REVIEW ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

A Review on IoT Intrusion Detection Systems Using Supervised Machine Learning: Techniques, Datasets, and Algorithms

Azeez Rahman Abdulla, Noor Ghazi M. Jameel

Technical college of Informatics, Sulaimani Polytechnic University, Sulaimani 46001, Kurdistan Region, Iraq

ABSTRACT

Physical objects that may communicate with one another are referred to "things" throughout the Internet of Things (IoT) concept. It introduces a variety of services and activities that are both available, trustworthy and essential for human life. The IoT necessitates multifaceted security measures that prioritize communication protected by confidentiality, integrity and authentication services; data inside sensor nodes are encrypted and the network is secured against interruptions and attacks. As a result, the issue of communication security in an IoT network needs to be solved. Even though the IoT network is protected by encryption and authentication, cyber-attacks are still possible. Consequently, it's crucial to have an intrusion detection system (IDS) technology. In this paper, common and potential security threats to the IoT environment are explored. Then, based on evaluating and contrasting recent studies in the field of IoT intrusion detection, a review regarding the IoT IDSs is offered with regard to the methodologies, datasets and machine learning (ML) algorithms. In This study, the strengths and limitations of recent IoT intrusion detection techniques are determined, recent datasets collected from real or simulated IoT environment are explored, high-performing ML methods are discovered, and the gap in recent studies is identified.

Index Terms: Internet of thing, Intrusion detection, Intrusion detection system techniques, Intrusion detection system datasets, Supervised machine learning

1. INTRODUCTION

A smart network called the Internet of Things (IoT) employs established protocols to link things to the Internet [1]. In an IoT network, smart tiny sensors join objects wirelessly. IoT devices can interact with one another without human involvement [2]. It uses distinctive addressing techniques to communicate, add more items and collaborate with them to develop new applications and services. Examples of IoT applications include smart environments, smart homes, and smart cities [3]. Thereby of the development of IoT applications, several obstacles have developed. One of these obstacles is IoT security that cannot be disregarded. IoT networks are subject to a range of malicious attacks because IoT devices can be accessed from anywhere over an unprotected network such as the Internet. The following security requirements should be considered when securing IoT environment:

- Confidentiality: IoT systems must ensure that unauthorized parties are prohibited from disclosing information [4].
- Integrity: Ensures that the messages must not have been modified in any manner [4].
- Availability: When data or resources are needed, they must be available [4]. Attackers can saturate a resource's bandwidth to degrade its availability.

Access this article online

DOI:10.21928/uhdjst.v7n1y2023.pp53-65 | E-ISSN: 2521-4217 P-ISSN: 2521-4209

Corresponding author's e-mail: Azeez Rahman Abdulla, Technical college of Informatics, Sulaimani Polytechnic University, Sulaimani 46001, Kurdistan Region, Iraq. Azeez.rahman.a@spu.edu.iq

Received: 18-10-2022 Accepted: 23-12-2022 Published: 01-03-2023

- Authenticity: The word "authenticity" relates to the ability to prove one's identity. The system should be able to recognize the identity of the entity with whom it is communicating [5].
- Non-repudiation: This guarantees that nothing can be rejected. In an IoT context, a node cannot reject a message or piece of data that has already been sent to another node or a user [6].
- Data freshness: Ensures that no outdated messages are retransmitted by an attacker [7].

In the last few years, advancement in artificial intelligent (AI) such as machine learning (ML) techniques has been used to improve IoT intrusion detection system (IDS). Numerous studies as [8,9], reviewed and compared different applied ML algorithms and techniques through various datasets to validate the development of IoT IDSs. However, it's still not clear a recent dataset collected from IoT environment, and which ML model was more effective for building an efficient IoT IDS. Therefore, the current requirement is to do an up-to-date review to identify these critical points.

In this study, a survey of the IoT IDSs is given. This paper aims to further the knowledge in regard to IoT cyber attacks' characteristics (motivation and capabilities). Then, strengths and limitations of different categories of IDSs techniques (hybrid, anomaly-based, signature-based, and specification-based) are compared. Moreover, the study presents a review on the recent researches in the area of IoT intrusion detection using ML algorithms for IoT network based on the datasets, algorithms and evaluation metrics to identify the recent IoT dataset and the outperformed ML algorithm in terms of accuracy used for IoT intrusion detection.

The paper is structured as follows: In section 2, common cyber-attacks in IoT the environment are clarified. In section 3 the strengths and limitations of IoT intrusion detection techniques are discussed. Section 4 discussed, analyzed and compared recent IoT intrusion detection researches' performance metrics, datasets and supervised ML algorithms. Finally, section 5 illustrates the conclusions of the paper.

2. IoT CYBER ATTACKS

Recently, IoT has developed quickly, making it the fastest-growing enormous impact of technology on social interactions and workplace environments, including education, healthcare and commerce. This technology is used for storing the private data of people and businesses, for financial data transactions, for product development and for marketing. Due to the widespread adoption of linked devices in the IoT, there is a huge global demand for strong security. Millions or perhaps billions of connected devices and services are now available [10-13]. Every day, there are more risks and assaults have gotten more frequent and sophisticated. In addition, sophisticated technologies are becoming more readily available to potential attackers [14,15]. To realize its full potential, IoT must be secured against threats and weaknesses [16]. By maintaining the confidentiality and integrity of information about the object and making that information easily accessible whenever it is needed, security is the act of avoiding physical injury, unauthorized access, theft, or loss to the item [17]. To ensure IoT security, it is crucial to maintain the greatest inherent value of both tangible items (devices) and intangible ones (services, information and data). System risks and vulnerabilities must be identified in order to provide a comprehensive set of security criteria to assess if the security solution is secure against malicious assaults or not [18]. Attacks are performed to damage a system or obstruct regular operations by utilizing various strategies and tools to exploit vulnerabilities. Attackers launch attacks to achieve goals, either for their personal satisfaction or to exact revenge [19]. Common IoT cyber-attack types are:

- Physical attacks: These assaults tamper with hardware elements. Most IoT devices often operate in outdoor areas which are extremely vulnerable to the physical assaults [20].
- Attacks known as reconnaissance include the illegal identification of systems, services, or vulnerabilities. The scanning of network ports is an example of a reconnaissance attack [21].
- Denial-of-service (DoS): This type of attack aims to prevent the targeted users from accessing a computer or network resource. The majority of IoT devices are susceptible to resource enervation attacks due to their limited capacity for memory and compute resources [22].
- Access attacks happen when unauthorized users get access to networks or devices that they are not allowed to use. Two types of access assaults exist: The first is physical access, in which a hacker gains access to a real object. The second is using IP-connected devices for remote access [22].
- Attacks on privacy: IoT privacy protection has grown more difficult as a result of the volume of information that is readily accessible via remote access techniques [14].
- Cyber-crimes: Users and data are used for hedonistic activities including fraud, brand theft, identity theft, and

theft of intellectual property using internet and smart products [14,15,23].
- Destructive attacks: Space is exploited to cause widespread disturbance and property and human life loss. Terrorism and retaliation are two examples of damaging assaults.
- Supervisory Control and Data Acquisition (SCADA) Attacks: SCADA systems are connected to industrial IoT networks; they are active devices in real-time industrial networks, which allow the remote monitoring and control of processes, even when the devices are located in remote areas. The most specific and common types of SCADA attacks are eavesdropping, man-in-the middle, masquerading, and malware [24].

3. IoT INTRUSION DETECTION SYSTEM

Despite the investment and potential it holds, there are still issues that prevent IoT from becoming a widely utilized technology. The security challenges with IoT are thought to be solvable via intrusion detection, which has been established for more than 30 years. Intrusion detection is often a system (referred to as IDS) which consists of tools or methods that analyze system activity to find assaults or unauthorized access. An IDS typically comprises of sensors, and a tool to evaluate the data from these sensors. Efficient and accurate intrusion detection solutions are necessary in the IoT environment to identify various security risks [25].

3.1. IoT Intrusion Detection Types

IDS types can be categorized in a variety of ways, particularly IDS for IoT as the majority of them are still being studied. According to Das *et al.*, [26] the research distinguishes three types of IDS:
- Host-based IDS (HIDS): To keep an eye on the system's harmful or malicious activity, HIDS is connected to the server. Specifically, HIDS examines changes in file-to-file communication, network traffic, system calls, running processes, and application logs. This sort of IDS's drawback is that it can only identify attacks on the systems it supports.
- Network-based IDS (NIDS): NIDS analyzes network traffic for attack activities and identifies harmful behavior on network lines.
- Distributed IDS (DIDS): DIDS will have a large number of linked and dispersed IDSs for attack detection, incident monitoring and anomaly detection. To monitor and respond to outside actions, DIDS needs a central server with strong computing and orchestration capabilities.

3.2. IoT Intrusion Detection Techniques

There are four basic types or methodologies for deploying IoT intrusion detection.
- Anomaly based IDS in IoT.
  It uses anomaly based IDS to find intrusions and monitor abusive behavior. It employs a threshold to determine if this behavior is typical or abnormal. These IDSs have the ability to monitor a typical IoT network's activity and set a threshold. To detect abnormalities, the network's activity is compared to a threshold and any deviation from this number is considered abnormal [27]. Table 1 compares and contrasts the strength and limitations of several anomaly-based IDSs methodologies based on resource and energy usage, detection accuracy and speed.
- Signature based IDS in IoT
  Signature based detections compare the network's current activity to pre-defined attack patterns. Each signature is connected to a particular assault since signatures are originally established and stored on the IoT device. Signature based approaches are commonly used and require a signature for each assault [27]. The strengths and limitations of different signature based IDSs techniques have been presented and compared in Table 2 based on resource consumption, energy, detection accuracy, and speed.
- Specification based IDS in IoT
  Specification-based approaches detect intrusions when network behavior deviates from specification definitions. Therefore, specification-based detection has the same purpose of anomaly-based detection. However, there is one important difference between these methods: In specification-based approaches, a human expert should manually define the rules of each specification [36]. The main aspects of specification-based IDSs have been outlined and then compared in Table 3 based on resource consumption, energy, detection accuracy, and speed.
- Hybrid IDS in IoT
  Signature based IDS has a large usable capacity and limited number of attack detections while anomaly based IDS has a high false positive rate and significant computation costs. A hybrid technique was suggested to solve the flaws of both systems [42]. The main characteristics of hybrid IDSs have been defined and then compared in Table 4 based on resource consumption, energy, detection accuracy, and speed.

TABLE 1: Comparison of different anomaly based IDS techniques

Reference No.	Technique	Strength	Limitations
[28]	Utilizing a fusion based technique to decrease the damage caused by strikes.	• Low communication overhead	• High energy consumption
[29]	Detecting Wormhole attacks using node position and neighbor information.	• Low resource consumption • Real time • Energy efficient	• Only One type of attack can be detected
[30]	Detecting sinkhole attacks by analyzing the behavior of devices	• Detection accuracy is high	• Detect limited number of attacks
[31]	A lightweight technique for identifying normal and deviant behavior	• Lightweight implementation • Detection accuracy is high	• High computational overhead
[32]	A request-response method's correlation functions are used to look for unusual network server activity	• Consuming modest resources • Lightweight detection system	• High computational overhead

IDS: Intrusion detection system

TABLE 2: Comparison of different signature based IDS techniques

Reference No.	Technique	Strength	Limitations
[33]	Detecting network attacks by signature code in IP based ubiquitous sensor networks	• High detection accuracy • Low energy and resource consumption	• Can detect limited number of intrusions
[34]	The pattern-matching engine is used to detect malicious nodes using auxiliary shifting and early decision techniques	• Low memory and computational complexity • Maximum speed up	• Not real-time • Can detect limited number of intrusions
[35]	Detection of malware signature detection using reversible sketch structure based on cloud.	• Fast • Low communication consumption • High detection accuracy	• High memory requirement • Has a limited ability to identify assaults

IDS: Intrusion detection system

TABLE 3: Comparison of different specification based IDS techniques

Reference No.	Technique	Strength	Limitations
[37]	Mitigation of black hole attacks Using an effective strategy in routing protocol for low-power and lossy (RPL) Networks	• Low delay • High detection accuracy of the infected node	• Only black hole attacks can be detected
[38]	Detecting internal attacks by designing a secure routing protocol based on reputation mechanism	• Detection accuracy is acceptable • Low delay	• Needs skilled administration
[39]	Topology assaults detection on RPL using semi-automated profiling tool.	• Detection accuracy is high • Low energy consumption • Low computation overhead	• High overhead
[40]	Sinkhole attacks are detected using a constraint based specification intrusion detection approach.	• Low overhead • Minimal energy usage	• Not real-time
[41]	Using a game-theoretic method to identify deceptive attacks in IoT network with honeypots.	• High detection accuracy • Real-time	• Needs additional resources. • High converge time

IDS: Intrusion detection system

4. SUPERVISED ML BASED IOT INTRUSION DETECTION

ML enables computer systems to predict events more correctly without being explicitly taught to do so. It is a subset of artificial intelligence (AI). ML algorithms use historical data as input to anticipate new output values. ML algorithms are mainly divided into three categories: reinforcement learning, unsupervised learning, and supervised learning. In this paper, recent researches using supervised ML algorithms

in the area of IoT intrusion detection were studied, analyzed and compared. Supervised learning emphasis on discovering patterns while utilizing labeled datasets. In supervised learning, the machine must be fed sample data with different characteristics (expressed as "X") and the right value output of the data (represented as "y"). The dataset is considered "labeled" because the output and feature values are known. Then, the algorithm analyzes data patterns to develop a model that can replicate the same fundamental principles with new data [46].

4.1. Datasets Used for IoT Intrusion Detection

Models for supervised ML are trained and evaluated using datasets. Any IDS's performance ultimately depends on the dataset's quality including whether it can reliably identify assaults or not [47]. Here, six datasets named NSL-KDD, UNSWNB15, CICIDS 2017, Bot-IoT, DS2OS, and IoTID20 are considered and used by researchers to train and test IoT intrusion detection models. Descriptions of the datasets are given below and their characteristics are summarized in Table 5.

- NSL-KDD
  The NSL-KDD dataset is an improved version of the KDD99. It does not include redundant records in the train set, so the classifiers will not be biased towards more frequent records. The number of selected records from each difficulty level group is inversely proportional to the percentage of records in the original KDD data set [47]. The NSL-KDD dataset has 41 characteristics, classified into three categories: Basic characteristics, content characteristics, and traffic characteristics.

TABLE 4: Comparison of different hybrid IDS techniques

Reference No.	Technique	Strength	Limitations
[42]	Employing a game theoretic approach to identify attackers by using anomaly detection only when a new attack pattern is anticipated and using signature based detection otherwise.	• Detection accuracy is high • Low energy consumption	• High resource consumption • Delay
[43]	The denial of service prevention manager is proposed, which uses aberrant activity detection and matching with attack signatures.	• Real time	• High resource consumption
[44]	Real-time attack detection using knowledgeable, self-adapting expert intrusion detection system.	• High detection accuracy • Real time • Low resource consumption	• High computational overhead
[45]	Attackers can be found by looking for timing irregularities while broadcasting the most recent rank to nearby nodes and using a timestamp.	• Real time • Low overhead • Low delay • High detection accuracy	• High computation overhead • High resource consumption
[27]	Targeting the routing attacks with an IDS with integrated mini-firewall which uses anomaly-based IDS in the intrusion detection and signature-based IDS in the mini-firewall	• Real Time • High availability • Low overhead	• Limited in dynamic network topology • High-resource consumption • Low detection accuracy

IDS: Intrusion detection system

TABLE 5: Dataset characteristics

Dataset	Year	Dataset link (URL)	No. of Instances	No. of Features	Dataset collection performed on IoT environment	Type of dataset
NSLKDD	2009	https://www.unb.ca/cic/datasets/nsl.html	148,519	41	No	Imbalanced
UNSW-NB15	2015	https://research.unsw.edu.au/projects/unsw-nb15-dataset	2,540,044	49	No	Imbalanced
CICIDS2017	2017	https://www.unb.ca/cic/datasets/ids-2017.html	2,830,743	83	No	Imbalanced
BoT- IoT	2019	https://ieee-dataport.org/documents/bot-iot-dataset	73,370,443	29	Yes	Imbalanced
DS2OS	2018	https://www.kaggle.com/datasets/francoisxa/ds2ostraffictraces	409,972	13	Yes	Imbalanced
IoTID20	2020	https://sites.google.com/view/iot-network-intrusion-dataset/home	625,783	83	Yes	Imbalanced

- UNSW-NB15
  The UNSW-NB15 dataset was published in 2015. It was created by establishing the synthetic environment at the UNSW cyber security lab. UNSW-NB15 represents nine major families of attacks by utilizing the IXIA Perfect Storm tool. IXIA tool has provided the capability to generate a modern representative of the real modern normal and the abnormal network traffic in the synthetic environment. There are 49 features and nine types of attack categories known as the analysis, fuzzers, Backdoors, DoS, exploits, reconnaissance, generic, shellcode, and worms [48].
- CICIDS 2017
  The CICIDS 2017 dataset generated in 2017. It includes benign and seven common family of attacks that met real worlds criteria such as DoS, DDoS, brute force, XSS, SQL injection, Infiltration, port scan, and botnet. The dataset is completely labeled with 83 network traffic features extracted and calculated for all benign and attack network flows [49].
- BoT-IoT
  The BoT-IoT dataset was created by designing a testbed network environment in the Research Cyber Range Lab of UNSW Canberra. This dataset consists of legitimate and simulated IoT network traffic along with various types of attacks such as information gathering (probing attacks), denial of service and information theft. It has been labeled with the label features indicating an attack flow, the attacks category and subcategory for possible multiclass classification purposes [50].
- DS2OS
  This dataset includes traces that were recorded using the IoT platform DS2OS. Labeled and unlabeled datasets come in two varieties. The only characteristics in an unlabeled dataset that can be used describe the data objects for unsupervised ML models. In addition, a labeled dataset includes information about each data instance's class and utilized for supervised ML models [51].
- IoTID20
  IoTID20 dataset is used for anomalous activity detection in IoT networks. The testbed for the IoTID20 dataset is a combination of IoT devices and interconnecting structures. The dataset consists of various types of IoT attacks and a large number of flow-based features. The flow-based features can be used to analyze and evaluate a flow-based IDS. The final version of the IoTID20 dataset consists of 83 network features and three label features [52].

4.2. Supervised ML Algorithms Used for IoT Intrusion Detection

For IoT intrusion detection, many supervised ML methods are employed. The list of used algorithms with corresponding descriptions is presented below:

- Logistic regression (LR): It is a probability-based method for predictive analysis. It is a more effective strategy for binary and linear classification issues because it employs the sigmoid function to translate expected values to probabilities between 0 and 1. It is a classification model that is relatively simple to implement and performs extremely well with linearly separable data classes [53].
- Naïve base (NB): Are a group of Bayes' Theorem-based categorization methods. It is a family of algorithms rather than a single method and they all operate under the same guiding principle in which each pair of characteristics is categorized standalone [53].
- Artificial neural networks (ANN): The biological neural network in the human brain served as the model for the widely used ML technology known as (ANN). Each artificial neuron's weight values are sent to the following layer as an output. Feed-forward neural network form of ANN that processes inputs from neurons in the previous layer. Multilayer perception is a significant type of feed forward neural networks (MLP). The most well-known MLP training method that modifies the weights between neurons to reduce error is called the back propagation algorithm. The system can display sluggish convergence and run the danger of a local optimum, but it can rapidly adapt to new data values [54].
- Support Vector Machine (SVM): This algorithm looks for a hyperplane to optimize the distance involving two classes. A learning foundation for upcoming data processing is provided by the categorization. The groups are divided into several configurations by the algorithm through hyperplanes (lines). A learning model that splits up new examples into several categories is produced by SVM. Based on these functions, SVMs are referred to as non-probabilistic, or binary linear classifiers. In situations that use probabilistic classification, SVMs can use methods such as Platt Scaling [53].
- Decision tree (DT) is a tree in which each internal node represents an assessment of an attribute. Each branch represents the result of an assessment and each leaf node denotes the classification outcome. Algorithms such as ID3, CART, C4.5, and C5.0 are frequently used to generate decision trees. By analyzing the samples, a decision tree is obtained and used to correctly classify new data [55].

- Random forest (RF) is a technique used to create a forest of decision trees. This algorithm is frequently used due to its fast operation. Countless decision trees can be used to create a random forest. By averaging the outcomes of each component tree's forecast, this method generates predictions. Random forests exhibit compelling accuracy results and are less likely to overfit the data than a traditional decision tree technique. This method works well while examining plenty of data [53].
- Ensemble Learning (which includes bagging and boosting). The boosting method is a well-known ensemble learning method for improving the performance and accuracy of ML systems. The fundamental idea behind the boosting strategy is the successive addition of models to the ensemble. Weak learners (base learners) are efficiently elevated to strong learners. As a consequence, it aids in reducing variation and bias and raising prediction accuracy. Boosting is an iterative method that alters the findings of an observation's weight depending on the most recent categorization. Adaboost (AB), gradient boosting machines (GBM), and extreme gradient boosting (XGBoost) are examples of boosting techniques. Bagging (also known as bootstrap aggregating). It is one of the earliest and most basic ensemble ML approaches and it works well for issues requiring little in the way of training data. In this approach, a collection of original models with replacement are trained using random subsets of data acquired using the bootstrap sampling method. The individual output models derived from bootstrap samples are combined by majority voting [56].

4.3. Evaluation Metrics

The efficiency of ML algorithms can be measured using metrics such as accuracy, precision, recall, and F1-score [57]. Performance metrics are calculated using different parameters called True positive (TP), False positive (FP), True negative (TN), and False negative (FN). For IDS s, these parameters are described as follow:

TP = The number of cases correctly identified as attack.

FP = The number of cases incorrectly identified as attack.

TN = The number of cases correctly identified as normal.

FN = The number of cases incorrectly identified as normal.

- **Precision** (also called positive predictive value) is the percentage of retrieved occurrences that are relevant. Model performance is considered better if the precision is higher [58]. Precision is computed using (1) [59].

$$\text{Precision} = \frac{\text{True postive}}{\text{True postive+False postive}} \quad (1)$$

- **Recall** (also known as sensitivity) is the percentage of occurrences that were found to be relevant. It also goes by the name True Positive Rate (TPR) and calculated using (2) [58].

$$\text{Recall} = \frac{\text{True postive}}{\text{True postive+False negative}} \quad (2)$$

- **Accuracy** is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations. Model accuracy is calculated using (3) [57].

$$\text{Accuracy} = \frac{\text{True postive+True negative}}{\text{Total}} \quad (3)$$

- **F1-Score** is the harmonic mean of recall and accuracy [60] which defines a the weighted average of recall and precision and calculated using (4) [57].

$$\text{F1 score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

- **ROC curve** is a receiver operating characteristic curve which shows the performance of a classifier at various thresholds level [57].
- **Area under curve (AUC):** is closely associated with the concept of ROC. It represents the area under the ROC curve. It has been extensively used as a performance measure for classification models in ML. Its values range from 0 to 1. The higher the value, the better the model is [61].

4.4. Analysis and Comparison of Supervised ML Algorithms for IoT Intrusion Detection

In this section, the analysis of the used ML algorithms has been presented and discussed. Researchers used many supervised ML algorithms specifically in classification and they performed well in some cases with very high accuracy. To review researches in the area of intrusion detection using ML in the IoT environment, various recent studies are examined and compared based on the ML algorithms (classifier), datasets, type of classification, and performance of the classifier. The performance of these algorithms depends on various metrics. In this study, the comparison among the algorithms is focused on accuracy metric. Detailed

review of 21 papers (published between 2019 and 2022) was analyzed in this section and compared in Table 6.

Mahmudul *et al.* [62] employed the DS2OS dataset with several ML algorithms (LR, SVM, DT, RF, ANN). Accuracy,

TABLE 6: Comparison of the selected supervised ML based IoT IDS

Reference No.	Year	ML algorithm (classifier)	Dataset	Classification type	Classifier accuracy
[62]	2019	LR, SVM, DT, RF, ANN	DS2OS	Multiclass	LR=0.983, SVM=0.982, DT=0.994, RF=0.994, ANN=0.994.
[63]	2019	RF	UNSW-NB15	Binary	RF=99.34
[64]	2019	LR, NB, DT, RF, KNN, SVM	KDD99, NSL-KDD, UNSW-NB15	Binary	Accuracy of the algorithms depend on the used dataset
[65]	2019	For the level-1 model, DT For level 2 model, RF	CICIDS2017, UNSW-15	2 level classification (binary then multiclass)	Both datasets' specificity was 100% for the model, while its precision, recall, and F score were all 100% for the CICIDS2017 dataset and 97% for the UNSW-NB15 dataset
[66]	2019	RF, AB, GBM, XGB, DT (Cart), MLP, extremely randomized trees (ETC)	CIDDS-001, UNSW-NB15, NSL-KDD	Binary	Average accuracy value for 4 datasets using holdout are: RF=94.94, GBM=92.98, XGB=93.15%, AB=90.37, CART=91.98, MLP=82.76, ETC=82.99
[67]	2019	DT, NN, SVM	UNSW-NB15	Multiclass	DT=89.76, NN=86.7و SVM=78.77, Proposed model: 88.92
[68]	2019	NB, QDA, RF, ID3, AB, MLP, KNN	BoT-IoT	Binary.	NB=0.78, QDA=0.88, RF=0.98, ID3=0.99, Adaboost=1.0, MLP=0.84, KNN=0.99
[69]	2019	SVM, LR, D T, KNN, RF	UNSW-NB15, their own dataset	Binary	The accuracy depends on the dataset and the algorithm
[58]	2020	RF, XGB, DT, MLP, GB, ET, LR	UNSW-NB15	Binary	Results with all features: RF=0.9516, XGB=0.9481, DT=0.9387, MLP=0.9371, GB=0.9331, ET=0.9501, LR=0.8984
[53]	2020	KNN, SVM, DT, NB, RF, ANN, LR	Bot-IoT	Binary, multiclass	On binary classification: KNN=0.99, SVM=0.99, DT=1.0, NB=0.99, RF=1.0 ANN=0.99, LR=0.99
[70]	2020	SVM, NB, DT, adaboost	Their own synthetic called (Sensor480)	Binary	SVM=0.9895, NB=0.9789, DT=1.0000, Adaboost=0.9895
[71]	2020	RF	IoTID20 dataset	Binary based on the attack type	The accuracy result depends on the attack type
[72]	2021	SVM	NSL-KDD, UNSW-NB15.	Binary, multiclass	The accuracy depends on the dataset, the type of classification and number of features
[55]	2021	RF, SVM, ANN	UNSW-NB15.	Binary, multiclass	All features: RF with Binary=98.67, Multi-class=97.37, SVM in Binary=97.69, Multiclass=95.67, ANN in Binary=94.78, multiclass=91.67
[73]	2021	LR, SVM, DT, ANN	IoTID20, BoT-IoT	Multiclass	The results are based on the dataset and the categories of attacks
[74]	2021	SLFN	IoTID20	Binary	The proposed model=0.9351
[75]	2021	SVM, GBDT, RF	NSL KDD	Binary	SVM=32.38, GBDT=78.01, RF=85.34
[76]	2021	B-stacking	CICIDS2017, NSL-KDD	Multiclass	Accuracy for CICIDS2017 is 99.11% Accuracy for NSL-KDD approximately is 98.5%
[77]	2022	DT, RF, GBM	IoT2020	Binary	DT=0.978305, RF=0.978443, GBM=0.9636
[78]	2022	Shallow neural networks (SNN), bagging trees (BT), DT, SVM, KNN	IoTID20	Binary, multiclass	For binary classification all models achieved 100% For multiclass: SNN=100%, DT=99.9%, BT=99.9%, SVM=99,8%, KNN=99.4%
[79]	2022	ANN, DT (C4.5), Bagging, KNN, Ensemble	IoTID20, NSL-KDD	Binary, multiclass	Accuracy depends on feature selection approaches, datasets, and attack type for multiclass classification

precision, recall, f1 score, and area under the receiver operating characteristic curve are the assessment measures used to compare performance. The measurements show that RF performs comparably higher performance, and the system acquired excellent accuracy (Ibrahim *et al.* [63]). An intelligent anomaly detection system called Anomaly Detection IoT (AD-IoT) which used the UNSW-NB15 dataset and RF to identify binary labeled categorization had been proposed. The results demonstrated that the AD-IoT could successfully produce the best classification accuracy while minimizing the false positive rate. Samir *et al.* in [64] used the datasets KDD99, NSL-KDD, and UNSW-NB15 to assess number of ML models. The KKN and LR algorithms produced the best results on the UNSW-NB15 dataset while the NB algorithm produced the worst results. On the NSL-KDD dataset, the DT classifier outperformed the others in terms of various metrics while on the KDD99 dataset, SVM and MLP produced a low false positive rate in comparison to other algorithms. The findings of this study showed that the DT and KNN algorithms outperformed the other algorithms. However, the KNN required more time to categorize data than the DT. Imtiaz and Qusay [65] conducted a two-level framework experiment for IoT intrusion detection. To determine the category of the anomaly, they chose a DT classifier for the level-1 model which categorized the network flow as normal or anomalous and forwarded the network flow to the level-2 model. RF was used as a level-2 model for multiclass categorization. Abhishek and Virender [66] employed both ensemble and single classifiers, two different types of classification techniques. The selection of the aforementioned classification algorithms was primarily influenced by the huge number of input characteristics that are vulnerable to overfitting. As a result, random search was used to determine the best input parameters for RF, AB, XGB, and GBM. In terms of precision, RF beats other classifiers. However, AB performs the worst of all the classifiers. Using Friedman test statistics and 10-fold validation, the results showed that the classifiers' performances are considerably varied. Following that, the average time required by several classifiers to categorize a single case, CART classifies instances of CIDDS-001, UNSW-NB15, KDDTrain+, and KDDTest+ faster than other classifiers. Vikash *et al.* [67] proposed (UIDS) an IDS using UNSW-NB15 dataset. Network traffic accuracy and assault detection rate were improved by the suggested approach. In addition, it examined data using several ML techniques (C5, neural network, SVM, and UIDS model) and came to the conclusion that UIDS compared favorably to other ML techniques. Analysis showed that the false alarm rate (FAR) of the UNSW-NB15 dataset was reduced with only 13 characteristics. Jadel and Khalid [68] tested seven ML algorithms. All the algorithms, except the Naive Bayes (NB) and Quadratic algorithm (QDA), achieved highest success in detecting almost all attack types. It can be seen that Adaboost was the best performance algorithm, followed by KNN and ID3. ID3 is noticeably faster than KNN. The accuracy of the algorithms depends on the entire dataset with the seven best features obtained in the feature selection step. Aritro *et al.* [69] analyzed the role of a set of chosen ML techniques for IoT intrusion detection based on dataset/flows two layers: Application layer (host based) and network layer (network based). For the application layer dataset, they created their own dataset from the IoT environment while for network layer they used UNSW-NB15 dataset. According to the results for both datasets, RF was the best algorithm in terms of accuracy and LR was the fastest in terms of speed. Mohammad [58] used different algorithms. The classifiers random forest (RF) and extra trees (ET) performed better than the others, and RF is the best of the two. Only 14 features were chosen by RF utilizing features selection, but the performance results were remarkably similar to those achieved with all features. In addition, compared to the others, the LR classifier had the lowest accuracy. Andrew *et al.* [53] employed different methods; nevertheless, the findings show that RF performed better with the non-weighted dataset regarding precision and accuracy in non-weighted dataset. However, ANN performed more accurately in binary classification using weighted dataset. KNN and ANN performed extremely well in multi-classification for weighted and non-weighted datasets, respectively. The findings made it clear that ANN accurately predicted the kind of attack. K. V. V. N. L *et al.* [70] tested four ML techniques on IoT traffic in order to distinguish between genuine and attack traffic. Using decision trees, all of the analyzed data may be precisely categorized into the correct classes. Decision trees also had the greatest accuracy compared to the other classifiers. Pascal *et al.* [71] suggested a new anomaly-based detection using hybrid feature selection for IoT networks using IoTID20 dataset. The relevant features were fed to the RF algorithm. Based on the attack category, the network traffic is classified as normal and attack category as DoS, Scan, or MITM. Nsikak *et al.* [72] tested SVM with dataset NSL-KDD and UNSW-NB15 datasets. The results using different numbers of features for both datasets were varied. The classification accuracy using binary classification was greater than multi-class according to the evaluation results. Muhammad *et al.* [55], the UNSW-NB15 dataset had been subjected to supervise ML including RF, SVM, and ANN.

The application of RF using mean imputation produced the greatest accuracy in binary classification. Overall, there were not many differences in accuracy across the different imputation strategies. By using RF on a regression-imputed dataset, the greatest accuracy in multi-class classification was also attained. In addition, as compared to other cutting-edge supervised ML-based techniques, RF achieved greater accuracy with less training time for clustered based classification. Khalid *et al.* [73] for classification objectives, the performance of four ML methods was assessed. The Bot-IoT dataset and the IoTID20 dataset were both utilized in the study, 5% of Bot-IoT dataset was selected with a full set of features, while the second dataset was fully selected in the experiment. The accuracy results were based on the dataset and the categories of attacks. Raneem *et al.* [74] developed an intrusion detection method using a single layer forward neural network (SLFN) classifier with IoTID20. The results showed that the SLFN classification approach outperformed other classification algorithms. Maryam *et al.* [75] proposed that three ML algorithms RF, GDBT, and SVM were applied to the NSL-KDD dataset using binary classification. The results showed that the RF obtained the highest accuracy on the fog layer while SVM obtained lowest accuracy. Souradipst *et al.* [76] proposed B-Stacking approach as an intrusion detection model to detect cyber-attacks and anomalies in IoT networks. B-Stacking is based on a combination of two ensemble algorithms; boosting and stacking. It chose KNN, RF, and XGBoost as the level-0 weak learners. XGboost is also used as the level-1 learner. The experimental results on two popular datasets showed that the model had a high detection rate and a low false alarm rate. Most importantly, the proposed model is lightweight and can be deployed on IoT nodes with limited power and storage capabilities. Jingyi *et al.* [77] used DT, RF, and GBM ML algorithms with a dataset generated from the IoTID20 dataset known as IoT2020 dataset. According to the results, the DT algorithm performed more accurately than the other algorithms, but RF had better AUC score. Abdulaziz *et al.* [78] proposed an anomaly intrusion detection in an IoT system. Five supervised ML models were implemented to characterize their performance in detecting and classifying network activities with feature engineering and data preprocessing framework. Based on experimental evaluation, the accuracy 100% recorded for the detection phase that distinguishes the normal and anomaly network activities. While for classifying network traffic into five attack categories, the implemented models of achieved 99.4-99.9%. Khalid *et al.* [79] proposed and implemented an IoT anomaly-based IDS based on novel feature selection and extraction approach. The model framework was trained and tested on IoTID20 and NSL-KDD datasets using four ML algorithms. The system scored a maximum detection accuracy of 99.98% for the proposed ML ensemble-based hybrid feature selection approach.

From the literature, it is observed that there are extensive efforts on developing IDS s for IoT. Several researchers have assessed the effectiveness of their systems using common datasets like NSL-KDD, UNSW-NB15, and CICIDS2017. These datasets were not used captured traffic from IoT environment. Hence, an extensive work should be conducted using recent datasets such as IoTID20 which consists of IoT network traffic features. The state of the art also shows that some models perform well, particularly tree-based algorithms such as boosting, random forest and decision trees. ML algorithms' performance outcomes vary depending on the used dataset, features, and classification category.

5. CONCLUSION

One of the most important technological progresses over the past decade was the widespread adoption of IoT devices across industries and societies. With the development of IoT, several obstacles have been raised. One of these obstacles is IoT security which cannot be disregarded. IoT networks are vulnerable to a variety of threats. Although the IoT network is protected by encryption and authentication, cyber attacks are still possible. Therefore, using IoT IDS is important and necessary. This paper conducted an in-depth comprehensive analysis and comparison of various recent researches which used different techniques, datasets, ML algorithms and their performance for detecting IoT intrusions. Based upon the analysis, the recent IoT dataset for intrusion detection is identified which is IoTID20 dataset. Furthermore, the ML algorithms that outperformed in most researches are tree-based algorithms such as DT, RF, and boosting algorithms. Many points were observed and needed further study like using and collecting real IoT intrusion detection datasets for training and testing ML models, real time, and lightweight IDSs are required that need less detection time and resources consumption. All these factors should be taken into account while developing new IoT IDSs. In addition, further study should be conducted to address recent IoT threats, and the need to identify the best IDS placement techniques that improve IoT security while lowering the risk of cyber attacks.

REFERENCES

[1] S. Chen, H. Xu, D. Liu, B. Hu and H. Wang. "A vision of IoT: Applications, challenges, and opportunities with china perspective."

IEEE Internet of Things Journal, vol. 1, no. 4, pp. 349-359, 2014.

[2] S. Li, L. D. Xu and S. Zhao. "The internet of things: A survey". *Information Systems Frontiers*, vol. 17, no. 2, pp. 243-259, 2015.

[3] T. Sherasiya and H. Upadhyay. "Intrusion detection system for internet of things". International Journal of Advance Research and Innovative Ideas in Education, vol. 2, no. 3, pp. 2244-2249, 2016.

[4] M. M. Patel and A. Aggarwal. "*Security Attacks in Wireless Sensor Networks: A Survey*". In: 2013 International Conference on Intelligent Systems and Signal Processing (ISSP). Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, pp. 329-333, 2013.

[5] S. N. Kumar. "Review on network security and cryptography". *International Transaction of Electrical and Computer Engineers System*, vol. 3, no. 1, pp. 1-11, 2015.

[6] R. S. M. Joshitta, L. Arockiam. "Security in IoT environment: A survey". *International Journal of Information Technology and Mechanical Engineering*, vol. 2, no. 7, pp. 1-8, 2016.

[7] M. M. Hossain, M. Fotouhi and R. Hasan. "*Towards an Analysis of Security Issues, Challenges, and Open Problems in the Internet of Things*". In: 2015 IEEE World Congress on Services. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, pp. 21-28, 2015.

[8] A. Khraisat and A. Alazab. "A critical review of intrusion detection systems in the internet of things: Techniques, deployment strategy, validation strategy, attacks, public datasets and challenges". *Cybersecurity*, vol. 4, no. 1, pp. 1-27, 2021.

[9] N. Mishra and S. Pandya. "Internet of things applications, security challenges, attacks, intrusion detection, and future visions: A systematic review". *IEEE Access*, vol. 9, pp. 59353-59377, 2021.

[10] L. Atzori, A. Iera and G. Morabito. "The internet of things: A survey," *Journal of Computer Network*, vol. 54, no. 15, pp. 2787-2805, 2010.

[11] S. Andreev and Y. Koucheryavy. "*Internet of things, smart spaces, and next generation networking*". vol. 7469. In: Lecture Notes in Computer Science. Springer, Berlin, Germany, p. 464, 2012.

[12] S. J. Kumar and D. R. Patel. "A survey on internet of things: Security and privacy issues". *International Journal of Computer Applications*, vol. 90, no. 11, pp. 20-26, 2014.

[13] J. Du and S. Chao. "*A Study of Information Security for M2M of IOT*". In: 2010 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE). Vol. 3. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, pp. V3-576-V3-579, 2010.

[14] B. Schneier. *Secrets and Lies: Digital Security in a Networked World*. John Wiley and Sons, Hoboken, New Jersey, 2015.

[15] J. M. Kizza. *Guide to computer network security*. Springer, Berlin, Germany, 2013.

[16] M. Taneja. "*An analytics framework to detect compromised IoT devices using mobility behavior*". In: 2013 International Conference on ICT Convergence (ICTC). Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, pp. 38-43, 2013.

[17] G. M. Koien and V. A. Oleshchuk. "*Aspects of Personal Privacy in Communications: Problems, Technology and Solutions*". River Publishers, Denmark, 2013.

[18] N. R. Prasad. "*Threat Model Framework and Methodology for Personal Networks* (*PNs*)". In: 2007 2nd International Conference on Communication Systems Software and Middleware. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, pp. 1-6, 2007.

[19] S. O. Amin, M. S. Siddiqui, C. S. Hong, and J. Choe. "*A novel coding scheme to implement signature based IDS in IP based Sensor Networks*". In: 2009 IFIP/IEEE International Symposium on Integrated Network Management-workshops. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, pp. 269-274, 2009.

[20] J. Deogirikar and A. Vidhate. "*Security Attacks in IoT: A Survey*". In: 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC). Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, pp. 32-37, 2017.

[21] S. Ansari, S. Rajeev and H. S. Chandrashekar. "Packet sniffing: A brief introduction". *IEEE Potentials*, vol. 21, no. 5, pp. 17-19, 2003.

[22] L. Liang, K. Zheng, Q. Sheng and X. Huang. "*A Denial of Service Attack Method for an IoT System*". In: 2016 8th International Conference on Information Technology in Medicine and Education (ITME). Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, pp. 360-364, 2016.

[23] C. Wilson. "*Botnets, Cybercrime, and Cyberterrorism: Vulnerabilities and Policy Issues for Congress*". Library of Congress, Congressional Research Service, Washington, DC, 2008.

[24] K. Tsiknas, D. Taketzis, K. Demertzis, and C. Skianis. "Cyber threats to industrial IoT: A survey on attacks and countermeasures". *IoT*, vol. 2, no. 1, pp. 163-186, 2021.

[25] N. Chakraborty and B. Research. "Intrusion detection system and intrusion prevention system: A comparative study". *International Journal of Computing and Business Research*, vol. 4, no. 2, pp. 1-8, 2013.

[26] N. Das, T. Sarkar. "Survey on host and network-based intrusion detection system". *International Journal of Advanced Networking and Applications*, vol. 6, no. 2, p. 2266, 2014.

[27] S. Raza, L. Wallgren and T. Voigt. "SVELTE: Real-time intrusion detection in the internet of things". *Ad Hoc Networks*, vol. 11, no. 8, pp. 2661-2674, 2013.

[28] P. Y. Chen, S. M. Cheng and K. C. Chen. "Information fusion to defend intentional attack in internet of things". *IEEE Internet of Things Journal*, vol. 1, no. 4, pp. 337-348, 2014.

[29] P. Pongle and G. Chavan. "Real time intrusion and wormhole attack detection in internet of things". *International Journal of Computer Applications*, vol. 121, no. 9, pp. 1-9. 2015.

[30] C. Cervantes, D. Poplade, M. Nogueira and A. Santos. "*Detection of Sinkhole Attacks for Supporting Secure Routing on 6LoWPAN for Internet of Things*". In: 2015 IFIP/IEEE International Symposium on Integrated Network Management (IM). Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, pp. 606-611, 2015.

[31] D. H. Summerville, K. M. Zach and Y. Chen. "*Ultra-lightweight Deep Packet Anomaly Detection for Internet of Things devices*". In: 2015 IEEE 34th International Performance Computing and Communications Conference (IPCCC). Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, pp. 1-8, 2015.

[32] V. Eliseev and A. Gurina. "*Algorithms for Network Server Anomaly Behavior Detection without Traffic Content Inspection*". In: Proceedings of the 9th International Conference on Security of Information and Networks. Association for Computing Machinery, New York, pp. 67-71, 2016.

[33] S. O. Amin, M. S. Siddiqui, C. S. Hong and S. Lee. "Implementing signature based IDS in IP-based sensor networks with the help of signature-codes". *IEICE Transactions on Communications*, vol. 93,

no. 2, pp. 389-391, 2010.

[34] D. Oh, D. Kim and W. W. Ro. "A malicious pattern detection engine for embedded security systems in the internet of things". *Sensors*, vol. 14, no. 12, pp. 24188-24211, 2014.

[35] H. Sun, X. Wang, R. Buyya and J. Su. "CloudEyes: Cloud-based malware detection with reversible sketch for resource-constrained internet of things (IoT) devices". *Journal of Software Practice and Experience*, vol. 47, no. 3, pp. 421-441, 2017.

[36] L. Santos, C. Rabadao and R. Gonçalves. "*Intrusion Detection Systems in Internet of Things: A Literature Review*". In: 2018 13th Iberian Conference on Information Systems and Technologies (CISTI). Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, pp. 1-7, 2018.

[37] F. Ahmed, Y. B. Ko. "Mitigation of black hole attacks in routing protocol for low power and lossy networks". *Security and Communication Network*s, vol. 9, no. 18, pp. 5143-5154, 2016.

[38] Y. Xia, H. Lin and L. Xu, "*An AGV Mechanism Based Secure Routing Protocol for Internet of Things*". In: 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, pp. 662-666, 2015.

[39] A. Le, J. Loo, K. K. Chai and M. Aiash. "A specification-based IDS for detecting attacks on RPL-based network topology". *Information*, vol. 7, no. 2, p. 25, 2016.

[40] M. Surendar and A. Umamakeswari. "*InDReS: An Intrusion Detection and Response System for Internet of Things with 6LoWPAN*." In: 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET). Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, pp. 1903-1908, 2016.

[41] Q. D. La, T. Q. S. Quek, J. Lee, S. Jin and H. Zhu. "Deceptive attack and defense game in honeypot-enabled networks for the internet of things". *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 1025-1035, 2016.

[42] H. Sedjelmaci, S. M. Senouci and M. Al-Bahri. "*A Lightweight Anomaly Detection Technique for Low-resource IoT Devices: A Game-theoretic Methodology*". In: 2016 IEEE International Conference on Communications (ICC). Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, pp. 1-6 2016.

[43] P. Kasinathan, C. Pastrone, M. A. Spirito and M. Vinkovits. "*Denial-of-Service detection in 6LoWPAN based Internet of Things*." In: 2013 IEEE 9th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob). Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, pp. 600-607, 2013.

[44] D. Midi, A. Rullo, A. Mudgerikar, and E. Bertino. "*Kalis-a System for Knowledge-driven Adaptable Intrusion Detection for the Internet of Things*". In: 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS). IEEE. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, pp. 656-666, 2017.

[45] T. Matsunaga, K. Toyoda and I. Sasase. "Low false alarm attackers detection in RPL by considering timing inconstancy between the rank measurements". *IEICE Communications Express*, vol. 4, no. 2, pp. 44-49, 2015.

[46] M. Praveena and V. Jaiganesh. "A literature review on supervised machine learning algorithms and boosting process". *International Journal of Computer Applications*, vol. 169, no. 8, pp. 32-35, 2017.

[47] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani. "*A Detailed Analysis of the KDD CUP 99 Data Set*". In: 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, pp. 1-6, 2009.

[48] N. Moustafa and J. Slay. "*UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems* (*UNSW-NB15 Network Data Set*)". In: 2015 Military Communications and Information Systems Conference (MilCIS). IEEE. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, pp. 1-6, 2015.

[49] I. Sharafaldin, A. H. Lashkari and A. A. Ghorbani. "*Toward generating a new intrusion detection dataset and intrusion traffic characterization*".In: The International Conference on Information Systems Security and Privacy. vol. 1, pp. 108-116, 2018.

[50] N. Koroniotis, N. Moustafa, E. Sitnikova and B. Turnbull. "Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-IoT dataset". *Future Generation Computer Systems*, vol. 100, pp. 779-796, 2019.

[51] F. X. Aubet. "Machine Learning-Based Adaptive Anomaly Detection in Smart Spaces". B.Sc. Thesis, Department of Informatics, Technische Universität München, Germany, 2018.

[52] I. Ullah and Q. H. Mahmoud. "*A Scheme for Generating a Dataset for Anomalous Activity Detection in IoT Networks*". In: Canadian Conference on Artificial Intelligence. Springer, Berlin, Germany, pp. 508-520, 2020.

[53] A. Churcher, R. Ullah, J. Ahmad, S. U. Rehman, F. Masood, M. Gogate, F. Alqahtani, B. Nour and W. J. Buchanan. "An experimental analysis of attack classification using machine learning in IoT networks". *Sensors*, vol. 21, no. 2, p. 446, 2021.

[54] R. Olivas. "Decision Trees," Rafael Olivas, San Francisco, 2007.

[55] M. Ahmad, Q. Riaz, M. Zeeshan, H. Tahir, S. A. Haider, M. S. Khan. "Intrusion detection in internet of things using supervised machine learning based on application and transport layer features using UNSW-NB15 data-set". *Journal on Wireless Communications and Networking*, vol. 2021, no. 1, pp. 1-23, 2021.

[56] J. Dou, A. P. Yunus, D. T. Bui, A. Merghadi, M. Sahana, Z. Zhu, C. W. Chen, Z. Han, B. T. Pham. "Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan". *Landslide*, vol. 17, no. 3, pp. 641-658, 2020.

[57] T. Saranya, S. Sridevi, C. Deisy, T. D. Chung, and M. K. A. A. Khan. "Performance analysis of machine learning algorithms in intrusion detection system: A review". *Procedia Computer Science*, vol. 171, pp. 1251-1260, 2020.

[58] M. Shorfuzzaman. "*Detection of Cyber Attacks in IoT using Tree-based Ensemble and Feedforward Neural Network*". In: 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC). Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, pp. 2601-2606, 2020.

[59] D. L. Streiner and G. R. Norman. "Precision" and "accuracy": Two terms that are neither". *Journal of Clinical Epidemiology*, vol. 59, no. 4, pp. 327-330, 2006.

[60] D. Chicco and G. Jurman. "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation". *BMC Genomics*, vol. 21, no. 1, p. 6, 2020.

[61] W. Ma and M. A. Lejeune. "A distributionally robust area under curve maximization model". *Operations Research Letters*, vol. 48, no. 4, pp. 460-466, 2020.

[62] M. Hasan, M. M. Islam, M. I. I. Zarif and M. M. A. Hashem. "Attack and anomaly detection in IoT sensors in IoT sites using machine

learning approaches". *Internet of Things*, vol. 7, p. 100059, 2019.

[63] I. Alrashdi, A. Alqazzaz, E. Aloufi, R. Alharthi, M. Zohdy and H. Ming. "*Ad-iot: Anomaly Detection of IOT Cyberattacks in Smart City Using Machine Learning*". In: 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC). Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, pp. 0305-0310, 2019.

[64] S. Fenanir, F. Semchedine and A. Baadache. "A machine learning-based lightweight intrusion detection system for the internet of things". *Revue D Intelligence Artificielle*, vol. 33, no. 3, pp. 203-211, 2019.

[65] I. Ullah and Q. H. Mahmoud. "*A Two-level Hybrid Model for Anomalous Activity Detection in IoT Networks*". In: 2019 16th IEEE Annual Consumer Communications and Networking Conference (CCNC). Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, pp. 1-6, 2019.

[66] A. Verma and V. Ranga. "Machine learning based intrusion detection systems for IoT applications". *Wireless Personal Communications*, vol. 111, no. 4, pp. 2287-2310, 2020.

[67] V. Kumar, A. K. Das, and D. Sinha. "UIDS: A unified intrusion detection system for IoT environment". *Evolutionary Intelligence*, vol. 14, no. 1, pp. 47-59, 2021.

[68] J. Alsamiri and K. Alsubhi. "Internet of things cyber attacks detection using machine learning". *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 12, pp. 628-634, 2019.

[69] A. R. Arko, S. H. Khan, A. Preety and M. H. Biswas. "*Anomaly Detection In IoT using Machine Learning Algorithms*". Brac University, Bangladesh, 2019.

[70] K. V. V. N. L. S. Kiran, R. N. K. Devisetty, N. P. Kalyan, K. Mukundini, and R. Karthi. "Building a intrusion detection system for IoT environment using machine learning techniques". *Procedia Computer Science*, vol. 171, pp. 2372-2379, 2020.

[71] P. Maniriho, E. Niyigaba, Z. Bizimana, V. Twiringiyimana, L. J. Mahoro and T. Ahmad. "*Anomaly-based Intrusion Detection Approach for IOT Networks Using Machine Learning*". In: 2020 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM). Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, pp. 303-308, 2020.

[72] N. P. Owoh, M. M. Singh, Z. F. Zaaba, and Applications. "A hybrid intrusion detection model for identification of threats in internet of things environment".*International Journal of Advanced Computer Science and Applications*, vol. 12, no. 9, pp. 689-697, 2021.

[73] K. Albulayhi, A. A. Smadi, F. T. Sheldon and R. K. Abercrombie. "IoT intrusion detection taxonomy, reference architecture, and analyses". *Sensors*, vol. 21, no. 19, p. 6432, 2021.

[74] R. Qaddoura, A. M. Al-Zoubi, H. Faris and I. Almomani. "A multi-layer classification approach for intrusion detection in iot networks based on deep learning". *Sensors*, vol. 21, no. 9, p. 2987, 2021.

[75] M. Anwer, S. M. Khan, M. U. Farooq and W. Nazir. "Attack detection in IoT using Machine Learning". *Engineering Technology and Applied Science Research*, vol. 11, no. 3, pp. 7273-7278, 2021.

[76] S. Roy, J. Li, B. J. Choi and Y. Bai. "A lightweight supervised intrusion detection mechanism for IoT networks". *Future Generation Computer Systems*, vol. 127, pp. 276-285, 2022.

[77] J. Su, S. He and Y. Wu. "Features selection and prediction for IoT attacks". *High Confidence Computing*, vol. 2, no. 2, p. 100047, 2022.

[78] A. A. Alsulami, Q. Abu Al-Haija, A. Tayeb, and A. Alqahtani, "An Intrusion Detection and Classification System for IoT Traffic with Improved Data Engineering". *Applied Sciences*, vol. 12, no. 23, p. 12336, 2022.

[79] K. Albulayhi, Q. A. Al-Haija, S. A. Alsuhibany, A. A. Jillepalli, M. Ashrafuzzaman and F. T. Sheldon. "IoT Intrusion Detection Using Machine Learning with a Novel High Performing Feature Selection Method". *Applied Sciences*, vol. 12, no. 10, p. 5015, 2022.