrial internet of things network using feature optimization and hybrid deep learning," *Discover Internet of Things,* vol. 6, no. 1, p. 34, 2026/02/12 2026, doi: 10.1007/s43926-026-00284-z.
[30] R. Ranpara, S. K. Patel, O. P. Kumar, and F. A. Al-Zahrani, "A computational framework for IoT security integrating deep learning-based semantic algorithms for real-time threat response," *Scientific Reports,* vol. 15, no. 1, p. 16794, 2025/05/14 2025, doi: 10.1038/s41598-025-93898-2.
[31] J. Alotaibi, "AI-driven intrusion detection and mitigation framework for software-defined IoT networks," *Peer-to-Peer Networking and Applications,* vol. 18, no. 6, p. 321, 2025/11/03 2025, doi: 10.1007/s12083-025-02151-0.
[32] M. M. Shtayat *et al.*, "An Improved Binary Spider Wasp Optimization Algorithm for Intrusion Detection for Industrial Internet of Things," *IEEE Open Journal of the Communications Society,* vol. 6, pp. 2926-2944, 2025, doi: 10.1109/OJCOMS.2024.3421647.
[33] M. A. Ferrag, O. Friha, D. Hamouda, L. Maglaras, and H. Janicke, "Edge-IIoTset: A new comprehensive realistic cyber security dataset of IoT and IIoT

applications for centralized and federated learning," *IEEe Access,* vol. 10, pp. 40281-40306, 2022.

[34] N. Moustafa and J. Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," in *2015 Military Communications and Information Systems Conference (MilCIS)*, 10-12 Nov. 2015 2015, pp. 1-6, doi: 10.1109/MilCIS.2015.7348942.

[35] A. Alsaedi, N. Moustafa, Z. Tari, A. Mahmood, and A. Anwar, "TON_IoT Telemetry Dataset: A New Generation Dataset of IoT and IIoT for Data-Driven Intrusion Detection Systems," *IEEE Access,* vol. 8, pp. 165130-165150, 2020, doi: 10.1109/ACCESS.2020.3022862.

[36] R. Molose, B. Isong, A. Abu-Mahfouz, and N. Dladlu, "Analysis of Privacy-Preserving Federated Learning's Resilience against Adverse Attack in Internet of Things Systems," in *2025 IEEE AFRICON*, 2025: IEEE, pp. 1-8.

[37] X. Yang, F. Tong, F. Jiang, and G. Cheng, "A lightweight and dynamic open-set intrusion detection for industrial Internet of Things," *IEEE Transactions on Information Forensics and Security,* 2025.

[38] W. Lian and A. Guerra-Manzanares, "MI^{2}DAS: A Multi-Layer Intrusion Detection Framework with Incremental Learning for Securing Industrial IoT Networks," *arXiv preprint arXiv:2602.23846,* 2026.

[39] Üstebay, Serpil. "Enhancing zero-day attack detection in IoT networks via isolation forest and ensemble tree models." *Electrica* 25 (2025): 1-8.

[40] Nitrat, Komcharn, Nopparuj Suetrong, and Natthanan Promsuk. "Zero-day attack detection in iot networks using a residual vision transformer-based approach with zero-shot learning." *IEEE Open Journal of the Communications Society* (2025).