SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V11I11P116

Volume 11 Issue 11, 179-195, November 2024
© 2024 Seventh Sense Research Group®

Review Article

The Role of Recent Datasets in Network Threat Classification and Intrusion Detection Systems

Priya Dasarwar

CSE, Symbiosis Institute of Technology, Nagpur Campus, Symbiosis International (Deemed University), Maharashtra, India.

Corresponding Author : priyadasarwar21@gmail.com

Received: 18 September 2024 Revised: 16 October 2024 Accepted: 15 November 2024 Published: 03 December 2024

Abstract *- As our reliance upon gadgets and technology grows, one of the most important challenges of the twenty-first century is developing safe networks, systems, and applications. The complexity of today's networks and services is growing, and with it, so too are the risks that individuals and businesses must manage. Researchers have created a variety of anomaly detection solutions in order to mitigate the impact of these threats; nevertheless, current methods find it difficult to keep updated with the constantly changing nature of modern architectures and associated threats, including zero-day attacks. This research addresses existing dataset weaknesses and research gaps and their implications for advancing Network Intrusion Detection Systems (NIDS) and the rise in complex attacks. For that purpose, the current paper presents researchers with a survey of well-known datasets UNSW-NB15, RPL NIDS-17 and N_BaIoT-18 and an analysis of their utilization, associated network hazards and various detection approaches. Current IDS research is highlighted in the manuscript.*

Keywords *- Data Set, Intrusion Detection, N_BaIoT-18, RPLNIDS-17, UNSW-NB15.*

1. Introduction

The Internet is transforming how people learn and work as it becomes more integrated into their daily lives, but it is also exposing us to ever-increasing security threats. The increasing reliance of the globe on networked actuators and sensors is impacting millions of people's lives. As a result, developing effective solutions to protect networks from security attacks is vital. The most difficult problem is detecting unidentified and disguised malware infections. Malware creation aims to undermine computer systems and exploit weaknesses in intrusion detection systems. Furthermore, safety concerns such as zero-day attacks aimed primarily at internet users have increased significantly [1]. Around the world, a large number of cybercriminals are enraged to steal information, illegally collect profits and find new targets. As a result, effective IDS that can identify sophisticated malware is essential. With a standard firewall, it is hard to identify various forms of malware as early as feasible. This is the aim of Intrusion Detection Systems (IDS). A recent, thorough taxonomy and evaluation related to this latest work is needed since machine learning has been utilized to improve intrusion detection during the last few decades. It is possible that with the help of deep learning technology, people can take advantage of more information, achieve more success, and realize their full potential. Its effects on society and the development of artificial intelligence are immense. AI has several applications beyond the three domains of image, sound, and behavior, including face identification, speech recognition, and robotics. Additionally, it shows how to use other notable cybersecurity technologies like intrusion detection and virus monitoring.

Early in the history of Artificial Intelligence (AI), employing Machine Learning (ML) technology was crucial in reducing cyberspace hazards. Despite its strength, machine learning depends a lot on feature extraction. This weakness is especially evident in the area of cybersecurity. Non-pre-defined features will not be recognized or found since machine learning algorithms only work with pre-defined features. One may claim that most machine learning algorithms' performance is determined by how well features are detected and extracted [2, 3]. Due to obvious issues with classical machine learning, researchers started focusing on Deep Neural Network (DNN), also called DL, a machine learning sub-domain. Many studies support the development of IDSs using either the DARPA 1999 or the KDD-Cup 99 datasets; nonetheless, the question of which deep learning techniques perform better remains unanswered. Despite being a crucial component of "online" intrusion detection systems' efficacy, the duration of time required to build intrusion detection systems is not considered when comparing different IDS approaches. The quality of the dataset and the selected learning model directly impact IDS efficacy. A dataset of outstanding quality can be characterized as enhancing real-world transaction performance metrics. Researchers have an issue with imbalanced datasets, as described in [4, 5]. When


This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

the distribution of classes is not uniform, a dataset is considered imbalanced [6]. Due to the datasets used, this can be a common issue in many classification situations. The utilized classifier biases towards the majority class in unbalanced datasets; nonetheless, most of them aim to find the minority class [7, 8]. Consequently, there is a significant classification error in the minority class samples, and significant goals might be missed. Datasets should be balanced based on data kinds to increase dataset quality.

The real-world usefulness of recent datasets for intrusion detection and network threat classification is diminished by problems such as a lack of diversity in attack tactics, protocols, and network topologies. Performance is hampered by issues like data imbalance, inadequate categorization, and static nature, particularly when it comes to new threats like APTs and zero-day exploits. Many datasets lack defined evaluation measures, cross-domain usability (such as cloud and IoT), and scalability. Improving intrusion detection systems requires filling these gaps using dynamic, extensive, and varied information [7].

The main objective of this research is to gather recent papers that offer a thorough summary of the data sets while also emphasizing the distinctive qualities of each data set. Particular emphasis was paid to the assault scenarios present in the data sets and how they interacted with one another. Every data set was also assessed in light of the characteristics of the classification scheme created in the first stage. The purpose of this long survey is to help researchers find IoT-based data sets that are pertinent to their objectives. The study of data sets conveys that the academic community has been observing a lack of publicly accessible network-based data sets. In recent years, it has tried to solve this lack by publishing a significant amount of data sets. This paper overviews the most recent datasets used in cybersecurity applications. This article aims to assist academics interested in learning more about network intrusion in the Internet of Things.

This research is presented in four sections, with the first being a summary introduction of the research and the second being the IDS datasets-related works section. In section 3, the research analysis of recent datasets is described. The conclusion and discussions of the data gathered are presented in Section 4.

2. Intrusion Detection Datasets Survey

Important datasets are summarized, and their shortcomings are mentioned in this section. In addition, recent IDSs are examined, focusing on the various attacks and machine learning approaches used on the datasets. Furthermore, throughout the last decade, changes in the algorithms utilized by research have been explored, revealing a definite shift in the employment of specific algorithms.

2.1. Intrusion Detection Datasets

To assess their findings, researchers used benchmark datasets. Nevertheless, real-world characteristics of recent network activity are absent from the datasets that are now accessible. Because of this, most anomalous intrusion detection systems should not be used in production environments. Moreover, IDS cannot respond to frequent network changes (such as the addition of new nodes, variations in traffic loads and network architecture, etc.). IDS cannot advance if only one relies on old datasets because networks are dynamic. It is important to consider this constant change characteristic when developing fresh datasets. The cost of building datasets from scratch will decrease by offering a standard dataset generation platform with expandable features that permit idea drift in network patterns.

Real-time network traffic can be captured to provide datasets or artificial traffic can be created through simulation. Synthetic attack injections can balance existing attack classes or introduce new attacks to an existing dataset. The following requirements must be met for a dataset to be taken into consideration, according to Viegas et al. [4]. (a) Genuine network traffic, similar to what is observed in production; (b) valid, as it encompasses all scenarios. (c) Labeled: every record is categorized as either normal or assault (d) varies, (e) is accurate and (f) is simply updated. (g) It is reproducible so researchers can compare information from different datasets and (h) shareable because it should not include private information. According to Anwer H.M. [12, 13], offering sufficient documentation for the feature along with the dataset collection environment constitutes a significant portion of the IDS dataset.

In the present study, two challenges that are relevant to research themes using synthetic datasets are identified. 1) Research in this field is limited since sharing datasets can be prohibited due to the sensitive nature of the data they contain. ii) The complexity of the requirements needed for the model to function well makes it challenging to simulate actual situations and the accompanying attacks. In contrast, this publication summarizes a list of the most widely utilized research and up-to-date datasets. The datasets that are currently accessible are arranged by domain in Table 1. Their advantages and disadvantages are also emphasized, including a few of the datasets mentioned earlier in [3]. A comprehensive description of NIDS datasets, including their main attributes, data format, anonymity, availability, size, recording environment, balancing, and more, is given by Hindy, Hanan et al. [1].

To help researchers make their own selections based on their use case and scenario, the authors present a list of datasets and the values corresponding to each criterion, as mentioned earlier. The authors look at attacks in the databases and offer a scientific comparison but don't go to great lengths about their research's wider implications. Furthermore, the

effort of analyzing and ranking datasets might produce unfair conclusions because of the scarcity of details supplementing the existing datasets. For example, it is better to have a dataset that appropriately depicts attack and background traffic than one that does not. But, since there isn't a fixed method for judging how realistic a generation is, this data isn't included in the dataset.

Table 1. Summary of existing datasets

Year of Creation	Data Set	Merits	Demerits
1998	DARPA	• The first standard for evaluating IDS in a simulated network environment. • Includes packet-based network activity spanning seven weeks. • Numerous threats, including port scanning, buffer overflows, DoS attacks, and rootkits, fall under this category.	• The models utilized for traffic generation were excessively basic. • Synthetic data cannot duplicate the background traffic present in real networks. • There's a lot of redundancy here.
1999	KDD99	• KDD99 is the most well-known and widely used. • Information is labelled, and each association is based on 41 highlights in addition to the label class. • Some types of attacks are Denial-of-Service, Remote-to-User, User-to-Root, and Probing. • • Provides a file containing network traffic (PCAP).	• KDD99 has been subjected to unbalanced arranging procedures. • The info is out of date. • •It is not intended for use with IoT frameworks.
2006	KYOTO	• The data was collected over three years. • It contains around 93 million sessions as a result of the exceptionally extended recording duration. • Ignored features with redundant information. • Represents actual real-world networks.	• Doesn't elaborate on particular attack types.
2009	NSL KDD	• It overcomes KDD99 constraints. An enhanced version of dataset KDD99 is called NSL-KDD. • There are no duplicate records in the test or preparation sets.	• Currently, there are no low-impression assault instances. • It is not intended for use with IoT frameworks.
2015	UNSW-NB15	• It provides a blend of real-world workouts and engineered materials for modern assault techniques. • Records network traffic in PCAP and CSV formats. • • Included are nine different sorts of attacks: analysis, fuzzers, backdoors, exploits, denial-of-service, reconnaissance, generic, worms, and shellcode.	• It is prone to the problem of high-class imbalance, which could lead to low accuracy and a high system failure rate.
2017	CICIDS2017	• Provides network traffic (PCAP) and CSV files. • Contains labelled information for artificial intelligence purposes. • DDoS, DoS, Brute Force FTP, Brute Force SSH, Infiltration, Heartbleed, Web Attack, and Botnet are some implemented attacks.	• This information is not available to the general public. • Incompatible with IoT frameworks.

2018	CICIDS2018	• It contains labelled data. • Records network traffic in PCAP and CSV formats. • Uses web, botnet, DDoS, Brute-force, Heartbleed, and DDoS attacks to penetrate local networks. • Dataset that is generated dynamically. • It's adaptable, expandable, and repeatable.	• This information is not available to the general public. • Incompatible with IoT frameworks.
2017	RPL NIDS17	• Network flows with labels using binary and multiclass • Based on 6LoWPAN networks with RPL protocol • Used for machine learning and deep learning. • Provides 21 features and 1 label in network traffic (PCAP), log, and CSV files. • The collection contains regular records and seven different types of attacks: clone-ID, hello flooding, sinkhole, blackhole, selective forwarding, local repair, and Sybil.	• This information is not available to the general public. • RPL protocol is the only one that can be used.
2018	N_BaIoT	• It is predicated on real-time data flow collected from commercial IoT devices. • Provides CSV files with 115 features and 1 label, as well as network traffic (PCAP). • Network flows that are labeled with binary and multiclass. • For deep learning and machine learning applications. • It comprises five Mirai attacks (UDPplain, UDP, Syn, Ack, and scan) and five BASHLITE assaults (COMBO, TCP, UDP, junk, and scan).	• Suitable for commercial devices alone, not for a wide IoT-based WSN

2.2. Discussion on Recent IDS Datasets: UNSW-NB15, RPL-NIDS-17 and N_BaIoT-18

Network attack behavioral patterns change over time, so updating standard datasets in a dynamic environment is necessary. This will facilitate the manifestation of diverse network traffic conditions and easily adaptable, redefining, and learning attack strategies. Choosing a suitable dataset is also an important task. While some datasets are made available to the public for research purposes, others may contain records that are out of date due to technological constraints. Specific groups develop these datasets for their own objectives.

The lack of an acceptable dataset is a problem that needs to be considered because many publicly available datasets are statistically insufficient. A few key factors were determined to build and analyze the framework of IDS datasets to produce a comprehensive and effective IDS dataset. These attributes include heterogeneity, feature set, labeled data samples, diversity of attacks, anonymity, available protocols, gathering all network traffic, capturing all network interactions, defining all network setups, and metadata. Several considerations are made when creating the UNSW-NB15, RPL NIDS-17, and N BaIoT-18 datasets.

The idea of profiles was used in the creation of these well-organized datasets. Each of these datasets demonstrates a conceptual understanding of the different application models, network devices, and protocols and a thorough understanding of the attacks conducted. Network data was recorded using the packet sniffer tool [22], and individual instances were classified as attack or normal. For every attack, packet captures were stored in separate CSV files. The complete dataset is then created by combining all of the CSV files.

2.3. Characteristics of the Dataset

Several scholars have expressed interest in using the UNSW-NB15 [5-37], RPL NIDS-17 [38-70], and N_ BaIoT 18 [71-102] datasets to create different classifiers. Table 1 displays the datasets' specifications. The dataset's files are used for multi-class and binary classification. The dataset's files should be combined so that there is a wide variety of attack categories in order to create an effective IDS [42]. The ability to precisely

detect every kind of attack is what defines an optimal intrusion detection system. Moreover, Table 2 lists the eleven characteristics of an ideal dataset employed in creating these datasets, as stated in [43].

Table 2. Specifications about UNSW-NB15, RPL NIDS-17 and N_BaIoT-18 Datasets

Parameters		Data Sets		
		UNSW- NB15	**RPL NIDS - 17**	**N_BaIoT-18**
Overview	Year of Traffic Creation	2015	2017	2018
	Public Available	Yes	No	Yes
	Normal Traffic	Yes	Yes	Yes
	Attack Traffic	Yes	Yes	Yes
	No. of Features	49	20	115
	Total No of Records	2,57,673	2,26,547	70,62,606
Type of Data	Meta data	Yes	Yes	Yes
	Format	Packet, Other	csv	csv
	Anonymity	None	None	None
Volume of Data	Count	2M Points	2M Points	5M Points
	Duration	31 Hours	Not specified	Not specified
Environment for Recording	Kind of Traffic	Emulated	Synthetic	Real
	Type of Network	Small Network	Small Network	Small Network
	Compl. Network	Yes	Yes	Yes
Assessment	Predef. Splits	Yes	Yes	Yes
	Balanced	No	No	No
	Label	Yes	Yes	Yes
Attack Categories	Number of Attacks	09	07	02

Table 3. Qualities for creating an optimal dataset

Characteristic	Description
Attacks	The dataset should include a diverse set of up-to-date attack categories.
Anonymity	The dataset should contain information from both the packet header and payload.
Capturing the Traffic	It describes the process of gathering both functional and non-functional network traffic to ascertain the DR and FPR of the IDS.
Features	The dataset must have all of the features that are well-defined in order to classify the attack.
Heterogeneity	To include all the specifics of the process used to identify the attacks, the dataset should be gathered from various sources.
Labeled Dataset	It describes the act of labeling data instances that are gathered from network traffic in order to achieve a comprehensive understanding of the network's interactions.
Metadata	A dataset should include detailed descriptions of the testing environment, the infrastructure of the attack system, the infrastructure of the victim system, and the attack scenario.
Network Configuration	Capturing real-world attack scenarios means deeply grasping the network topology and how networking devices are connected in the testing environment.
Network Traffic	It refers to capturing all network packets from the host, destination, firewall, and web applications for the purposes of flow analysis and dataset generation.
Network Interaction	It means keeping a thorough log of every network communication within and outside the network.
Protocols	Any beneficial and detrimental communication using a variety of protocols would be included in an ideal dataset.

2.4. Attacks Related to Datasets

A broad and modular taxonomy of security risks is necessary to assist academics and cybersecurity experts in developing tools to recognize a range of assaults, from well-known to zero-day ones. There are many different attack scenarios in IDS data sets. This feature indicates whether or not a data set contains hostile network activity; it is set to yes if the data set contains at least one assault. These data sets blend real, current, and regular network traffic with fake, artificial, and modern network traffic assault operations [1, 41 61]. Table 3 lists the exact attack types linked to UNSW-NB15, RPL NIDS-17, and N BaIoT-18 datasets.

Table 4. Attacks related to datasets

DataSet	Attack	Description
UNSW NB-15	Fuzzers	An attack is when the attacker temporarily suspends or crashes the operating system, software, or network in an effort to find security holes.
	Analysis	Attacks that compromise web applications by a variety of techniques, such as spam email distribution, malicious web scripting, and port scanning.
	Backdoors	An attacker can use this method to get around standard authentication and get unauthorized remote access to a system.
	DoS	An intrusion is when the attacker attempts to overburden computer resources by making them extremely busy to prevent authorized access to the resources.
	Exploits	Attacks that profit from defects, malfunctions, or software in Operating Systems (OS) or applications.
	Generic	This attack aims to decode the security system's key, targeting a cryptographic system.
	Reconnaissance	A probe is an attack designed to get beyond security measures on a target computer network by gathering information about it.
	Shellcode	In a malware attack, the attacker gains control of the compromised system by breaking through a small code segment that originates from a shell.
	Worm	Depending on the security flaws on the target computer it wants to access, malware can replicate itself and propagate to other computers by using the network.
RPL NIDS 17	Clone ID	An attack where a malicious node is created by copying the ID of one logical node to another. Data that was intended to reach the victim node instead ends up at the malicious node. As a result, the attacker now controls a sizable chunk of the wireless network.
	Hello Flooding	A malevolent node attempts to convince each node that it is its neighbor by sending out high-quality route information, such as favorable routing metrics (rank), to other nodes in the network.
	Local Repair	This attack is carried out through poisoning. The rogue node broadcasts a message to the whole network and sets its rank to infinite. To get to the root (gateway) node, other valid nodes now need to look for the new parent. When this occurs frequently, the network performs worse because the topology has to be adjusted every time a node changes.
	Selective Forwarding	The rogue node drops packets in a selective forwarding attack, one packet at a time. In an Internet of Things environment, malicious nodes may discard data packets while carrying forwarding control packets, enabling them to remain undetected.
	Sink hole	An attack is when a hostile node sends out favorable routing information, directing traffic towards it from other nodes.
	Black hole	A routing attack wherein every packet that arrives at the malicious node is dropped so that the packet's true sender is not informed that it did not make it to its intended recipient.
	Sybil	A kind of infiltration occurs when the ID of one node is similar to multiple nodes. This method allows a single node to get information from a large network. This assault causes a rapid degradation in the system's performance.
N_BaIoT 18	Mirai Ack	Flooding of Ack
	Mirai Scan	Scan for vulnerable devices automatically.
	Mirai Syn	Flooding of Syn
	Mirai UDP	Flooding of UDP
	Mirai Plain UDP	UDP flooding with fewer options, optimized for higher PPS
	Bashlite Combo	Spam material is sent, and a connection is established to a certain IP address and port.
	Bashlite Junk	Spam data transmission
	Bashlite Scan	Looking for susceptible devices on the network
	Bashlite TCP	Flooding of TCP
	Bashlite UDP	Flooding of UDP

2.5. Machine Learning Techniques for IDS

Table 5. Summary of proposed implementation of machine learning techniques

Author/Reference	Data Set	Techniques Used	Feature Selection	Performance Parameters	Classification
Chowdhury et al. [7]	UNSW NB-15	Support Vector Machine	Simulated Anneling	Accuracy98.76%, FPR 0.09%, FNR 1.35%	Binary
Anwer et al. [12]	UNSW NB-15	J48 and Naive Bayes	Filter and Wrapper Method	Accuracy 88%	Binary
Idham mad et al. [14]	UNSW NB-15	Feed-forward Neural Network	Correlation-based Feature Selection	Accuracy 97.1%	Binary
Hajisa lem et al. [15]	UNSW NB-15	Artificial Bee Colony (ABC) Artificial Fish Swarm (AFS) algorithms	Correlation-based Feature Selection Fuzzy C-Means Clustering	DR 0.13% Accuracy 98.6% FPR 98.9%	Binary
Guha et al. [16]	UNSW NB-15	Artificial Neural Network	Genetic algorithm	Accuracy.95.46	Binary
Kamar udin et al. [17]	UNSW NB-15	Ensemble classifier (Logitboost & Random Forest)	Filter and Wrapper	DR 99.10% Accuracy 99.45% FAR 0.18%	Binary
Nahiy an [22]	UNSW NB-15	K-means	Statistical Techniques	Recall 92% Precision 91% F1 –score 91%	Binary
Moust afa et al. [24]	UNSW NB-15	AdaBoost, Decision Tree (DT), Naive Bayes (NB) Artificial Neural Network (ANN)	Correlation Coefficient	Accuracy 99.54%, DR 98.93%, FPR 1.38%	Binary
Verma, Abhishek et al. [38]	RPL NIDS-17	Naive Bayes (NB) Decision Tree (DT) Logistic Regression (LR) Artificial Neural Networks Expectation-Maximization Clustering	Correlation analysis: Pearson's correlation coefficient technique and Gain ratio technique.	Accuracy 85.14% FAR 21.65%	Binary
Verma, Abhishek et al. [39]	RPL NIDS-17	Ensemble Classifiers (Boosted Trees, Bagged Trees, Subspace Discriminant and RUSBoosted Trees)	Principal Component Analysis (PCA)	Accuracy 94.5%, AUC.0.96%	Binary
P. Jaya Prakash et al. [40]	RPL NIDS-17	Ensemble Classifiers (Support Vector Machine, Decision Trees, K-Nearest Neighbour, Logistic Regression)	Novel feature selection technique (SA-improved SSA)	Accuracy 0.88 Precision 0.69 ADR 0.79 F-measure 0.73 Specificity 0.91 FAR 0.088	Binary
Musa Osman, Jingsha He, Fawaz Mahiuob Mohammed Mokbal and Nafei Zhu [42]	RPL NIDS-17	Artificial Neural Network (ANN) model	Random Forest Classifier	Accuracy 97.14 % Precision 97.03%, False- positive rate 0.36% AUC-ROC 8%	Multiclass

Foley, John et al.[45]	RPL NIDS-17	Multilayer Perceptron and Random Forest	Sampling	RSME 0.19 MAPE 21.38 ROC 0.97 DR 87.08 %	Binary
Verma, Abhishek and V. Ranga. [43]	RPL NIDS-17	Ensemble learning	Random Search Algorithm	Accuracy 0.98 Specificity 0.95 Sensitivity 0.94 FPR 0.85 AUC 0.005	Binary
Napiah, Mohamad Nazrin et al. [57]	RPL NIDS-17	CHA-IDS (Multi-agent system framework)	Correlation-based Feature Selection	TPR 99% FPR 0.002	Multiclass
Kfoury, Elie F. et al. [59]	RPL NIDS-17	Neural Network	Feature Scaling	U-Matrix (unified distance matrix)	Binary
Meidan, Yair et al. [76]	N_BaIoT	Deep Autoencoders	Statistical technique	Accuracy 80%	Binary
Joshi, Shreehar and Eman Abdelfattah. [87]	N_BaIoT	Decision Trees, Extra Trees Classifiers, Random Forests, and Support Vector Machines	Random Selection	Precision 85 % Recall 84 % F-score 85 %	Binary
Kim, Jiyeon et al. [88]	N_BaIoT	RNN and LSTM Model	Random Selection	Accuracy 96%	Binary
Kim, Ji Yeon et al. [89]	N_BaIoT	ML (Naïve BAYES (NB), K-Nearest Neighbors (KNN), Logistic Regression (LR), Decision Tree (DT) and RANDOM FOREST (RF))) and DL (Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Long Short-Term Memory (LSTM)) Model	Random Selection	F1-score 0.99	Multiclass
Desai, Madhuri Gurunathrao et al. [90]	N_BaIoT	Decision Tree (DT), Random Forest Classifier (RFC), and Support Vector Machines (SVM)	Principle Component Analysis	Accuracy 99%	Multiclass
Alqahtani, Mnahi et al. [91]	N_BaIoT	Genetic Algorithm (GA) Extreme Gradient Boosting (XGBoost)	Fisher-score-based Feature Selection	Accuracy 99.96%	Multiclass
Alkahtani, Hasan and Theyazn H. H. Aldhyani. .[94]	N_BaIoT	Convolutional neural network and long short-term memory	Random Selection	Accuracy 89%	Multiclass
Bagui, Sikha et al. [98]	N_BaIoT	Logistic Regression, Support-Vector Machine and Random Forest	Random Selection	Accuracy 99% Precision 99% Recall 99% F1-Score 99%	Multiclass

This section examines various machine learning-based intrusion detection algorithms in detail using the most recent benchmark datasets, UNSW-NB15, RPL NIDS-17, and N BaIoT-18. Several benefits of IDS that rely on machine learning include the following:

- IDS based on machine learning that uses supervised algorithms can detect attack variants in real time as they watch traffic flow behaviour.
- New threats can be identified by machine learning-based intrusion detection systems that use unsupervised learning

methods.

- The machine learning-based IDS has a low to moderate CPU load.
- IDSs based on machine learning can identify the intricate aspects of an assault. Additionally, it increases detecting speed and accuracy.

A wide range of attacks are still evolving. Database modifications are not required for IDS based on machine learning that uses clustering and outlier detection. Many academics have conducted in-depth analyses of various machine learning techniques, either with or without feature selection, to detect intrusive behavior.

Their suggested intrusion detection techniques have different characteristics and provide different outcomes. Current research indicates that no single intrusion detection method can identify every type of attack. As a result, using a specific intrusion detection technology is recommended to detect a specific set of attacks. Table 5 summarises various intrusion detection algorithms based on current datasets.

3. IDS Datasets Analysis

This section provides an overview of recent ML IDS and a critique of the latest datasets' flaws. The terms "Intrusion Detection System*" OR "IDS*" were used to filter the dates so that past articles from IEEE Xplore and Google Scholar searches were included. Various datasets, machine learning techniques, and known attacks were filtered out.

Between 2011 and 2021, a total of 99 published articles were examined. Table 4 presents a summary of the most widely cited IDS for UNSW-NB15 [5-37], RPL NIDS-17 [38-70], and N BaIoT-18 [71-102] datasets during the last decade. Each IDS is described in detail, including a description of the methods used and the datasets against which it was tested. In addition, the assaults that have been detected are listed. Figure 1 shows the distribution of current datasets used in literature-based research. Because these datasets include representations of contemporary network traffic and attack scenarios, they have been widely utilized for assessing intrusion detection systems for Internet of Things networks. UNSW-NB15 is the dataset of choice, as seen in Figure 1. The limitations of the earlier datasets are thoroughly examined by the authors in [1, 3].

The results of the trials indicate that UNSW-NB15 is a more complex dataset and deserves to be employed as a replacement benchmark for NIDS assessment. [7, 12, 14, 15, -17, 22, 24, 38-40, 42, 43, 45, 57, 59, 76, 87-91, 94, 98] all discuss the analysis and evaluation of the UNSW-NB15, RPL NIDS-17, and N BaIoT-18 data sets.

Figure 2 visualizes the attack types and number of instances of attacks for UNSW-NB15, RPL NIDS-17, and N BaIoT-18 data sets presented in Table 3. [5, 54, 76] UNSW-NB15 [5-37], RPL NIDS-17 [38-70], and N BaIoT-18 [71-102] summarized the attacks in these datasets. Figure 3 depicts the analysis of binary and multi-class classification done on IDS by the researchers in the literature.

For the UNSW-NB15 dataset, around 37% of papers have demonstrated binary classification of IDS where, and only 11% of work is done on multi-class classification. In the RPL NIDS-17 dataset,38% and 3% of papers have mentioned binary and multi-class classification, respectively. Finally, for the N_BaIoT-18 dataset, 4% of research papers include binary classification, and 7% of research papers perform multi-class classification.

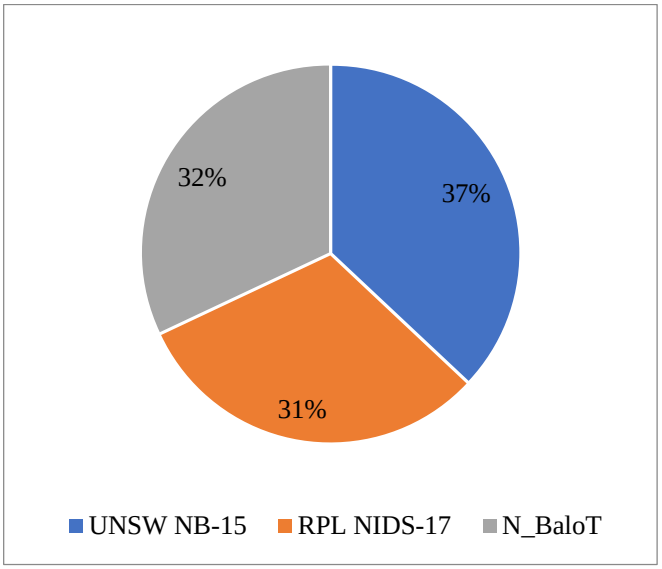

Fig. 1 Distribution of recent datasets used for IDS evaluation

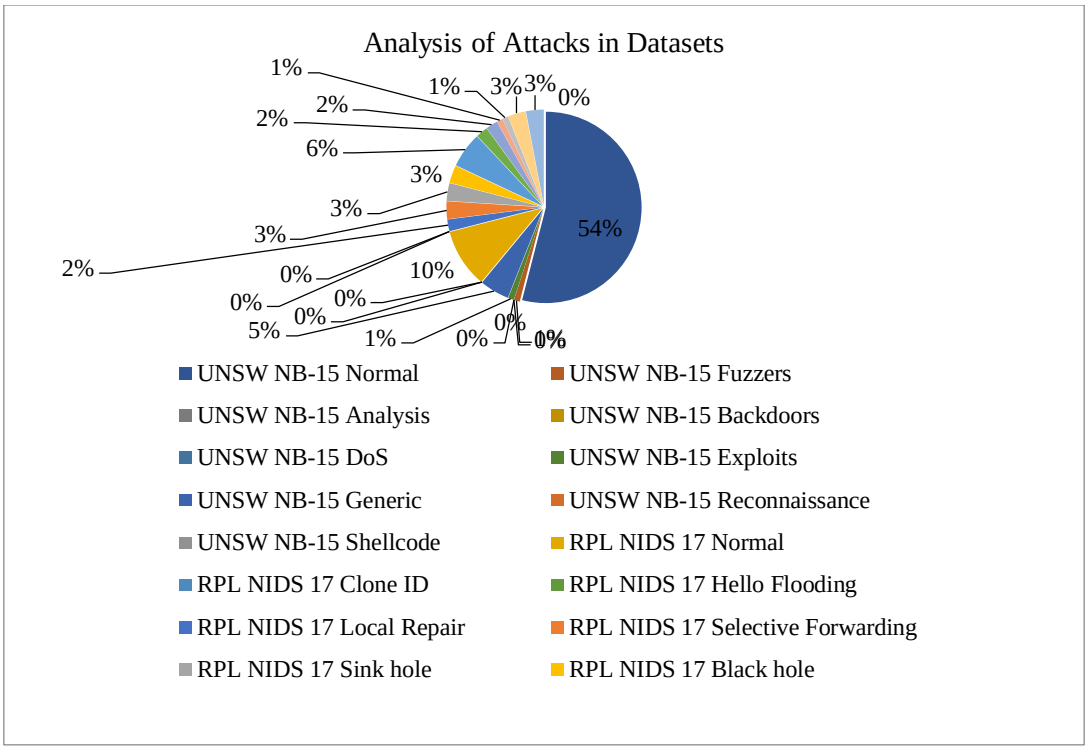

Fig. 2 Distribution of attacks for recent datasets

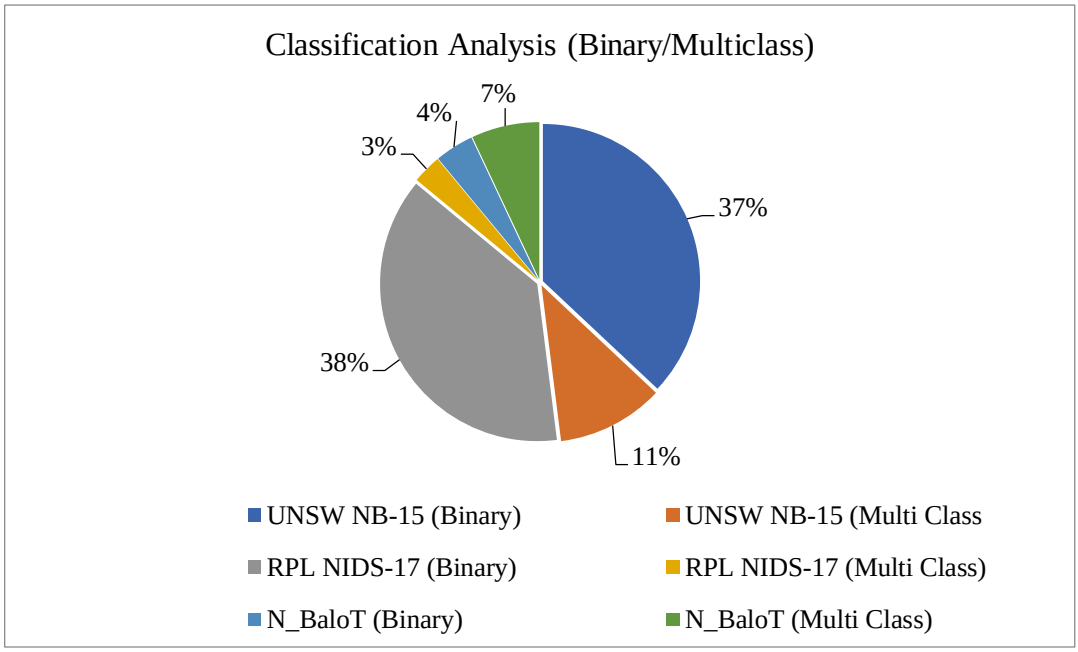

Fig. 3 Analysis of binary and multi-class classification

It is crucial to consider the algorithms employed in earlier IDS research before proceeding. Anomaly-based IDSs work by seeing patterns that distinguish between normal and anomalous traffic. When creating an IDS, Figure 3 shows how ML methods dominate for the UNSW-NB15, RPL NIDS-17, and N BaIoT-18 data sets. Statistical and knowledge-based algorithms are both underrepresented, as demonstrated in the graph. This supremacy is due to the widespread usage of machine learning techniques in various academic fields. Figures 4, 5 and 6 show the dispersion of machine learning algorithms utilised by the IDSs for the datasets UNSW-NB15, RPL NIDS-17, and N BaIoT-18. The ability of Naive Bayes, Decision Trees, SVM, and Neural Networks to discriminate between benign and attack classes given a feature set explains their popularity as the most

commonly employed algorithms. However, it is worth noting that using new machine learning techniques and adopting ones from other domains will help progress the development of IDSs in the coming decade.

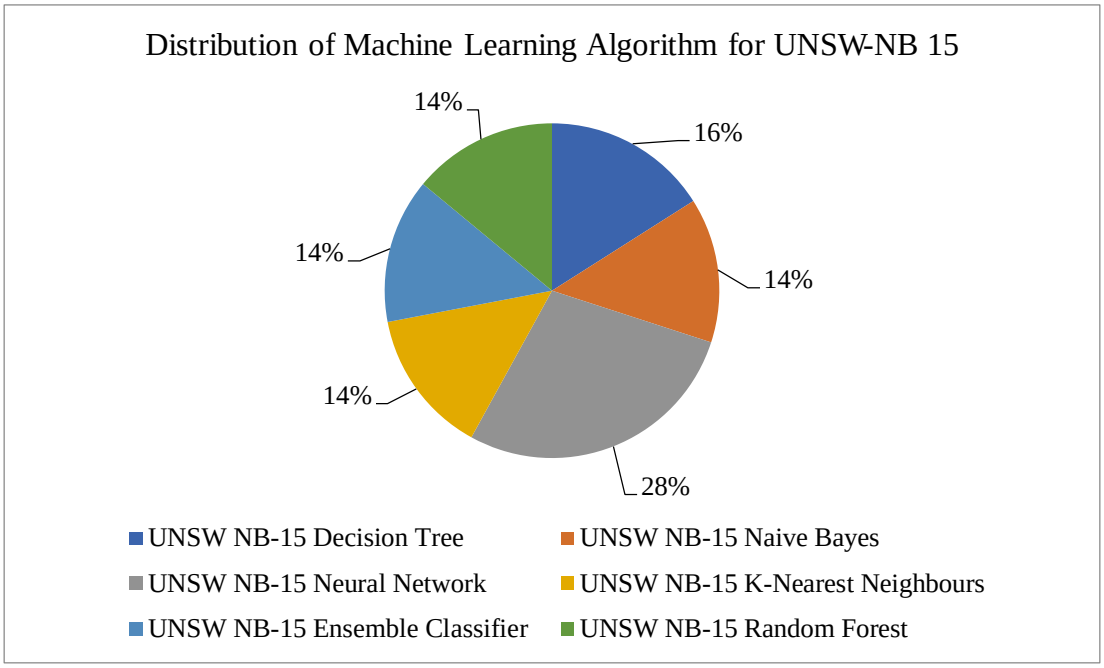

Fig. 4 Distribution of ML algorithms for UNSW NB-15 dataset

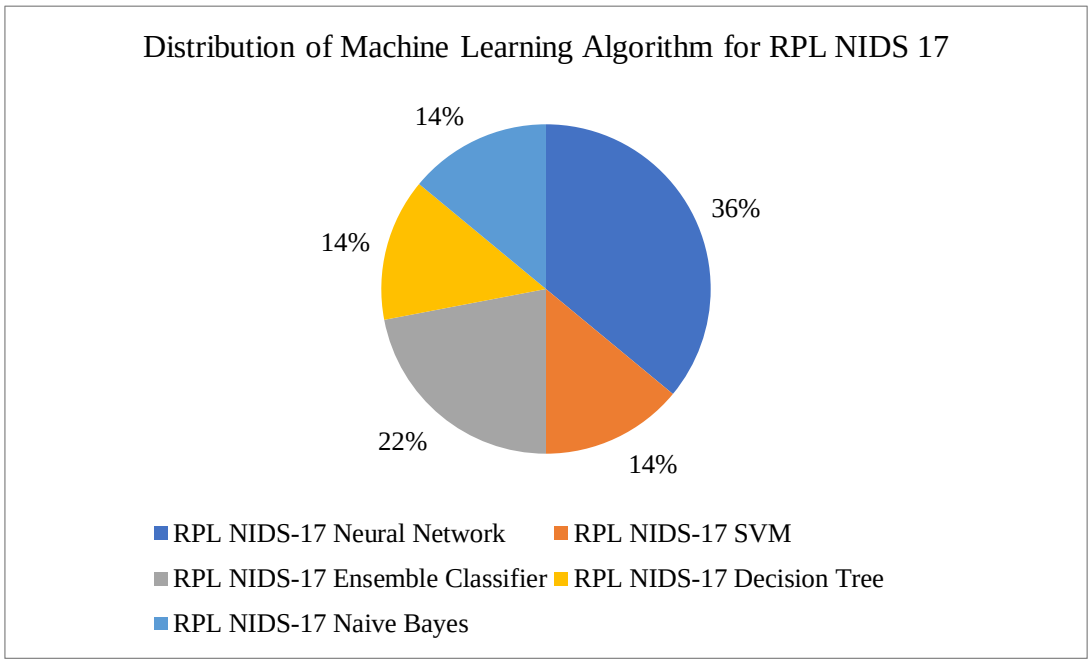

Fig. 5 Distribution of ML algorithms for RPL NIDS 17 dataset

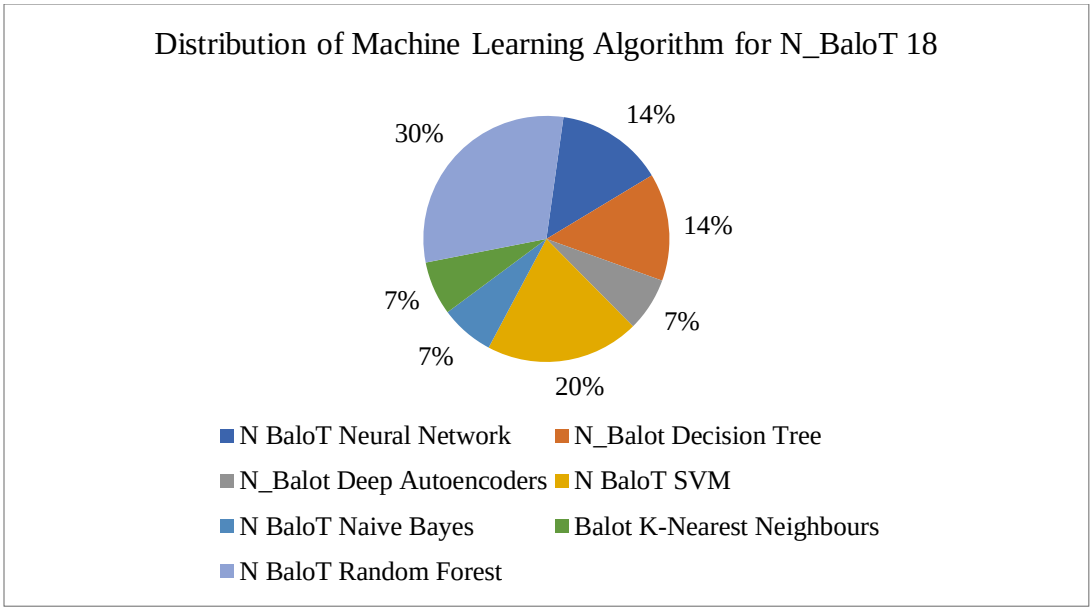

Fig. 6 Distribution of ML algorithms for N_BaIoT-18 dataset

Table 6. Strengths and weaknesses of recent datasets

Dataset	Strengths	Weaknesses
UNSW-NB15	Current traffic, a variety of attack methods, intricate features, and widespread use	Class disparity, superfluous features, and inadequate representation of IoT traffic
RPL-NIDS-17	LLN emphasis, IoT-specific attacks, and labeled data	Fewer features at the packet level, limited scalability, and restricted scope
N-BaIoT-18	Device-specific, real-world scenarios and IoT botnets	Botnet detection only, Controlled environment only, and no encrypted traffic

Table 6 contrasts three datasets that are utilized in studies on intrusion detection. N-BaIoT-18 is perfect for IoT botnet detection but lacks adaptability for other applications; RPL-NIDS-17 concentrates on IoT-specific attacks but has a narrow scope; and UNSW-NB15 is excellent at current traffic variety but lacks IoT relevance. Intrusion Detection System (IDS) research frequently uses the UNSW-NB15, RPL-NIDS-17, and N-BaIoT-18 datasets, each of which has unique advantages and disadvantages. UNSW-NB15 is a widely used benchmark because it incorporates 49 detailed features, various actual and synthetic data, and contemporary network traffic with various attack techniques (such as DoS and fuzzing). It has limitations regarding IoT traffic representation, redundant features, and class imbalance.

IoT-specific risks are the focus of RPL-NIDS-17, which employs labeled data for supervised learning and is especially effective in Low-power and Lossy Networks (LLNs). However, it lacks diversity, is not scalable, and provides fewer packet-level features. Although it offers real-world device-specific traffic and behavioral data and is perfect for IoT botnet detection, N-BaIoT-18 is limited to botnet attacks and controlled conditions and does not represent encrypted traffic. The suitability of each dataset depends on how well its features match particular research objectives. The intended use will determine which of these datasets is best. While RPL-NIDS-17 concentrates on IoT-specific threats in LLNs, UNSW-NB15 is well-suited for general-purpose intrusion detection with various attack types. Detecting IoT botnets is where N-BaIoT-18 shines, but it is less effective against other kinds of network threats. Aligning the dataset's features with the particular needs of the study or application is necessary when choosing one.

4. Conclusion

The study examines the shortcomings in datasets produced by the Intrusion Detection Systems (IDS) industry. The effectiveness of the ML-based IDS was evaluated using these datasets. The results show that updating the underlying dataset is necessary to detect new attacks in the field of enhanced performance intrusion detection systems. This is because attackers employ various procedures and technological tools to execute their attacks.

Moreover, the multiple assault pattern duplicates the need for datasets with real-world network circumstances. In order to fulfill the need to develop an intrusion detection dataset with realistic network traffic and updated network attacks, the UNSW-NB15, RPL NIDS-17, and N BaIoT-18

datasets have been introduced. This study looks at some of these datasets' shortcomings and their characteristics. Examined is the most well-known IDS study covered in the literature. The study yields three main results. Researchers' usage of recent benchmark datasets is highlighted in the first section. Second, we review the binary and multi-class classification work and provide a taxonomy of threats observed in datasets. Lastly, we go over the many machine learning algorithms that have been used to assess IDS.

References

[1] Hanan Hindy et al., "A Taxonomy of Network Threats and the Effect of Current Datasets on Intrusion Detection Systems," *IEEE Access*, vol. 8, pp. 104650-104675, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[2] Markus Ring et al., "A Survey of Network-based Intrusion Detection Data Sets," *Computers & Security*, vol. 86, pp. 147-167, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[3] Priya R. Maidamwar, Mahip M. Bartere, and Prasad P. Lokulwar, *Intrusion Detection Systems in IoT: Techniques, Datasets, and Challenges*, 1st ed., Computing Technologies and Applications, pp. 1-40, 2021. [Google Scholar] [Publisher Link]

[4] Eduardo K. Viegas, Altair O. Santin, and Luiz S. Oliveira, "Toward A Reliable Anomaly-Based Intrusion Detection in Real-World Environments," *Computer Networks*, vol. 127, pp. 200-216, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[5] Nour Moustafa, and Jill Slay, "A Hybrid Feature Selection for Network Intrusion Detection Systems: Central Points," *arXiv*, pp. 5-13, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[6] Hossein Gharaee, and Hamid Hosseinvand, "A New Feature Selection IDS Based on Genetic Algorithm and SVM," *8th International Symposium on Telecommunications*, Tehran, Iran, pp. 139-144, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[7] Chowdhury Md. Nasimuzzaman, Ken Ferens, and Mike Ferens, "Network Intrusion Detection using Machine Learning," *Proceedings of the International Conference on Security and Management*, pp. 30-35, 2016. [Google Scholar] [Publisher Link]

[8] Deval Bhamare et al., "Feasibility of Supervised Machine Learning for Cloud Security," *International Conference on Information Science and Security*, Pattaya, Thailand, pp. 1-5, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[9] Mirza M. Baig, Mian M. Awais, and El-Sayed M. El-Alfy, "A Multiclass Cascade of Artificial Neural Network for Network Intrusion Detection," *Journal of Intelligent & Fuzzy Systems*, vol. 32, no. 4, pp. 2875-2883, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[10] Mustapha Belouch, Salah El Hadaj, and Mohamed Idhammad, "A Two-Stage Classifier Approach using RepTree Algorithm for Network Intrusion Detection," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 6, pp. 389-394, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[11] Malek Al-Zewairi, Sufyan Almajali, and Arafat Awajan, "Experimental Evaluation of a Multi-Layer Feed-Forward Artificial Neural Network Classifier for Network Intrusion Detection System," *IEEE International Conference on New Trends in Computing Sciences*, Amman, Jordan, pp. 167-172, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[12] Hebatallah Mostafa Anwer, Mohamed Farouk, and Ayman Abdel-Hami, "A Framework for Efficient Network Anomaly Intrusion Detection with Features Selection," *9th International Conference on Information and Communication Systems*, Irbid, Jordan, pp. 157-162, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[13] M.A. Mithun Aravind, and V.K.G. Kalaiselvi, "Design of an Intrusion Detection System Based on Distance Feature Using Ensemble Classifier," *4th International Conference on Signal Processing, Communications and Networking*, Chennai, India, pp. 1-6, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[14] Mohamed Idhammad, Karim Afdel, and Mustapha Belouch, "DoS Detection Method based on Artificial Neural Networks," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 4, pp. 465-471, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[15] Vajiheh Hajisalem, and Shahram Babaie, "A Hybrid Intrusion Detection System Based on ABC-AFS Algorithm for Misuse and Anomaly Detection," *Computer Network*, vol. 136, pp. 37-50, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[16] Sayantan Guha, Stephen S. Yau, and Arun Balaji Buduru, "Attack Detection in Cloud Infrastructures Using Artificial Neural Network with Genetic Feature Selection," *IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th International Conference on Pervasive Intelligence and Computing, 2nd International Conference on Big Data Intelligence and Computing and Cyber Science and Technology Congress*, Auckland, New Zealand, pp. 414-419, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[17] Muhammad Hilmi Kamarudin et al., "A LogitBoost-Based Algorithm for Detecting Known and Unknown Web Attacks," *IEEE Access*, vol. 5, pp. 26190-26200, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[18] Nour Moustafa, Jill Slay, and Gideon Creech, "Novel Geometric Area Analysis Technique for Anomaly Detection using Trapezoidal Area Estimation on Large-scale Networks," *IEEE Transactions on Big Data*, vol. 5, no. 4, pp. 481-494, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[19] Khoi Khac Nguyen et al., "Cyberattack Detection in Mobile Cloud Computing: A Deep Learning Approach," *IEEE Wireless Communications and Networking Conference*, Barcelona, Spain, pp. 1-6, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[20] Rifkie Primartha, and Bayu Adhi Tama, "Anomaly Detection using Random Forest: A Performance Revisited," *International Conference on Data and Software Engineering*, Palembang, Indonesia, pp. 1-6, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[21] Sana Siddiqui, Muhammad Salman Khan, and Ken Ferens, "Multiscale Hebbian Neural Network for Cyber Threat Detection," *International Joint Conference on Neural Networks*, Anchorage, AK, USA, pp. 1427-1434, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[22] K. Nahiyan et al., "A Multi-agent Based Cognitive Approach to Unsupervised Feature Extraction and Classification for Network Intrusion Detection," *International Conference on Advances on Applied Cognitive Computing*, 2017. [Google Scholar]

[23] Bipraneel Roy, and Hon Cheung, "A Deep Learning Approach for Intrusion Detection in Internet of Things using Bi-Directional Long Short- Term Memory Recurrent Neural Network," *28th International Telecommunication Network and Applications Conference*, Sydney, NSW, Australia, pp. 1-6, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[24] Nour Moustafa, Benjamin Turnbull, and Kim-Kwang Raymond Choo, "An Ensemble Intrusion Detection Technique based on proposed Statistical Flow Features for Protecting Network Traffic of Internet of Things," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4815-4830, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[25] Bayu Adhi Tama, and Kyung-Hyune Rhee, "Attack Classification Analysis of IoT Network via Deep Learning Approach," *Research Briefs on Information & Communication Technology Evolution*, vol. 3, pp. 150-158, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[26] Mustapha Belouch, Salah El Hadaj, and Mohamed Idhammad, "Performance Evaluation of Intrusion Detection Based on Machine Learning Using Apache Spark," *Procedia Computer Science*, vol. 127, pp. 1-6, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[27] Yiyun Zhou et al., "Deep Learning Approach for Cyberattack Detection," *IEEE Conference on Computer Communications Workshops*, Honolulu, HI, USA, pp. 262-267, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[28] Nour Moustafa et al., "Collaborative Anomaly Detection Framework for handling Big Data of Cloud Computing," *Military Communications and Information Systems Conference*, Canberra, ACT, Australia, pp. 1-6, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[29] Muna AL-Hawawreh, Nour Moustafa, and Elena Sitnikova, "Identification of Malicious Activities in Industrial Internet of Things Based on Deep Learning Models," *Journal of Information Security and Application*, vol. 41, pp. 1-11, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[30] Nour Moustafa et al., "A New Threat Intelligence Scheme for Safeguarding Industry 4.0 Systems," *IEEE Access*, vol. 6, pp. 32910-32924, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[31] Valentina Timčenko, and Slavko Gajin, "Ensemble Classifiers for Supervised Anomaly Based Network Intrusion Detection," *13th IEEE International Conference on Intelligent Computer Communication and Processing*, Cluj-Napoca, Romania, pp. 13-19, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[32] Nour Moustafa, Gaurav Misra, and Jill Slay, "Generalized Outlier Gaussian Mixture technique based on Automated Association Features for Simulating and Detecting Web Application Attacks," *IEEE Transactions on Sustainable Computing*, vol. 6, no. 2, pp. 245-256, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[33] Tian Yingjie et al., "Ramp Loss One-Class Support Vector Machine; A Robust and Effective Approach To Anomaly Detection Problems," *Journal Neorocomputing*, vol. 310, pp. 223-235, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[34] Mukrimah Nawir et al., "Performances of Machine Learning Algorithms for Binary Classification of Network Anomaly Detection System," *1st International Conference on Big Data and Cloud Computing*, pp. 1-8, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[35] Hung Nguyen Viet et al., "Using Deep Learning Model for Network Scanning Detection," *Proceedings of the 4th International Conference on Frontiers of Educational Technologies*, pp. 117-121, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[36] Souhail Meftah, Tajjeeddine Rachidi, and Nasser Assem, "Network Based Intrusion Detection Using the UNSW-NB15 Dataset," *International Journal of Computing and Digital Systems*, vol. 8, no. 5, pp. 477-487, 2019. [CrossRef] [Google Scholar]

[37] Ramy Elhefnawy, Hassan Abounaser, and Amr Badr, "A Hybrid Nested Genetic-Fuzzy Algorithm Framework for Intrusion Detection and Attacks," *IEEE Access*, vol. 8, pp. 98218-98233, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[38] Abhishek Verma, and Virender Ranga, "Evaluation of Network Intrusion Detection Systems for RPL Based 6LoWPAN Networks in IoT," *Wireless Personal Communications*, vol. 108, pp. 1571-1594, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[39] Abhishek Verma, and Virender Ranga, "ELNIDS: Ensemble Learning based Network Intrusion Detection System for RPL based Internet of Things," *4th International Conference on Internet of Things: Smart Innovation and Usages*, Ghaziabad, India, pp. 1-6, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[40] P. Jaya Prakash, and B. Lalitha, "A Novel Intrusion Detection System for RPL Based IoT Networks with Bio-Inspired Feature Selection and Ensemble Classifier," *Research Square*, pp. 1-23, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[41] George Simoglou et al., "Intrusion Detection Systems for RPL Security: A Comparative Analysis," *Computers & Security*, vol. 104, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[42] Musa Osman et al., "Artificial Neural Network Model for Decreased Rank Attack Detection in RPL Based on IoT Networks," *International Journal of Network Security*, vol. 23, no. 3, pp. 496-503, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[43] Abhishek Verma, and Virender Ranga, "Mitigation of DIS Flooding Attacks in RPL-based 6LoWPAN Networks," *Emerging Telecommunications Technologies*, vol. 31, no. 2, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[44] Sarumathi Murali, and Abbas Jamalipour, "A Lightweight Intrusion Detection for Sybil Attack under Mobile RPL in the Internet of Things," *IEEE Internet of Things Journal*, vol. 7, no. 1, pp. 379-388, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[45] John Foley, Naghmeh Moradpoor, and Henry Ochenyi "Employing a Machine Learning Approach to Detect Combined Internet of Things Attacks against Two Objective Functions Using a Novel Dataset," *Security and Communication Networks*, vol. 2020, no. 1, pp. 1-17, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[46] Abhishek Verma, and Virender Ranga, "Machine Learning Based Intrusion Detection Systems for IoT Applications," *Wireless Personal Communications*, vol. 111, pp. 2287-2310, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[47] Mohammed Al Qurashi, Constantinos Marios Angelopoulos, and Vasilios Katos, "An Architecture for Resilient Intrusion Detection in IoT Networks," *ICC 2020 - 2020 IEEE International Conference on Communications*, Dublin, Ireland, pp. 1-7, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[48] Gaurav Soni, and R. Sudhakar, "A L-IDS against Dropping Attack to Secure and Improve RPL Performance in WSN Aided IoT," *7th International Conference on Signal Processing and Integrated Networks*, Noida, India, pp. 377-383, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[49] Arun Kumar Bediya, and Rajendra Kumar, "Real Time DDoS Intrusion Detection and Monitoring Framework in 6LoWPAN for Internet of Things," *IEEE International Conference on Computing, Power and Communication Technologies*, Greater Noida, India, pp. 824-828, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[50] K.N. Ambili, and Jimmy Jose, "TN-IDS for Network Layer Attacks in RPL based IoT Systems," *IACR Cryptology ePrint Archive*, pp. 1-24, 2020. [Google Scholar] [Publisher Link]

[51] Himanshu B. Patel, and Devesh C. Jinwala, "Blackhole Detection in 6LoWPAN Based Internet of Things: An Anomaly Based Approach," *TENCON 2019 - 2019 IEEE Region 10 Conference*, Kochi, India, pp. 947-954, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[52] Sarika Choudhary, and Nishtha Kesswani, "Cluster-Based Intrusion Detection Method for Internet of Things," *IEEE/ACS 16th International Conference on Computer Systems and Applications*, Abu Dhabi, United Arab Emirates, pp. 1-8, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[53] Emre Aydogan et al., "A Central Intrusion Detection System for RPL-Based Industrial Internet of Things," *15th IEEE International Workshop on Factory Communication Systems*, Sundsvall, Sweden, pp. 1-5, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[54] Ahmed Raoof, Ashraf Matrawy, and Chung-Horng Lung, "Routing Attacks and Mitigation Methods for RPL-Based Internet of Things," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1582-1606, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[55] Behnam Farzaneh, Mohammad Ali Montazeri, and Shahram Jamali, "An Anomaly-Based IDS for Detecting Attacks in RPL-Based Internet of Things," *5th International Conference on Web Research*, Tehran, Iran, pp. 61-66, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[56] Snehal Deshmukh-Bhosale, and Santosh S. Sonavane, "A Real-Time Intrusion Detection System for Wormhole Attack in the RPL based Internet of Things," *Procedia Manufacturing*, vol. 32, pp. 840-847, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[57] Mohamad Nazrin Napiah et al., "Compression Header Analyzer Intrusion Detection System (CHA - IDS) for 6LoWPAN Communication Protocol," *IEEE Access*, vol. 6, pp. 16623-16638, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[58] Ashwini Nikam, and Dayan Ambawade, "Opinion Metric Based Intrusion Detection Mechanism for RPL Protocol in IoT," *3rd International Conference for Convergence in Technology*, Pune, India, pp. 1-6, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[59] Elie Kfoury et al., "A Self Organizing Map Intrusion Detection System for RPL Protocol Attacks," *International Journal of Interdisciplinary Telecommunications and Networking*, vol. 11, no. 1, pp. 30-43, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[60] Faiza Medjek et al., "A Trust-Based Intrusion Detection System for Mobile RPL Based Networks," *IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, Exeter, UK, pp. 735-742, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[61] Minalini Goyal, and Maitreyee Dutta, "Intrusion Detection of Wormhole Attack in IoT: A Review," *International Conference on Circuits and Systems in Digital Enterprise Technology*, Kottayam, India, pp. 1-5, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[62] Furkan Yusuf Yavuz, Devrim Ünal, and Ensar Gül, "Deep Learning for Detection of Routing Attacks in the Internet of Things," *International Journal of Computational Intelligence Systems*, vol. 12, pp. 39-58, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[63] R. Darwin, "Implementation of Advanced IDS in Contiki for Highly Secured Wireless Sensor Network," *International Journal of Applied Engineering Research*, vol. 13, no. 6, pp. 4214-4218, 2018. [Google Scholar] [Publisher Link]

[64] Fatma Gara et al., "An Intrusion Detection System for Selective Forwarding Attack in IPv6-based Mobile WSNs," *13th International Wireless Communications and Mobile Computing Conference*, Valencia, Spain, pp. 276-281, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[65] Anhtuan Le et al., "A Specification-Based IDS for Detecting Attacks on RPL-Based Network Topology," *Information*, vol. 7, no. 2, pp. 1-19, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[66] Ting Miao, Ramnath Teja Chekka, and Ki-Hyung Kim, "GIDPS: A Game Theory-Based IDPS for RPL-Networked Low Power Lossy Networks with Energy Limitation," *Sixth International Conference on Ubiquitous and Future Networks*, Shanghai, China, pp. 278-283, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[67] Anhtuan Le et al., "Specification-Based IDS for Securing RPL from Topology Attacks," *IFIP Wireless Days*, Niagara Falls, ON, Canada, pp. 1-3, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[68] Erdem Canbalaban, and Sevil Sen, "A Cross-Layer Intrusion Detection System for RPL-Based Internet of Things," *Ad-Hoc, Mobile, and Wireless Networks*, pp. 214-227, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[69] Behnam Farzaneh et al., "A New Method for Intrusion Detection on RPL Routing Protocol Using Fuzzy Logic," *6th International Conference on Web Research*, Tehran, Iran, pp. 245-250, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[70] Alexander Vodyaho et al., "Data Collection Technology for Ambient Intelligence Systems in Internet of Things," *Electronics*, vol. 9, no. 11, pp. 1-26, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[71] Matija Stevanovic, and Jens Myrup Pedersen, "On the Use of Machine Learning for Identifying Botnet Network Traffic," *Journal of Cyber Security and Mobility*, vol. 4, pp. 1-32, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[72] Ahmad Azab, Mamoun Alazab, and Mahdi Aiash, "Machine Learning Based Botnet Identification Traffic," *IEEE Trustcom/BigDataSE/ISPA*, Tianjin, China, pp. 1788-1794, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[73] Sean Miller, and Curtis Busby-Earle, "The Role of Machine Learning In Botnet Detection," *11th International Conference for Internet Technology and Secured Transactions*, Barcelona, Spain, pp. 359-364, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[74] Qianru Zhou et al., "An Assessment of Intrusion Detection using Machine Learning on Traffic Statistical Data," *TechRxiv*, vol. 14, no. 8, pp. 1-10, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[75] K.V. Pradeepthi, and Arputharaj Kannan, "Detection of Botnet traffic by using Neuro-fuzzy based Intrusion Detection," *Tenth International Conference on Advanced Computing*, Chennai, India, pp. 118-123, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[76] Yair Meidan et al., "N-BaIoT—Network-Based Detection of IoT Botnet Attacks Using Deep Autoencoders," *IEEE Pervasive Computing*, vol. 17, no. 3, pp. 12-22, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[77] Jadel Alsamiri, and Khalid Alsubhi, "Internet of Things Cyber Attacks Detection using Machine Learning," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 12, pp. 627-634, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[78] Nour Moustafa, Benjamin Turnbull, and Kim-Kwang Raymond Choo, "An Ensemble Intrusion Detection Technique Based on Proposed Statistical Flow Features for Protecting Network Traffic of Internet of Things," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4815-4830, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[79] Duc C. Le, and Nur Zincir-Heywood, "Learning from Evolving Network Data for Dependable Botnet Detection," *15th International Conference on Network and Service Management*, Halifax, NS, Canada, pp. 1-5, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[80] Reem Alhajri, and Rachid Zagrouba, and Fahd Al-Haidari, "Survey for Anomaly Detection of IoT Botnets Using Machine Learning Auto-Encoders," *International Journal of Applied Engineering Research*, vol. 14, no. 10, pp. 2417-2421, 2019. [Google Scholar] [Publisher Link]

[81] Zainab Al-Othman, Mouhammd Alkasassbeh, and Sherenaz AL-Haj Baddar, "A State-of-the-Art Review on IoT Botnet Attack Detection," *arXiv*, pp. 1-6, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[82] V. Kanimozhi, and Thangavel Prem Jacob, "Artificial Intelligence Outflanks All Other Machine Learning Classifiers in Network Intrusion Detection System on the Realistic Cyber Dataset CSE-CIC-IDS2018 Using Cloud Computing," *ICT Express*, vol. 7, no. 3, pp. 366-370, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[83] Mustafa Alshamkhany et al., "Botnet Attack Detection using Machine Learning," *14th International Conference on Innovations in Information Technology(IIT)*, Al Ain, United Arab Emirates, pp. 203-208, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[84] A. Sankaran et al., "Botnet Detection Using Machine Learning," *International Research Journal of Engineering and Technology*, vol. 7, no. 7, pp. 5116-5121, 2020. [Google Scholar] [Publisher Link]

[85] MohammadNoor Injadat, Abdallah Moubayed, and Abdallah Shami, "Detecting Botnet Attacks in IoT Environments: An Optimized Machine Learning Approach," *32nd International Conference on Microelectronics*, Aqaba, Jordan, pp. 1-4, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[86] Aaya F. Jabbar, and Imad J. Mohammed, "Development of an Optimized Botnet Detection Framework based on Filters of Features and Machine Learning Classifiers using CICIDS2017 Dataset," *IOP Conference Series: Materials Science and Engineering*, vol. 928, pp. 1-13, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[87] Shreehar Joshi, and Eman Abdelfattah, "Efficiency of Different Machine Learning Algorithms on the Multivariate Classification of IoT Botnet Attacks," *11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference*, New York, NY, USA, pp. 517-521, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[88] Jiyeon Kim et al., "Feature Analysis of IoT Botnet Attacks based on RNN and LSTM," *International Journal of Engineering Trends and Technology*, vol. 68, no. 4, pp. 43-47, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[89] Jiyeon Kim et al., "Intelligent Detection of IoT Botnets Using Machine Learning and Deep Learning," *Applied Sciences*, vol. 10, no. 19, pp. 1-22, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[90] Madhuri Gurunathrao Desai, Yong Shi, and Kun Suo, "IoT Bonet and Network Intrusion Detection using Dimensionality Reduction and Supervised Machine Learning," *11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference*, New York, NY, USA, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[91] Mnahi Alqahtani, Hassan Mathkour, and Mohamed Maher Ben Ismail, "IoT Botnet Attack Detection Based on Optimized Extreme Gradient Boosting and Feature Selection," *Sensors*, vol. 20, no. 21, pp. 1-21, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[92] Alejandro Guerra-Manzanares et al., "MedBIoT: Generation of an IoT Botnet Dataset in a Medium-sized IoT Network," *Proceedings of the 6th International Conference on Information Systems Security and Privacy*, Valletta, Malta, vol. 1, pp. 207-218, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[93] Christian Dietz et al., "Towards Adversarial Resilience in Proactive Detection of Botnet Domain Names by using MTD," *NOMS 2020 - 2020 IEEE/IFIP Network Operations and Management Symposium*, Budapest, Hungary, pp. 1-5, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[94] Hasan Alkahtani, and Theyazn H.H. Aldhyani, "Botnet Attack Detection by Using CNN-LSTM Model for Internet of Things Applications," *Security and Communication Networks*, vol. 2021, no. 1, pp. 1-23, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[95] Rana Faek et al., "Exposing Bot Attacks Using Machine Learning and Flow Level Analysis," *International Conference on Data Science, E-learning and Information Systems*, pp. 99-106, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[96] Amritanshu Pandey et al., "Identification of Botnet Attacks Using Hybrid Machine Learning Models," *Hybrid Intelligent Systems*, pp. 249-257, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[97] P. Jithu et al., "Intrusion Detection System for IoT Botnet Attacks Using Deep Learning," *SN Computer Science*, vol. 2, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[98] Sikha Bagui, Xiaojian Wang, and Subhash Bagui, "Machine Learning Based Intrusion Detection for IoT Botnet," *International Journal of Machine Learning and Computing*, vol. 11, no. 6, pp. 399-406, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[99] Khlood Shinan et al., "Machine Learning-Based Botnet Detection in Software-Defined Network: A Systematic Review," *Symmetry*, vol. 13, no. 5, pp. 1-28, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[100] Han Wang et al., "Non-IID Data Re-Balancing at IoT Edge with Peer-To-Peer Federated Learning for Anomaly Detection," *Proceedings of the 14th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, pp. 153-163, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[101] Segun I. Popoola et al., "SMOTE-DRNN: A Deep Learning Algorithm for Botnet Detection in the Internet-of-Things Networks," *Sensors*, vol. 21, no. 9, pp. 1-20, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[102] Erman Özer, Murat İskefiyeli, and Jahongir Azimjonov, "Toward Lightweight Intrusion Detection Systems Using the Optimal and Efficient Feature Pairs of the Bot-IoT 2018 Dataset," *International Journal of Distributed Sensor Networks*, vol. 17, no. 10, 2021. [CrossRef] [Google Scholar] [Publisher Link]