Measurement: Digitalization 4 (2025) 100015

Contents lists available at ScienceDirect

Measurement: Digitalization

journal homepage: www.elsevier.com/locate/meadig

Full Length Article

Comparative evaluation of a novel IDS dataset for SDN-IoT using deep learning models against InSDN, BoT-IoT, and ToN-IoT

Heba Dhirar *, Ali Hamad

Information and Communication Engineering, AL-Khwarizmi College of Engineering, University of Baghdad, Baghdad, Iraq

ARTICLE INFO

Keywords:
Intrusion Detection System (IDS)
Deep Learning (DL)
Internet of Things (IoTs)
Software-defined Network (SDN)
Dataset
ToN-IoT
BoT-IoT
InSDN

ABSTRACT

The proliferation of Software-Defined Networking (SDN) and Internet of Things (IoT) has introduced new security challenges, necessitating effective Intrusion Detection Systems (IDS) tailored to the unique characteristics of SDN-IoT environments. This study presents a comprehensive evaluation of a newly generated IDS dataset specifically designed for SDN-IoT networks. The dataset is benchmarked against three widely used IDS datasets—InSDN, BoT-IoT, and ToN-IoT—using four deep learning architectures: Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN), and Deep Neural Network (DNN). Each model was trained and tested on all four datasets to assess performance across key metrics, including accuracy, precision, recall, F1-score, and computational efficiency. The results highlight the strengths and limitations of the proposed dataset in comparison to existing benchmarks, demonstrating the suitability of various deep learning models for anomaly detection in SDN-IoT contexts. This comparative analysis provides valuable insights for researchers and practitioners aiming to design robust, intelligent security systems in evolving network architectures.

1. Introduction

The convergence of Software-Defined Networking (SDN) and the Internet of Things (IoT) is reshaping modern network architectures by enhancing flexibility, scalability, and centralized management. SDN decouples the control and data planes, enabling dynamic, programmable network configuration [1], while IoT connects billions of heterogeneous devices, generating massive volumes of data. However, this integration introduces new and complex security challenges [2], particularly due to the distributed, resource-constrained, and heterogeneous nature of IoT environments coupled with the centralized control logic of SDN.

Intrusion Detection Systems (IDS) are critical components in safeguarding SDN-IoT networks. They monitor traffic to detect malicious activities and policy violations [3]. Traditional IDS approaches, however, often fall short when applied to the dynamic and diverse nature of SDN-IoT networks. Deep Learning (DL) techniques have emerged as powerful tools to enhance intrusion detection by automatically learning complex patterns from large-scale data [17,22,25]. Yet, the performance of DL-based IDS solutions is heavily influenced by the quality, relevance, and structure of the datasets used for training and evaluation [22,25].

Numerous publicly available datasets have been developed for IDS research, including BoT-IoT [4], ToN-IoT [5], and InSDN [6], each addressing different aspects of network behavior and attack types. Despite their contributions, these datasets may not fully capture the intricacies of modern SDN-IoT architectures or may lack balanced representation across attack categories. In response, this study introduces a novel, synthetically generated Network Intrusion Detection Systems (NIDS) dataset specifically designed for SDN-IoT networks, incorporating diverse attack vectors and realistic traffic patterns.

To evaluate the effectiveness of the proposed dataset, we conduct a comparative analysis using four DL models: Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN), and Deep Neural Network (DNN). These models are chosen for their proven success in time-series analysis, anomaly detection, and pattern recognition within network traffic.

The main contributions of this paper are as follows:

- We present a new IDS dataset tailored for SDN-IoT environments, addressing gaps in existing public datasets.

* Corresponding author at: University of Baghdad, Baghdad, Baghdad IRAQ.
E-mail address: heba.d@kecbu.uobaghdad.equ.iq (H. Dhirar).

https://doi.org/10.1016/j.meadig.2025.100015
Received 8 March 2025; Received in revised form 4 September 2025; Accepted 8 September 2025
Available online 9 September 2025

- We benchmark the proposed dataset against three well-known datasets (BoT-IoT, ToN-IoT, and InSDN) using consistent DL models and evaluation metrics.
- We analyze the performance of CNN, LSTM, RNN, and DNN models across all datasets, highlighting trends and identifying the most suitable models for SDN-IoT intrusion detection.

The remainder of this paper is organized as follows: Section 2 reviews related work; Section 3 details the dataset generation and description; Section 4 and 5 outlines the DL models and experimental metrics; Section 6 presents the results and discussion; and Section 7 concludes the paper with future directions.

2. Related work

The vast data volume and device diversity in IoT environments present serious security challenges. To address these, numerous Intrusion Detection Systems (IDSs) have been proposed using ML and DL techniques, especially within Software-Defined Networking (SDN) frameworks. Table 1 summarizes recent studies targeting IDS development in SDN and IoT networks.

Alzahrani et al. [7], applied XGBoost to the NSL-KDD [8] dataset for SDN-based NIDS, achieving 95.5 % accuracy using only five selected features. Despite its effectiveness, NSL-KDD is outdated and lacks SDN-specific patterns.

Elsayed et al. [9], developed SATIDS using an enhanced LSTM and evaluated it on ToN-IoT and InSDN datasets. Their model accurately classified multiple attack types but required high memory during training.

Mohsin et al. [10], assessed ML models (RF, KNN, NB, LG) for DDoS detection across various SDN topologies. RF and KNN showed strong performance, while NB and LG suffered from low accuracy and high false positives.

Jose et al. [11], compared DNN, CNN, and LSTM for IoT IDS using CIC-IDS2017 [12]. CNN and LSTM achieved high accuracies (98.61 % and 97.67 %, respectively), outperforming traditional methods.

Chaganti et al. [13], proposed an LSTM-based model for detecting SDN-IoT attacks such as port scanning and DoS, showing strong classification capabilities in hybrid environments.

Ali et al. [14], introduced a federated ANN-based IDS to detect Low-Rate DDoS attacks using the CAIDA dataset, achieving 98.85 % accuracy. However, CAIDA lacks IoT traffic, limiting relevance.

Raza et al. [15], used FL with the Edge-IIoTset dataset for SDN intrusion detection but did not discuss architectural details or dataset composition, raising concerns about scalability.

Alkhamisi et al. [16], proposed a GCNN-GRU model for Multi-Controller SDN, extending NSL-KDD with synthetic SDN traffic. While promising, its reliance on outdated data hinders real-world applicability.

Chatzimiltis et al. [17], introduced a Smart Meter IDS using Split Learning (SL) and FL for NAN security in SDN applications. The study lacked empirical validation in SDN-aware environments.

Kazmi et al. [18], proposed FCVAE to improve non-IID threat detection in SDN using the InSDN dataset. Accuracy improved from 92 % to 97 %, but the model's performance in real-time SDN-IoT settings remains unverified.

In summary, while prior work presented on Table 1, has advanced IDS development for SDN and IoT separately, few solutions effectively address the hybrid SDN-IoT context with realistic datasets and architectures. This motivates our proposed unified dataset and DL-based IDS framework tailored for dynamic SDN-IoT networks.

Table 1
Survey of the most related work of IDSs on the SDN and IoT network.

Ref.	Year	Network	Dataset	Technique	Accuracy
Alzahrani et al. [7]	2021	SDN	NSL-KDD [8]	XGBoost	Detection: 95.5 Classification: 95.95 %
ElSayed et al. [9]	2021	SDN	InSDN ToN-IoT	CNN+RF	99.28 %
Mohsin et al. [10]	2022	SDN	Custom	RF KNN NB LG	RF: 100 % KNN: 99.99–100 % NB: 72.11–83.5 % LG: 59.44–92.74 %
Jose et al. [11]	2023	IoT	CIC-IDS 2017 [12]	DNN LSTM CNN	94.61 % 97.67 % 98.61 %
Chaganti et al. [13]	2023	SDN-IoT	SDN IoT-focused	LSTM	97.1 %
Ali et al. [14]	2023	SDN-IoT	CAIDA [19]	ANN	98.85 %
Raza et al. [15]	2024	SDN-IoT	Edge-IIoTset [20]	ANN	98.65 %
Alkhamisi et al. [16]	2024	MC-SDN	NSL-KDD	GCNN-GRU	97.78٪ 92.15٪ 96.12٪
Chatzimiltis et al. [17]	2024	SDN	NSL-KDD	SL-NLS	99 %
Kazmi et al. [18]	2025	SDN	InSDN	DL	97 %
Our work	2025	SDN-IoT	Custom BoT-IoT ToN-IoT InSDN	LSTM CNN RNN DNN	98.48 %−96.65 %- 95.25 %−95.25 %; 83.84 %−86.94 % 78.19 %−70.63 %; 70.28 %−75.00 % 46.99 %−56.74 %; 53.92 %−58.29 % 57.90 %−53.42 %

3. Methodology

High-quality datasets are essential for training accurate AI-based IDS. However, privacy concerns often restrict access to real network traffic. While legacy datasets remain common for benchmarking, they are inadequate for real-time detection, especially in SDN environments, as they fail to capture modern attack behavior. This section details a custom dataset generation named SDN-IoT for NIDS, including network traffic simulation, feature extraction from SDN flow tables, and normalization techniques to ensure model-ready data. Fig. 1 illustartes the dataset generation steps.

3.1. Attack types analysis

The initial step in building the dataset is identifying the attack classes affecting the SDN network and their detection level. Table 2 presents a complete study of the attack activities and their detection classification [21,22]. The attacks can be divided into 7 primary groups: DDoS, DoS, Probe Attack, Web attacks, Malware, Remote to Local (R2L), and User to Root (U2R). To provide a reliable and functional NIDS dataset, only network-level detectable attacks (DoS, DDoS, Probe, Web attack) were included. Current network emulators like Mininet-WiFi primarily generate basic connectivity traffic, which lacks the complexity and variability of real-world network behavior [23]. This constrained normal traffic profile risks interference with other traffic types.

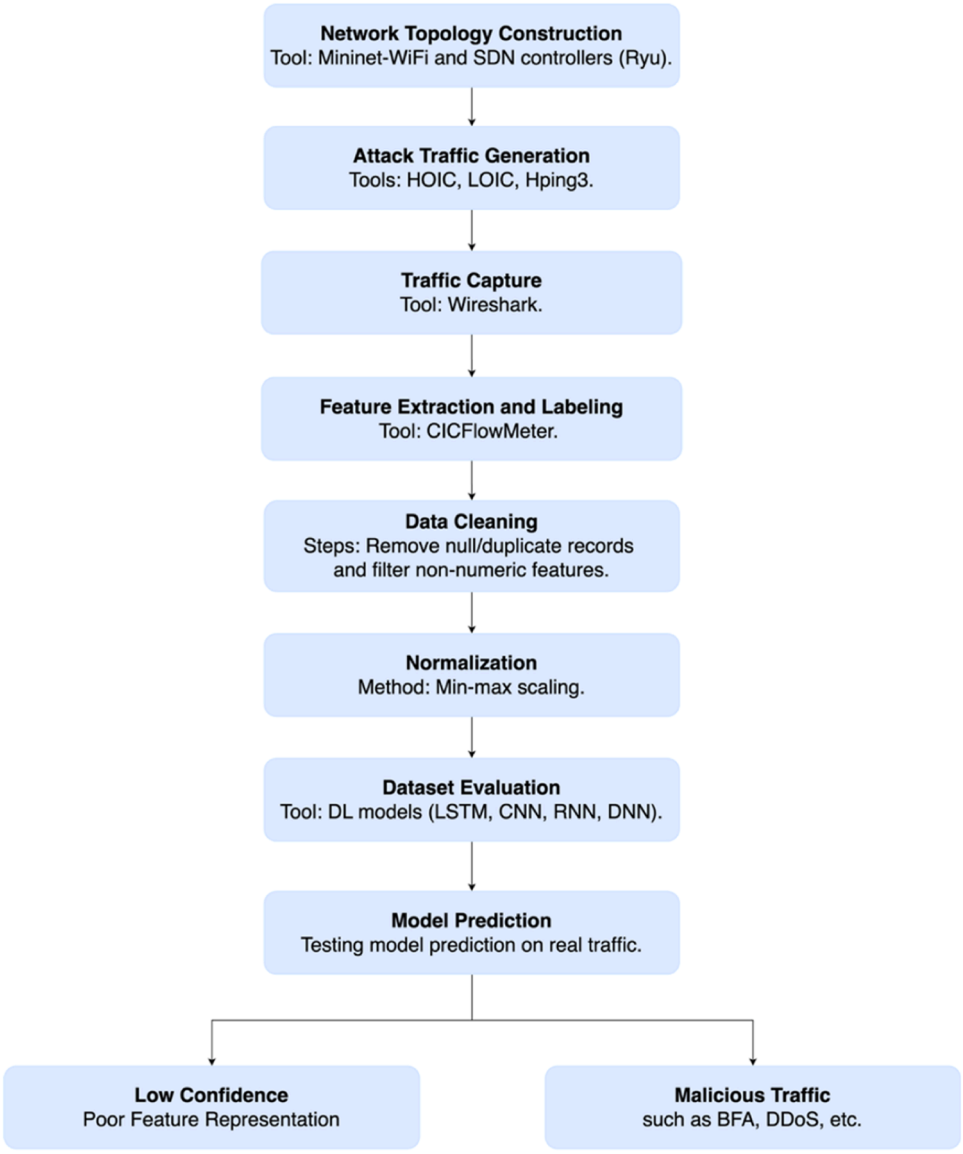


Fig. 1. Dataset creation process diagram.

Table 2
Comprehensive analysis of attack activities and detection classification.

Class	Attack Activities	Key Network Indicators	Host Indicators	Classification
DoS	UDP/HTTP/SYN Floods	High flow duration, packet length anomalies.	–	Network
DDoS	ICMP Flood, TCP SYN Attack	High packet forwarding, low active time.	–	Network
Probe	Port/IP Scanning	Short flows, high packet variance.	–	Network
Web Attacks	SQLi, XSS, CSRF	Small packets, high backward packet deviations.	Session/request anomalies.	Both
Malware	Ransomware, Spyware	–	Irregular flows, high throughput.	Host
R2L	Phishing, Backdoors	–	Small packets, irregular byte patterns.	Host
U2R	Buffer Overflows	–	Abnormal timing, high backward packets.	Host

3.2. Attack types analysis

Various open-source tools were utilized to simulate a range of attack activities originating from diverse sources and targeting multiple destinations. An overview of these tools is presented in Table 3. The resulting traffic was captured using Wireshark and stored locally as PCAP files for subsequent analysis. Wireshark, an open-source packet analyzer, enables the identification of potential network threats by recording and analyzing trace files generated in real-time network environments [24].

3.3. Dataset creation

To generate a comprehensive intrusion detection dataset, each attack class was simulated and captured as individual PCAP files using Wireshark within the controller's virtual environment. Approximately 400 million packets were collected, as detailed in Table 4, ensuring a robust representation of malicious traffic patterns. The attacks were systematically orchestrated using specialized tools deployed on dedicated attacker machines, each targeting victim devices with predefined IP addresses within the network. This structured approach ensures precise

Table 3
Overview of prominent network attack tools.

Tools	Network
High Orbit Ion Cannon (HOIC)	An open-source testing program that intentionally employs mobile devices to engage in ongoing attacks [25]
Low Orbit Ion Cannon (LOIC)	
Metasploit	Allows its users to access its source code and incorporate their bespoke modules to execute operations ranging from scanning to exploitation across various other environments [26].
Nmap	A network reconnaissance tool for scanning hosts, auditing services, and monitoring, widely leveraged in hacking for vulnerability assessment [26].
Hping3	Regarded as one of the most publicly accessible tools for DDoS attacks, it is utilized to produce various types of flooding attacks [6].

attack characterization while maintaining reproducibility for IDS evaluation.

3.4. Feature extraction and normalization

The captured PCAP files were processed using CICFlowMeter, that generates structured CSV outputs, each containing 83 statistical features per network flow such as duration, packet size, and protocol metrics. Each CSV file represents a distinct attack class, with manual labeling applied to ensure accurate classification for intrusion detection analysis. From this processing pipeline, a total of 859,253 flow records were generated, resulting in a dataset size of approximately 430 MB, forming a comprehensive foundation for attack detection and analysis. The raw dataset was carefully processed, preprocessing involved removing non-numeric values and IP addresses to ensure privacy and compatibility with algorithmic processing, eliminating duplicate records to prevent model bias, and handling infinite and undefined values through systematic imputation. Following this feature refinement, min-max normalization was applied to standardize the dataset while preserving relative relationships between features. This comprehensive preprocessing pipeline transformed the raw network flows into model-ready data suitable for both training and evaluation of IDSs.

3.5. Dataset performance comparison

To evaluate the quality of the generated dataset in representing various attacks, we selected the popular SDN dataset InSDN, along with two notable realistic IoT datasets, ToN-IoT and BoT-IoT, for comparison. We then assessed the dataset quality by measuring the performance of various ML/DL-based IDS models in detecting attacks when trained on each of these datasets.

3.6. Dataset partitioning

To enable a fair and computationally efficient comparison with the large-scale BoT-IoT and ToN-IoT datasets, a representative subset of 860,000 records was extracted using stratified sampling. Then each datasets were then split into training (70 %), testing (20 %), and validation (10 %) sets based on balanced data distributions such that the new dataset contain an equal number of samples from each class, this balanced situation demonstrates superior performance [27], as present on Table 5.

4. Deep learning-based intrusion detection model

Empirical studies' methods demonstrate that DL models consistently outperform conventional techniques in accuracy, scalability, and adaptability to evolving threats, especially when trained on extensive datasets. Table 6 highlights the most effective DL-based IDS approaches [28].

The model achieved higher accuracy with a simpler architecture, likely due to better generalization on the balanced distributed dataset, which may not require the complexity of deeper networks to learn effectively. The model was optimized using the Adam optimizer with a categorical cross-entropy loss function, trained over 5 epochs per round with a batch size of 64. The initial learning rate was set to 0.001 with a decay rate ($\gamma = 0.9$), complemented by L2 wt decay ($\lambda = 10^{-4}$) to regularize network parameters and prevent overfitting. Detailed architectural specifications of each trained model are summarized in Table 7.

5. Statistical metrics

Performance evaluation of a classification model often involves several key metrics, such as accuracy, precision, recall, and F1-score. These metrics are derived from the confusion matrix, and include True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN), which are the key metrics in the evaluation of classification performance [56]:

- **True positive:** It refers to the number of correctly classified positive instance, where the actual and predicted outcomes are both positive.
- **True negative:** It represents the correctly classified negative instance, where both the actual and predicted outcomes are negatives.
- **False positive:** It occurs when the model incorrectly classifies a negative event as positive.
- **False negative:** It happens when the model incorrectly classifies a positive event as negative.

These metrics are essential for calculating other performance measures and to assess the accuracy and effectiveness of multi-class classification models [57,58]. For multi-class classification, let C_i denote the i_{th}class ($i = 1,..,l$), with TP_i, FP_i, FN_i, and TN_i representing the TP, FP, FN, and TN for C_i, respectively. The generalized evaluation for multi-class using μ micro-averaging that pools all class predictions globally can then be formulated as [59]:

Accuracy: Calculates the ratio of accurate forecasts to the total number of predictions.

$$Accuracy = \frac{\sum_{i=1}^{l} \frac{TP_i + TN_i}{TP_i + FP_i + FN_i + TN_i}}{l} \quad (2)$$

Precision: Calculates the ratio of accurately predicted positive observations to the total anticipated positives.

$$Precision = \frac{\sum_{i=1}^{l} TP_i}{\sum_{i=1}^{l} (TP_i + FP_i)} \quad (3)$$

Recall: Measures the ratio of true positives accurately detected by the model.

Table 4
Generated IDS dataset specifics.

Attack class	Activities	Tools	Captured Packets	Records Number
DDoS/DoS	Http, TCP, UDP, ICMP, TCP-SYN.	HOIC, LOIC, metaspliot, hping3	362,430,567	262,652
Probe	Port scanning, vulnerability scanning, IP sweeps, Host discovery.	nmap	23,358,817	305,675
Web attack	Brute force attack (BFA).	LOIC	12,628,681	290,929

Table 5
The most effective DL model for NIDS.

Dataset	Samples number	Training	Testing	Validation
SDN-IoT	Training: 590,899	Probe: 36.09 %	Probe: 36.09 %	Probe: 36.09 %
	Validation: 84,583	BFA: 34.18 %	BFA: 34.18 %	BFA: 34.18 %
	Testing: 168,660	DoS/DDoS: 29.72 %	DoS/DDoS: 29.72 %	DoS/DDoS: 29.72 %
InSDN	Training: 237,698	DDoS: 32.37 %	DDoS: 32.37 %	DDoS: 32.36 %
	Validation: 34,024	Probe: 28.89 %	Probe: 28.89 %	Probe: 28.88 %
	Testing: 67,847	Normal: 20.14 %	Normal: 20.15 %	Normal: 20.14 %
		DoS: 15.78 %	DoS: 15.78 %	DoS: 15.78 %
		DDoS: 2.26 %	DDoS: 2.26 %	DDoS: 2.26 %
		BFA: 0.41 %	BFA: 0.41 %	BFA: 0.41 %
		Web-Attack: 0.05 %	Web-Attack: 0.05 %	Web-Attack: 0.05 %
		BOTNET: 0.04 %	BOTNET: 0.04 %	BOTNET: 0.04 %
		U2R: 0.005 %	U2R: 0.005 %	U2R: 0.004 %
ToN-IoT	Training: 601,999	Benign: 46.98 %	Benign: 46.98 %	Benign: 46.98 %
	Validation: 86,172	XSS: 40.16 %	XSS: 40.16 %	XSS: 40.16 %
	Testing: 171,828	Password: 6.35 %	Password: 6.35 %	Password: 6.35 %
		Injection: 5.19 %	Injection: 5.18 %	Injection: 5.19 %
		Scanning: 0.67 %	Scanning: 0.67 %	Scanning: 0.67 %
		Backdoor: 0.50 %	Backdoor: 0.50 %	Backdoor: 0.50 %
		Ransomware: 0.09 %	Ransomware: 0.09 %	Ransomware: 0.09 %
		MIMT: 0.009 %	MIMT: 0.01 %	MIMT: 0.009 %
		DDoS: 0.003 %	DDoS: 0.003 %	DDoS: 0.0003 %
		DoS: 0.002 %	DoS: 0.002 %	DoS: 0.0002 %
BoT-IoT	Training: 601,998	DDoS: 36.59٪	DDoS: 36.59٪	DDoS: 36.59٪
	Validation: 86,172	DoS: 36.56٪	DoS: 36.56٪	DoS: 36.56٪
	Testing: 171,828	Reconnaissance: 26.16٪	Reconnaissance: 26.16٪	Reconnaissance: 26.16٪
		Benign:0.66٪	Benign:0.66٪	Benign:0.66٪
		Theft:0.01٪	Theft:0.01٪	Theft:0.01٪

Table 6
The most effective DL model for NIDS.

Studies	Models	Accuracy range (%)
[9,16,29–38]	CNN	74.71–100
[31,36,39–44]	LSTM	87.0–99.97
[45–49]	CNN+LSTM	76.8–98.8
[29,50,51]	RNN	89.0–99.71
[52]	LSTM+RNN	98.8
[29,31,53–55]	DNN	89.0–99.9

Table 7
Architecture specifications and training parameters for DL models.

Model	Layers	Neurons	Hyperparameters
LSTM	Input → LSTM (5x) → Output	256 → 128 → 128 → 64 → 32 → Output	Activations: Tanh (hidden), Softmax (output).
CNN	Input → Conv1D → MaxPool → Conv1D → MaxPool → Conv1D → MaxPool → Flatten → Dense (2x) → Output	256 → 128 →128 → 64 → 32 → Output	Activations: ReLU (hidden), Softmax (output), Kernal size = 2,3.
RNN	Input → GRU → Dense (2x) → Output	64 → 64 → 32 → Output	Activations: ReLU (hidden), Softmax (output).
DNN	Input → Dense (5x) → Output	256 → 128 → 128 → 64 → 32 → Output	Activations: ReLU (hidden), Softmax (output).

$$Recall = \frac{\sum_{i=1}^{l} TP_i}{\sum_{i=1}^{l} (TP_i + FN_i)} \tag{4}$$

F1 Score: It represents the harmonic mean of precision and recall, establishing a balance between the two metrics.

$$F1\ score = \frac{(B^2) * Precision * Recall}{B^2 * Precision + Recall} \tag{5}$$

Confusion matrix: Is an essential tool for evaluating classification algorithm. It provides a summary of the predictive outcomes for a categorization task, as shown on Fig. 2.

Loss functions quantify model prediction errors during training, guiding optimization algorithms to adjust parameters. They are essential for improving accuracy in tasks. The standard weighted categorical cross-entropy loss is given by [61]:

$$Loss\ function = -\frac{1}{M} \sum_{k=1}^{K} \sum_{m=1}^{M} w_k * y_m^k * \log(h_\theta(x_m, k)) \tag{6}$$

Where M is the total training smaples, K is the number of classes, w_k represents the class weight, y_m^k represents the label, x_m is the input vector, and h_θ represent a neural network model with parameters θ.

6. Results

The analysis of model performance across the four datasets—BoT-IoT, InSDN, SDN-IoT, and ToN-IoT—reveals distinct learning behaviors and dataset characteristics. In the BoT-IoT dataset, the DNN model demonstrates the strongest early performance, achieving over 89 % accuracy by the fifth epoch, while LSTM and RNN show moderate but consistent improvement, indicating effective learning but at different rates. In contrast, models trained on the InSDN dataset, particularly the RNN, begin with very low accuracy and precision but improve rapidly, reflecting the dataset's complexity and potential class imbalance. For the SDN-IoT dataset, DNN again exhibits excellent generalization, with both training and validation accuracies surpassing 90 % early in the training process, suggesting a balanced and learnable dataset. However, in the ToNIoT dataset, LSTM performance starts extremely low, with initial accuracy below 10 %, improving slowly due to the severe class imbalance (as previously noted), which hinders early convergence. Overall, DNN consistently provides strong baseline performance across datasets, while LSTM and RNN require more epochs to stabilize, particularly in datasets with skewed class distributions, Figs. 3–6 present the expermentail result of each DL model for the datasets.

Across all datasets, model performance varied significantly based on data balance and attack diversity. On the SDN-IoT dataset, all models performed strongly, with DNN achieving the highest accuracy of 98.48 %, followed by LSTM at 96.65 %, demonstrating robust detection of BFA, DoS/DDoS, and Probe attacks. In contrast, the InSDN dataset

		PREDICTED classification				
	Classes	a	b	c	d	Total
ACTUAL classification	a	6	0	1	2	9
	b	3	9	1	1	14
	c	1	0	10	2	13
	d	1	2	1	12	16
	Total	11	11	13	17	**52**

Fig. 2. Multi-class confusion matrix [60].

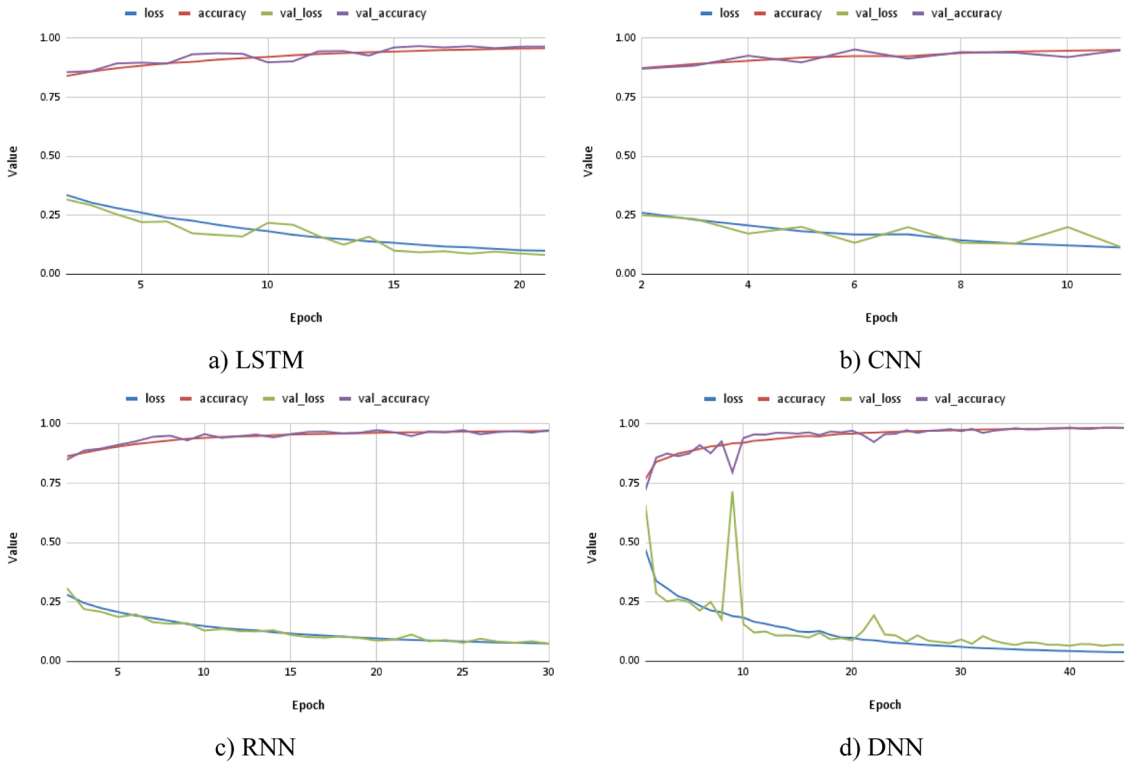


Fig. 3. The expermentail results of SDN-IoT dataset.

showed more variability due to class imbalance; LSTM achieved 86.94 % accuracy, while minority classes like Web-Attack and U2R were poorly detected across all models. The ToN-IoT dataset further highlighted imbalance issues—LSTM performed best at 75.00 %, effectively identifying common attacks like XSS and Benign, whereas CNN and RNN performed poorly on rare attack types. The BoT-IoT dataset proved the most challenging; although LSTM reached 58.29 % accuracy by detecting Reconnaissance well, all models struggled with minor classes like Theft and Benign. Overall, DNN and LSTM consistently outperformed CNN and RNN, particularly on balanced datasets, Figs. 7–10 present the classification reports of each DL model for the datasets.

7. Real-Time detection performance evaluation

The experimental outcome in Table 8 show that models achieve strong real-time detection on the SDN-IoT and InSDN datasets, particularly for DDoS/DoS and Probe attacks. DNN consistently maintained stable high-confidence classifications, while RNN often misclassified Probe traffic as DDoS. In contrast, performance on the ToN-IoT and BoT-IoT datasets was markedly weaker, with low-confidence or incorrect predictions, especially for BFA and Probe attacks. Despite these differences, inference latency remained below 17 ms across all datasets and packet sizes, demonstrating that real-time requirements can be satisfied. These results highlight a methodological limitation: evaluating models solely within the same dataset may overstate performance. The high accuracy observed in SDN-IoT may reflect dataset-specific simplicity rather than realistic attack complexity, as evidenced by the weaker results on ToN-IoT and BoT-IoT. To address this concern, further quality assessment of the proposed dataset should include cross-dataset evaluations, augmentation studies, and statistical similarity analyses. Such steps are essential to determine whether the synthetic dataset truly captures real-world network behavior and to ensure robust

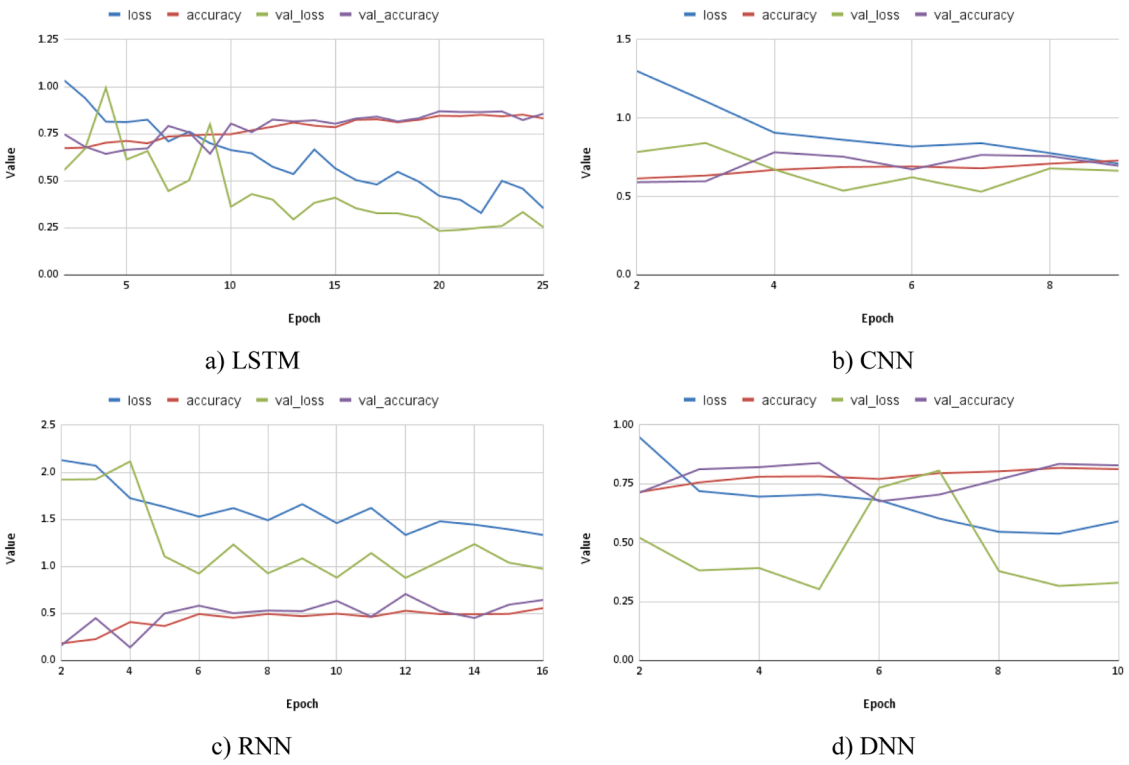


Fig. 4. The expermentail results of InSDN dataset.

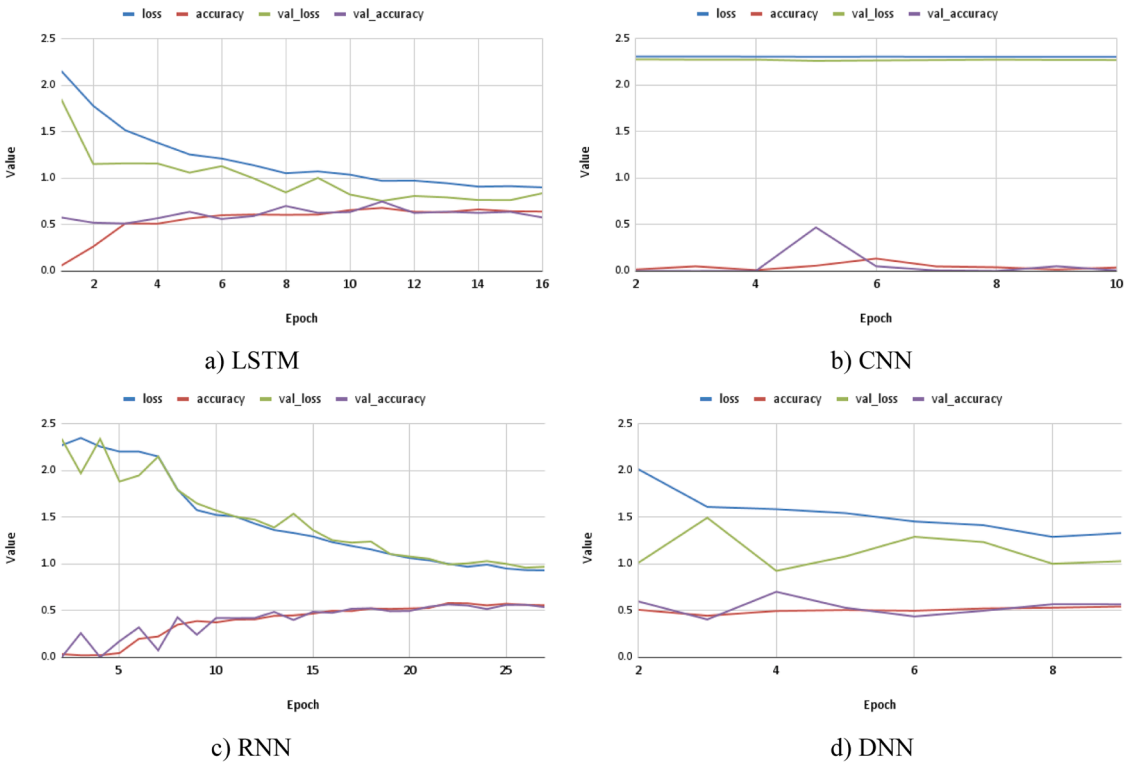


Fig. 5. The expermentail results of ToN-IoT dataset.

machine learning-based IDS deployment.

8. Conclusion

Based on the experimental results, including both offline training performance and real-time traffic classification, the custom-generated SDN-IoT dataset demonstrates clear superiority over benchmark datasets such as InSDN, ToN-IoT, and BoT-IoT. Its balanced class distribution—Probe (36.09 %), BFA (34.18 %), and DoS/DDoS (29.72 %)—ensures equal representation across attack types, enabling deep learning

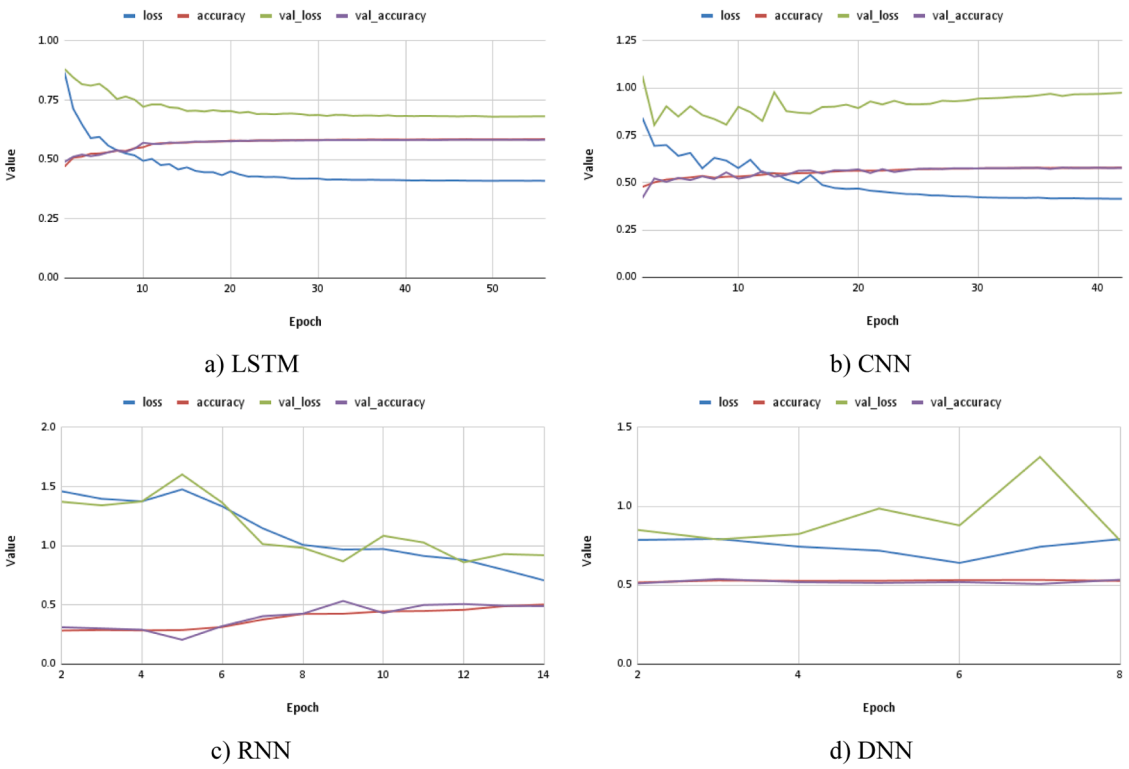


Fig. 6. The expermentail results of BoT-IoT dataset.

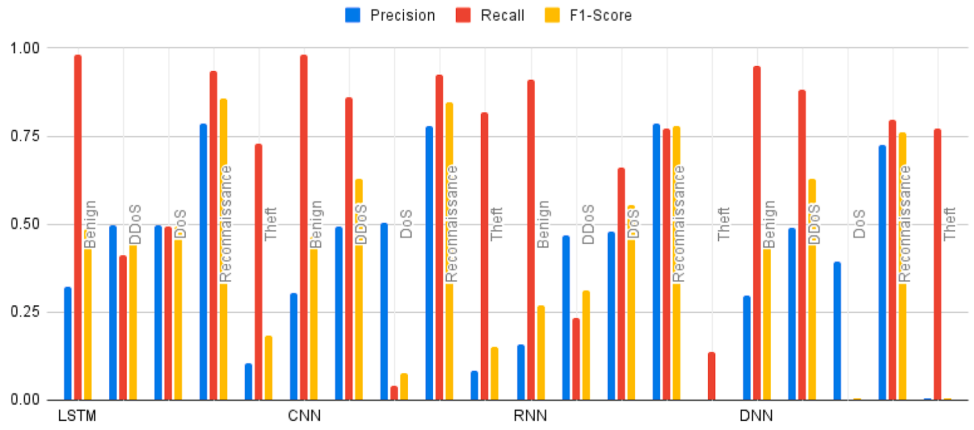


Fig. 7. The classification report of BoT-IoT dataset.

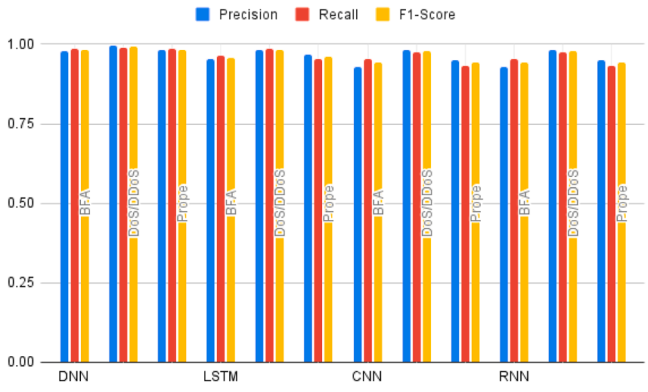


Fig. 8. The classification report of SDN-IoT dataset.

models (especially DNN and LSTM) to consistently achieve high accuracy, exceeding 96 %. In the real-time prediction tests with 300 packets, the SDN-IoT dataset maintained robust performance across all models, with LSTM, CNN, RNN, and DNN achieving high confidence scores ($\geq$0.98) and low inference times (9–16 ms). This shows the dataset's practical usability for real-world IDS deployment, where fast and reliable detection is critical. In contrast, InSDN exhibited frequent misclassifications, with attacks like BFA and DoS often being mislabeled as Probe, despite high confidence values, reflecting its class imbalance problem. ToN-IoT performed poorly in real-time detection, where models either failed to classify (low confidence for LSTM and CNN) or consistently misclassified traffic as unrelated classes (e.g., Ransomware or Benign). BoT-IoT also showed inconsistent results, with models like LSTM and RNN producing low-confidence or incorrect labels, while only CNN and DNN maintained high-confidence predictions in certain cases. Overall, unlike the benchmark datasets that suffer from severe imbalance or limited generalizability, the proposed SDN-IoT dataset provides

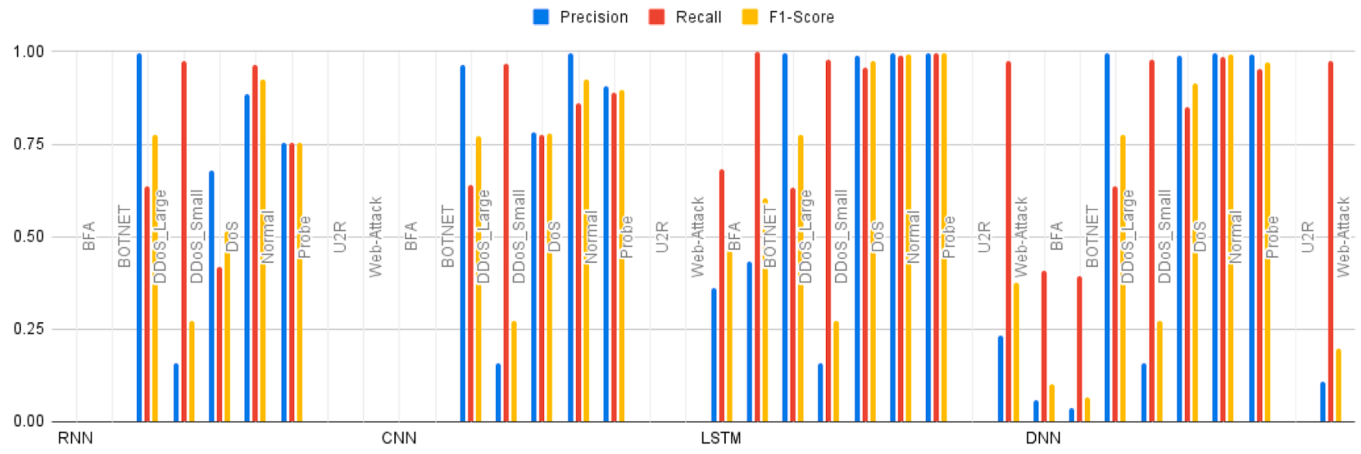


Fig. 9. The classification report of InSDN dataset.

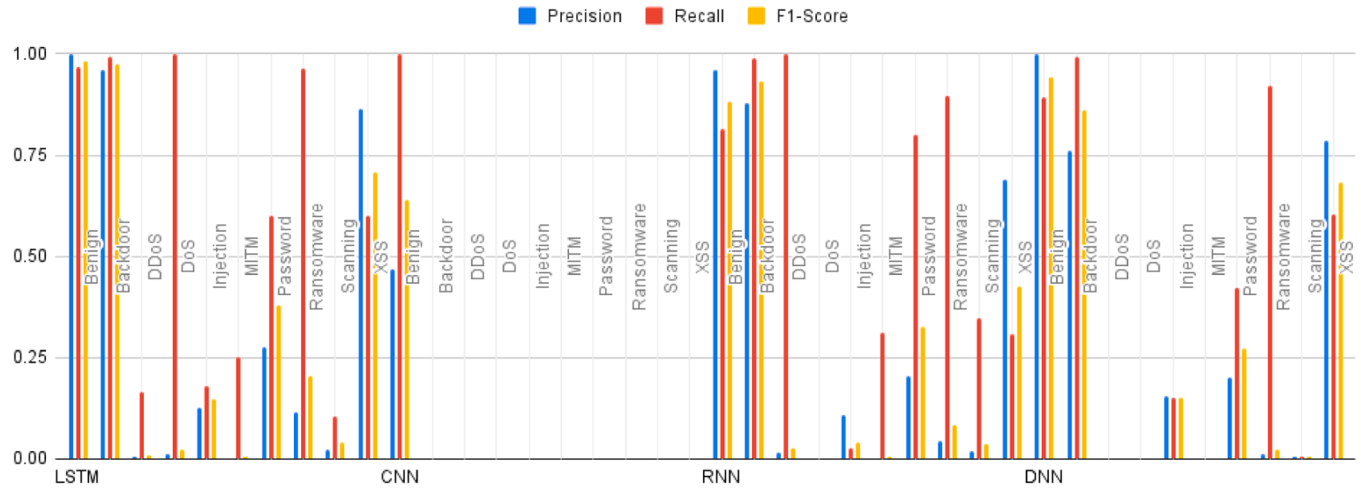


Fig. 10. The classification report of ToN-IoT dataset.

Table 8
Classification testing of real-time traffic on each dataset for 300 packets.

Dataset	Generated Attack	Predicted class, confidence, and time, on each model			
		LSTM	**CNN**	**RNN**	**DNN**
SDN-IoT	BFA	BFA - 0.9983, (11 ms)	DDoS/DoS - 0.9875, (11 ms)	Probe - 0.9834, (12 ms)	Probe - 0.9829, (10 ms)
	DDoS/DoS	DDoS/DoS - 0.9999, (13 ms)	DDoS/DoS - 1.0, (12 ms)	DDoS/DoS - 1.0, (15 ms)	DDoS/DoS - 1.0, (11 ms)
	Probe	Probe - 0.9996, (15 ms)	Probe - 0.9988, (15 ms)	DDoS/DoS - 1.0, (16 ms)	Probe - 1.0, (9 ms)
InSDN	BFA	DoS - 0.9993, (11 ms)	Probe - 1.0, (11 ms)	Probe - 1.0, (12 ms)	Probe - 1.0, (9 ms)
	DDoS/DoS	Probe - 0.9999, (13 ms)	Probe - 1.0, (12 ms)	Probe - 1.0, (16 ms)	Probe - 1.0, (12 ms)
	Probe	Probe - 0.9999, (11 ms)	Probe - 1.0, (11 ms)	Probe - 1.0, (12 ms)	Probe - 1.0, (10 ms)
ToN-IoT	BFA	None – 0.4545 (11 ms)	None – 0.1196 (11 ms)	Ransomware – 0.9999 (12 ms)	Benign – 1.0 (9 ms)
	DDoS/DoS	None – 0.4198 (13 ms)	None – 0.1196 (13 ms)	Ransomware – 0.9951 (14 ms)	Benign – 1.0 (12 ms)
	Probe	None – 0.4667 (11 ms)	None – 0.1195 (11 ms)	Ransomware – 0.9945 (13 ms)	Benign – 1.0 (10 ms)
BoT-IoT	BFA	None - 0.4303, (12 ms)	DoS - 1.0, (11 ms)	None - 0.5856, (14 ms)	DDoS - 1.0, (10 ms)
	DDoS/DoS	Reconnaissance - 0.9287, (13 ms)	DoS - 1.0, (13 ms)	None - 0.5407, (15 ms)	Reconnaissance - 1.0, (11 ms)
	Probe	None - 0.5079, (11 ms)	DoS - 1.0, (11 ms)	None - 0.5320, (13 ms)	DDoS - 1.0, (10 ms)

both high offline accuracy and reliable real-time detection**.** Its balanced and diverse attack representation makes it more suitable for developing accurate, resilient, and realistic IDS solutions in SDN-based IoT environments.

CRediT authorship contribution statement

Heba Dhirar: Writing – review & editing, Supervision. **Ali Hamad:** Writing – original draft, Software, Resources, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The dataset used is publicly access at Kaggle.

References

[1] S. Soltani, A. Amanlou, M. Shojafar, R. Tafazolli, Security of topology discovery service in SDN: vulnerabilities and countermeasures, IEEE Open J. Commun. Society 5 (2024) 3410–3450, https://doi.org/10.1109/OJCOMS.2024.3406489.

[2] Z. Wen, R. Yang, P. Garraghan, T. Lin, J. Xu, M. Rovatsos, Fog orchestration for internet of things services, IEEe Internet. Comput. (2017), https://doi.org/10.1109/MIC.2017.36.

[3] S. Akbar, T.S. Rao, M.Ali Hussain, A hybrid scheme based on big data analytics using intrusion Detection System, Indian J. Sci. Technol. 9 (Sep. 2016), https://doi.org/10.17485/ijst/2016/v9i33/97037.

[4] "The bot-IoT dataset | UNSW Research." Accessed: May 27, 2025. [Online]. Available: https://research.unsw.edu.au/projects/bot-iot-dataset.

[5] "ToN_IoT datasets | IEEE DataPort." Accessed: Nov. 24, 2024. [Online]. Available: https://ieee-dataport.org/documents/toniot-datasets.

[6] M.S. Elsayed, N.A. Le-Khac, A.D. Jurcut, InSDN: a novel SDN intrusion dataset, IEEe Access. (2020), https://doi.org/10.1109/ACCESS.2020.3022633.

[7] A.O. Alzahrani, M.J.F. Alenazi, Designing a network intrusion detection system based on machine learning for software defined networks, Future Internet. (2021), https://doi.org/10.3390/fi13050111.

[8] "NSL-KDD | IEEE DataPort." Accessed: Nov. 24, 2024. [Online]. Available: https://ieeedataport.org/documents/nsl-kdd-0.

[9] M.S. ElSayed, N.A. Le-Khac, M.A. Albahar, A. Jurcut, A novel hybrid model for intrusion detection systems in SDNs based on CNN and a new regularization technique, J. Network Comput. Appl. (2021), https://doi.org/10.1016/j.jnca.2021.103160.

[10] M.A. Mohsin, A.H. Hamad, Performance evaluation of SDN DDoS attack detection and mitigation based random forest and K-nearest neighbors machine learning algorithms, Revue d'Intelligence Artificielle (2022), https://doi.org/10.18280/ria.360207.

[11] J. Jose, D.V. Jose, Deep learning algorithms for intrusion detection systems in internet of things using CIC-IDS 2017 dataset, Int. J. Electr. Comput. Eng. 13 (1) (2023) 1134–1141, https://doi.org/10.11591/IJECE.V13I1.PP1134-1141.

[12] "IDS 2017 | datasets | research | Canadian Institute for Cybersecurity | UNB." Accessed: Nov. 24, 2024. [Online]. Available: https://www.unb.ca/cic/datasets/ids-2017.html.

[13] R. Chaganti, W. Suliman, V. Ravi, and A. Dua, "Deep learning approach for SDN-enabled intrusion detection system in IoT networks," Information 2023, Vol. 14, Page 41, vol. 14, no. 1, p. 41, Jan. 2023, doi: 10.3390/INFO14010041.

[14] M.N. Ali, M. Imran, M.S. ud din, B.S. Kim, Low rate DDoS detection using weighted federated learning in SDN control plane in IoT network, Applied Sciences 13 (2023) 1431, https://doi.org/10.3390/APP13031431. Vol.Pagevol. 13, no. 3, p. 1431, Jan. 2023.

[15] M. Raza, M. Jasim Saeed, M.B. Riaz, M. Awais Sattar, Federated learning for privacy-preserving intrusion detection in software-defined networks, IEEe Access. 12 (2024) 69551–69567, https://doi.org/10.1109/ACCESS.2024.3395997, vol.

[16] A. Alkhamisi, I. Katib, S.M. Buhari, Federated learning-based security attack detection for multi-controller software-defined networks, Algorithms. 17 (Jul. 2024), https://doi.org/10.3390/a17070290.

[17] S. Chatzimiltis, M. Shojafar, M.B. Mashhadi, R. Tafazolli, A collaborative software defined network-based smart grid intrusion detection system, IEEE Open J. Commun. Society 5 (2024) 700–711, https://doi.org/10.1109/OJCOMS.2024.3351088.

[18] S.H.A. Kazmi, R. Hassan, F. Qamar, K. Nisar, M.A. Al-Betar, Federated Conditional Variational Auto Encoders for Cyber Threat Intelligence: Tackling Non-IID Data in SDN Environments, IEEE Access, Jan. 2025, https://doi.org/10.1109/ACCESS.2025.3529894.

[19] "CAIDA UCSD DDoS 2007 Attack Dataset | IEEE DataPort." Accessed: May 27, 2025. [Online]. Available: https://ieee-dataport.org/documents/caida-ucsd-ddos-2007-attackdataset.

[20] M.A. Ferrag, O. Friha, D. Hamouda, L. Maglaras, H. Janicke, Edge-IIoTset: a new comprehensive realistic cyber security dataset of IoT and IIoT applications for centralized and federated learning, IEEe Access. (2022), https://doi.org/10.1109/ACCESS.2022.3165809.

[21] I. Sharafaldin, A.H. Lashkari, A.A. Ghorbani, Toward generating a new intrusion detection dataset and intrusion traffic characterization, in: ICISSP 2018 - Proceedings of the 4th International Conference on Information Systems Security and Privacy, SciTePress, 2018, pp. 108–116, https://doi.org/10.5220/0006639801080116.

[22] P. García-Teodoro, J. Díaz-Verdejo, G. Maciá-Fernández, E. Vázquez, Anomaly-based network intrusion detection: techniques, systems and challenges, Comput. Secur. 28 (2009) 18–28, https://doi.org/10.1016/j.cose.2008.08.003.

[23] A. Hirsi, L. Audah, A. Salh, M.A. Alhartomi, S. Ahmed, Detecting DDoS threats using supervised machine learning for traffic classification in software defined networking, IEEe Access. (2024), https://doi.org/10.1109/ACCESS.2024.3486034.

[24] G. Jain, Anubha, Application of SNORT and wireshark in network traffic analysis, IOP. Conf. Ser. Mater. Sci. Eng. 1119 (1) (2021) 012007, https://doi.org/10.1088/1757- 899X/1119/1/012007.

[25] S. Bhatia, S. Behal, I. Ahmed, Distributed denial of service attacks and defense mechanisms: current landscape and future directions, Adv. Inf. Security 72 (2018) 55–97, https://doi.org/10.1007/978-3-319-97643-3_3.

[26] S. Raj, N.K. Walia, A study on Metasploit Framework: a pen-testing tool2020 International Conference on Computational Performance Evaluation, ComPE (2020) 296–302, https://doi.org/10.1109/COMPE49325.2020.9200028. Jul. 2020.

[27] E.M. Campos, et al., Evaluating federated Learning for intrusion detection in Internet of Things: review and challenges, Comput. Netw. 203 (2022), https://doi.org/10.1016/j.comnet.2021.108661.

[28] O.H. Abdulganiyu, T.Ait Tchakoucht, Y.K. Saheed, A systematic literature review for network intrusion detection system (IDS), Int. J. Inf. Secur. 22 (5) (2023) 1125–1162, https://doi.org/10.1007/S10207-023- 00682-2, vol. 22, no. 52023.

[29] O. Friha, M.A. Ferrag, L. Shu, L. Maglaras, K.K.R. Choo, M. Nafaa, FELIDS: federated learning-based intrusion detection system for agricultural Internet of Things, J. Parallel. Distrib. Comput. 165 (Jul. 2022) 17–31, https://doi.org/10.1016/j.jpdc.2022.03.003.

[30] M.H. Bhavsar, Y.B. Bekele, K. Roy, J.C. Kelly, D. Limbrick, FL-IDS: federated learning-based intrusion detection system using edge devices for transportation IoT, IEEe Access. 12 (2024) 52215–52226, https://doi.org/10.1109/ACCESS.2024.3386631.

[31] J. Mateus, G.A.L. Zodi, A. Bagula, Federated learning-based solution for DDoS detection in SDN, in: 2024 International Conference on Computing, Networking and Communications, ICNC 2024, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 875–880, https://doi.org/10.1109/ICNC59896.2024.10556115.

[32] Y. Xiao, C. Xing, T. Zhang, Z. Zhao, An intrusion detection model based on feature reduction and convolutional neural networks, IEEe Access. 7 (2019) 42210–42219, https://doi.org/10.1109/ACCESS.2019.2904620.

[33] J. Kim, J. Kim, H. Kim, M. Shim, and E. Choi, "CNN-based network intrusion detection against denial-of-service attacks," Electr. (Basel) 2020, Vol. 9, Page 916, vol. 9, no. 6, p. 916, Jun. 2020, doi: 10.3390/ELECTRONICS9060916.

[34] Z. Wu, J. Wang, L. Hu, Z. Zhang, H. Wu, A network intrusion detection method based on semantic re-encoding and deep learning, J. Network Comput. Appl. 164 (2020) 102688, https://doi.org/10.1016/J.JNCA.2020.102688.

[35] M.T. Nguyen, K. Kim, Genetic convolutional neural network for intrusion detection systems, Future Gener. Comput. Syst. 113 (2020) 418–427, https://doi.org/10.1016/J.FUTURE.2020.07.042.

[36] N. Gupta, V. Jindal, P. Bedi, LIO-IDS: handling class imbalance using LSTM and improved one-vs-one technique in intrusion detection system, Comput. Netw. 192 (2021) 108076, https://doi.org/10.1016/J.COMNET.2021.108076.

[37] L. Ashiku, C. Dagli, Network intrusion Detection System using Deep Learning, Procedia Comput. Sci. 185 (2021) 239–247, https://doi.org/10.1016/J.PROCS.2021.05.025.

[38] R.H. Hwang, M.C. Peng, C.W. Huang, P.C. Lin, V.L. Nguyen, An unsupervised deep learning model for early network traffic anomaly detection, IEEe Access. 8 (2020) 30387–30399, https://doi.org/10.1109/ACCESS.2020.2973023.

[39] P. Singh, G.S. Gaba, A. Kaur, M. Hedabou, A. Gurtov, Dew-cloud-based hierarchical federated learning for intrusion detection in IoMT, IEEe J. Biomed. Health Inform. 27 (2023) 722–731, https://doi.org/10.1109/JBHI.2022.3186250.

[40] M. Kumar, S. Kim, Securing the internet of Health things: embedded federated learning-driven long short-term memory for cyberattack detection, Electr. (Switzerland) 13 (2024), https://doi.org/10.3390/electronics13173461.

[41] S.M.S. Bukhari, et al., Secure and privacy-preserving intrusion detection in wireless sensor networks: federated learning with SCNN-Bi-LSTM for enhanced reliability, Ad. Hoc. Netw. 155 (2024) 103407, https://doi.org/10.1016/j.adhoc.2024.103407.

[42] A. Wani, S. Revathi, R. Khaliq, SDN-based intrusion detection system for IoT using deep learning classifier (IDSIoT-SDL), CAAI. Trans. Intell. Technol. (2021), https://doi.org/10.1049/cit2.12003.

[43] R.A. Elsayed, R.A. Hamada, M.I. Abdalla, S.A. Elsaid, Securing IoT and SDN systems using deep-learning based automatic intrusion detection, Ain Shams Engineering Journal (2023), https://doi.org/10.1016/j.asej.2023.102211.

[44] G. Sri vidhya, R. Nagarajan, A novel bidirectional LSTM model for network intrusion detection in SDN-IoT network, Computing (2024), https://doi.org/10.1007/s00607-024-01295-w.

[45] J. Zhang, Y. Ling, X. Fu, X. Yang, G. Xiong, R. Zhang, Model of the intrusion detection system based on the integration of spatial-temporal features, Comput. Secur. 89 (2020) 101681, https://doi.org/10.1016/J.COSE.2019.101681.

[46] K. Jiang, W. Wang, A. Wang, H. Wu, Network intrusion detection combined hybrid sampling with deep hierarchical Network, IEEe Access. 8 (2020) 32464–32476, https://doi.org/10.1109/ACCESS.2020.2973730.

[47] P.R. Kanna, P. Santhi, Unified Deep learning approach for efficient intrusion detection system using integrated spatial–Temporal features, Knowl. Based. Syst. 226 (2021) 107132, https://doi.org/10.1016/J.KNOSYS.2021.107132.

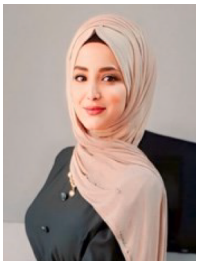
[48] T. Thilagam, R. Aruna, Intrusion detection for network based cloud computing by custom RC-NN and optimization, ICT Express 7 (4) (2021) 512–520, https://doi.org/10.1016/J.ICTE.2021.04.006.

[49] N. Niknami, J. Wu, DeepIDPS: an adaptive DRL-based intrusion detection and prevention system for SDN, in: IEEE International Conference on Communications, 2024, https://doi.org/10.1109/ICC51166.2024.10622849.

[50] A.U.H. Qureshi, H. Larijani, J. Ahmad, N. Mtetwa, A novel random neural network based approach for intrusion detection systems, in: 2018 10th Computer Science and Electronic Engineering Conference, 2019, pp. 50–55, https://doi.org/10.1109/CEEC.2018.8674228. CEEC 2018 - Proceedings.

[51] T.A. Tang, D. McLernon, L. Mhamdi, S.A.R. Zaidi, M. Ghogho, Intrusion detection in sdn-based networks: deep recurrent neural network approach. Advanced

Sciences and Technologies for Security Applications, Springer, 2019, pp. 175–195, https://doi.org/10.1007/978- 3-030-13057-2_8.
[52] W. Elmasry, A. Akbulut, A.H. Zaim, Evolving deep learning architectures for network intrusion detection using a double PSO metaheuristic, Comput. Netw. 168 (2020) 107042, https://doi.org/10.1016/J.COMNET.2019.107042.
[53] K. Narayana Rao, K. Venkata Rao, P.R. Prasad, A hybrid intrusion Detection system based on sparse autoencoder and Deep Neural Network, Comput. Commun. 180 (2021) 77–88, https://doi.org/10.1016/J.COMCOM.2021.08.026.
[54] P. Devan, N. Khare, An efficient XGBoost–DNN-based classification model for network intrusion detection system, Neural Comput. Appl. 32 (16) (2020) 12499–12514, https://doi.org/10.1007/S00521-020-04708-X/METRICS.
[55] Y. Yang, K. Zheng, C. Wu, Y. Yang, Improving the classification effectiveness of intrusion detection by using improved conditional variational AutoEncoder and deep neural network, Sensors 19 (2019) 2528, https://doi.org/10.3390/S19112528. Vol.Pagevol. 19, no. 11, p. 2528, Jun. 2019.
[56] J. Wu, Y. Peng, M. Song, M. Cui, and L. Zhang, Link congestion prediction using machine learning for Software-defined-network data plane, in IEEE CITS 2019: proceeding of the 2019 International Conference on Computer, Information and Telecommunication Systems, Beijing, China, 2019, pp. 1–5. doi: 10.1109/CITS.2019.8862098.
[57] T.E. Ali, Y.W. Chong, S. Manickam, Comparison of ML/DL approaches for detecting DDoS attacks in SDN, Appl. Sci. (Switzerland) 13 (5) (2023), https://doi.org/10.3390/app13053033 vol.no.
[58] N.S. Soud, N.A.S. Al-Jamali, Intelligent congestion control of 5G traffic in SDN using dual-spike neural network, J. Eng. 29 (1) (Jan. 2023) 110–127, https://doi.org/10.31026/j.eng.2023.01.07.
[59] J. Opitz, A closer look at classification evaluation metrics and a critical reflection of common evaluation practice, Trans. Assoc. Comput. Linguist. 12 (2024) 820–836, https://doi.org/10.1162/TACL_A_00675/122720/A-CLOSER-LOOK-AT-CLASSIFICATIONEVALUATION-METRICS.
[60] M. Grandini, E. Bagli, and G. Visani, "Metrics for multi-class classification: an overview," 2020, Accessed: May 10, 2025. [Online]. Available: https://arxiv.org/pdf/2008.05756.
[61] Y. Ho, S. Wookey, The real-world-weight cross-entropy loss function: modeling the costs of mislabeling, IEEe Access. 8 (2020) 4806–4813, https://doi.org/10.1109/ACCESS.2019.2962617.



Heba Dhirar received her B.S. in information and communication engineering, from Al Khwarizmi College of Engineering, Baghdad University, Iraq in 2019. Currently, she is a master's student in information and communication engineering at the University of Baghdad, Iraq. Her research interests include Computer Engineering, machine learning, and information security. Email: heba.d@kecbu.uobaghdad.edu.iq

Ali H. Hamad Faculty member of Information and communication engineering department university of Baghdad since 2003. He received his Ph. D in Control and systems engineering from the University of Basrah Iraq in 2015. M. Sc. and B. Sc. from the University of Technology Iraq in 2000 and 1997 respectively. Research interests include IoT, machine learning, information security, wireless sensor networks, and blockchain. Email: ahamad@kecbu.uobaghdad.edu.iq