Journal of Applied Engineering and Technological Science
Vol 6(1) 2024: 684-703

ATTACKS DETECTION IN INTERNET OF THINGS USING MACHINE LEARNING TECHNIQUES: A REVIEW

Amer Dawood Saleem*[1], Amer Abdulmajeed Abdulrahman[2]
Computer Science Dep., College of Science, University of Baghdad, Baghdad, Iraq[12]
aamer.dawood2201m@sc.uobaghdad.edu.iq[1], amer.abdulrahman@sc.uobaghdad.edu.iq[2]

Received: 21 April 2024, Revised: 28 July 2024, Accepted: 29 July 2024
**Corresponding Author*

ABSTRACT
The proliferation of IoT devices across sectors such as home automation, business, healthcare, and transportation has led to the generation of vast amounts of sensitive data. This widespread adoption has introduced significant security challenges and vulnerabilities. This study aims to analyze and evaluate machine learning (ML) and deep learning (DL) models for detecting malicious activities in IoT networks, with a focus on improving cybersecurity measures. We conducted a comprehensive review of various ML and DL models, including Random Forest, Decision Tree, HTA-GAN, Hybrid CNN-LSTM, and SVM. The study also includes an evaluation of the datasets used for identifying harmful data, ensuring effective detection of large-scale attacks in IoT ecosystems. Our findings indicate that these models enhance IoT security by deploying efficient intrusion detection systems (IDS) using reliable, large-scale datasets. The study highlights the performance of these models in balancing security and resource management, given the constraints of IoT devices.ML and DL approaches offer significant security benefits for IoT networks, despite the challenges associated with their implementation. The study underscores the importance of future research to address these challenges and further improve IoT security. The results provide valuable insights into the application of ML/DL models in IoT security, contributing to both theoretical knowledge and practical solutions for enhancing cybersecurity in IoT ecosystems.
Keywords : *IoT Devices, Machine Learning, Security Attacks, Deep Learning, Intrusion Detection Systems.*

1. Introduction

The twenty-first century has witnessed significant advancements in computer networks, particularly in the realm of wireless communications and connectivity. The term "Internet of Things" (IoT) was initially coined by Kevin Ashton in 1999(Taherdoost, 2023) . The IoT refers to a network composed of interconnected objects and information-sensing services, such as infrared and global positioning systems, that can communicate and exchange information with each other(Al-Hadhrami & Hussain, 2021a).

The growing popularity of IoT devices has brought numerous benefits to our daily lives, but it has also introduced new security challenges. The connectivity and data transmission capabilities of IoT devices make them vulnerable to complex and large-scale cyberattacks, such as Distributed Denial of Service (DDoS) attacks targeting critical websites and infrastructure(Ahmad et al., 2022b). For instance, notable cyberattacks on IoT devices have resulted in significant breaches and disruptions. The lack of security measures and dedicated anomaly detection systems in heterogeneous IoT networks contributes to their vulnerability to various attacks, including flooding, spoofing, disruption of service, and energy drain. These attacks can have severe consequences, ranging from the malfunctioning of devices to threats to human life, such as compromising a wireless car steering wheel or disrupting the oxygen supply in a medical device(Ahmad et al., 2022b) .

With the increasing number of IoT devices containing sensitive, private, and valuable customer data, security has become a major concern. Most IoT devices have limited resources (battery, bandwidth, memory, computation), making traditional algorithm-based and highly adjustable security approaches impractical. Machine learning (ML) and deep learning (DL) based technologies offer a promising solution to enhance the security of IoT devices. ML is an advanced form of artificial intelligence that operates without explicit programming and can function in dynamic networks. ML techniques can train machines to identify and classify

various types of threats, detect sophisticated and initial-level assaults, and potentially identify new attacks through their learning capabilities (Anthi et al., 2019).

Despite the potential of ML and DL approaches, there remains a significant gap between the security needs and the actual implementation within the IoT device environment(Gatea & Hameed, 2022). This study aims to address this gap by proposing a system that monitors the network, identifies attacks, detects anomalies, especially large-scale attacks, analyzes the behavior of attacks on applications in real time, and utilizes ML and DL methods. The dataset plays a crucial role in intrusion detection systems, as constraints in computing capabilities and the presence of heterogeneity in hardware, software, and security protocols contribute to the lack of safety in IoT devices. For instance, a survey by Synopsys in May 2017 found that 67% of medical device manufacturers believe an attack on a medical device is probable, yet only 17% are actively implementing the required measures to prevent such attacks (Anthi et al., 2019).

Research Gap: There is a notable gap between the security requirements and the security capabilities of existing IoT devices, mainly due to their limited processing capability and the heterogeneity in terminology for devices, software, and protocols.

Contribution of this Review Paper:

1. The review presents different ML and DL techniques and their applications to address common IoT attacks. It includes comparisons and summary tables for ML and DL approaches, sharing lessons learned.
2. A comprehensive evaluation of modern security solutions for IoT devices is provided, focusing on utilizing various ML methods to construct a security model for identifying and classifying attacks.
3. The authors describe the basic challenges and limitations of IoT devices and ML algorithms.
4. This study examines existing solutions applied to mitigate typical attacks on IoT networks and explores their constraints from the standpoint of limited-capacity devices.

2. Literature Review

The proliferation of Internet of Things (IoT) devices, the growth of interconnected networks, the diverse range of devices, and their integration into various sectors such as healthcare and transportation have attracted a significant number of attackers. These attackers aim to exploit vulnerabilities in networks to illegally acquire sensitive data, potentially compromising the functionality of these networks. As a result, researchers have put forth various methodologies employing machine learning (ML) and deep learning (DL) to mitigate anomalous behaviors within the network.

Bi-directional Long Short-Term Memory Recurrent Neural Network (BLSTM RNN): The authors in(Roy & Cheung, 2018) present a deep learning methodology for identifying intrusions in IoT networks using BLSTM RNN. The proposed technique is trained using a multi-layer Deep Learning Neural Network on the benchmark dataset UNSW-NB15. The main focus of the paper is the binary categorization of IoT network attacks and normal patterns, with the reported accuracy of the classifier being 95.71%. This method shows promise but may face challenges in scalability and generalization across diverse IoT environments.

Intrusion Detection System (IDS): In their study, the researchers in(Anthi et al., 2019) proposed an IDS employing a supervised method for detecting common network-based cyber-attacks on IoT networks. The IDS is designed to detect the normal behavior of IoT devices, discover malicious packets, and identify the type of attack. Evaluated using 12 types of attacks, the system's core functions achieve an F-measure of 96.2%, 90.0%, and 98.0%, respectively. This showcases the IDS's capability to autonomously differentiate between malicious and benign activities, although the complexity of real-world IoT environments may pose additional challenges.

Lightweight Fuzzy Logic-Based Intrusion Detection (secure-MQTT): By presenting a proposed model that includes a lightweight fuzzy logic-based intrusion detection scheme, the vulnerabilities in MQTT communication between IoT devices are addressed in(Fadhil et al.,

2021). The proposed secure-MQTT method controls intruder activities with low configuration during communication. This approach highlights the need for lightweight solutions tailored to the resource constraints of IoT devices.

Multilayer Attack Detection System (IoT DDOS): The authors (Chen et al., 2020) developed a multilayer attack detection system using ML techniques, specifically targeting DDoS attacks. The system employs feature extraction methods tailored to different types of DDoS attacks and proposes an IoT authentication mechanism for non-IP devices. The system achieves an accuracy level of 97% and detects attacks within 0.36 seconds. While effective, the reliance on specific datasets and attack types may limit its generalizability.

Distributed System Architecture for IDS: In (Larriva-Novo et al., 2020), a distributed system architecture is proposed to manage extensive datasets, such as UGR16, for developing IDS. The study employs ML techniques to improve the pre-processing model's response and effectiveness. The model achieves high accuracy rates for DOS attacks (99.97%) and other types of attacks but shows low effectiveness for botnet and blacklist traffic detection. This highlights the ongoing challenge of developing comprehensive solutions that address all attack types.

Deep Neural Network (DNN) for NIDS: The paper(Hussien & Dhannoon, 2020a) presents a method for identifying anomalies in Network Intrusion Detection Systems (NIDS) using DNN methodology. Dropout is used as a regularization technique to reduce overfitting. The study achieves 99.45% accuracy on the NSL-KDD dataset, demonstrating the potential of DNNs in intrusion detection. However, the application of such models to real-time and diverse IoT environments requires further investigation.

Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) Model: The researchers in (Sahu et al., 2021) introduced a security measure utilizing a CNN to extract data features and an LSTM for classification. The study, using data from intentionally infected Raspberry Pi devices, achieved a notable accuracy rate of 96%. This method effectively combines feature extraction and sequence learning, though its applicability to larger, more diverse datasets needs further validation.

Review of IoT Security Vulnerabilities and DDOS Attacks: The research in(Mishra & Pandya, 2021) provides a multi-faceted review of IoT security vulnerabilities and focuses on DDOS attacks. The review covers the impact, solutions, and anomaly detection methods for DDOS attacks, emphasizing the critical need for robust security measures across all IoT layers.

Generative Adversarial Network (GAN) for Anomaly Detection: The study in(Chen et al., 2022) proposes a GAN-based predictive model for anomaly detection in high-dimensional data. Evaluated on synthetic datasets and popular anomaly benchmarks, the GAN-based methods showed average performance, with HTA-GAN demonstrating the best detection results. This highlights the potential and limitations of GANs in anomaly detection, suggesting room for improvement.

Intrusion Detection System Using Data Mining Algorithms: The paper(Gharkan & Abdulrahman, 2023) proposes an IDS for detecting malicious traffic using data mining algorithms. Evaluated on the CICDDoS2019 dataset, the Random Forest algorithm outperforms other models, achieving higher accuracy. This study underscores the importance of selecting appropriate algorithms for specific datasets and attack types.

Botnet Attack Detection Model Using ML: In (Alissa et al., 2022) ,the authors developed a botnet attack detection model using the Bot-IoT and UNSW datasets. Naïve Bayes, K-Nearest Neighbour, Support Vector Machine, and Decision Trees were employed, with Decision Trees achieving the highest accuracy (99.89%). This study demonstrates the effectiveness of feature selection and dimensionality reduction techniques in improving ML model performance.

Framework for Malicious Network Traffic Detection: The authors in(Anwer et al., 2021) developed a framework for detecting malicious network traffic using a Support Vector Machine (SVM), gradient-boosted decision trees (GBDT), and Random Forest (RF). The RF algorithm achieved a significantly higher accuracy (85.34%). This highlights the importance of employing robust classification techniques in network security.

The table below illustrates the summarization of the state-of-the-art models of cybersecurity attacks.

Table 1 -Summarization of The State-of-the-art Models In Cybersecurity.

Reference	Dataset	Method	Attacks type	Accuracy
(Roy & Cheung, 2018)	UNSW-NB15	Bi-directional long short-short-term memory Recurrent Neural network (BlSTM RNN)	Analysis Backdoor Denial-of-Service Worms Reconnaissance	95%
(Anthi et al., 2019)	"IoT-23" dataset	Naïve Bayes Bayesian Network J48 Decision Tree Zero R OneR Simple Logistic Support V M MLP Random Forest	Reconnaissance (Quick Scan, Intense Scan) DoS/DDoS (TCP Flood, UDP Flood, Hello flood) MITM (Ettercap, SSL Strip, Burp suit) Replay (MITM, framework suite) Spoofing (DNS, ARP)	(79%) (96%) (98.8%) (17%) (79.0%) (96.0%) (89.0%) (N/A) (96.0%)
(Larriva-Novo et al., 2020)	UGR’16	multilayer perceptron (MLP) decision tree(DT)	DOS Netscan spam Botnet Blacklist No attack	D.T(99.97 %) D.T(98.94 %) D.T(94.91 %) D.T (36.62%) D.T (1.01%) D.T (99.98%)
(Shafiq et al., 2020)	Bot-IoT dataset	Bayes Net Decision tree Naive Bayes Random forest Random tree	Information gathering (service and OS) TCP, UDP DoS/DDOS HTTP DoS/DDoS, Information theft (Keylogging, theft data)	(99.77%) (99.79%) (99.79%) (99.99%) (99.99%)
(Chen et al., 2020)	datasets(wireless or wired networks).	a multi-layer system using “decision tree” and SVM(“Support Vector Machine”)	DDoS data flood, ICMP flood, SYN flood, UDP flood	0.97%
(Alissa et al., 2022b),	Bot-IoT UNSW	DT Naive Bayes K-NN SVM	DDoS, DoS, OS Scan, DDoS, DoS, OS Scan, DDOS DOS Scan OS Fuzzers, Backdoor, Reconnaissance and Worm attacks Backdoor Fuzzers Worms Reconnaissance	99.87% 85.91% 82.14% 99.25%

(Anwer et al., 2021)	NSL KDD	SVM RF GBDT	Unauthorized to remote (R2L) Denial-of-Service (DoS Unauthorized to root super user privileges (U2R attack) Port scanning attack (Probe)	SVM, GDBT(32.38) RF(85.34)
(Sahu et al., 2021)2021	IOT-23	Model Architecture Using Hybrid Convolution Neural Network and Long Short-Term Memory Models. (CNN-LSTMM	Command and Control DDOS File Download Heart Beat Part Of A Horizontal Port Scan Mirai Torii Okiru	0.96%
(Tharewal et al., 2022	real dataset of the natural gas pipeline	"Deep Reinforcement Learning IDS intrusion detection model". DRL-IDS	NMRI CMRI MSCI MPCI MFCI DOS Reconnaissance	0.99 %
(Chen et al., 2022	'verteral' 'optdigits' EURO-Arg Synthetic (1,2,3,4)	"a novel GAN-based predictive model, called HTA-GAN"	detecting operational anomalies in large-scale IoT data infrastructures using multivariate time series data.	(0.97%) 0.57% 0.49 (0.97,0.76 %, 0.76%, 0.97%)
(Neto et al., 2023) 2023	CICIOT2023	Logistic Regression perceptron Adobos Random Forest DeepNeuraNetwork (DNN)	DDOS (Family) DOS(Family) Recon(Family) Web-Based(Family) Brute force(Family) Spoofing (Family) Mirai (Family)	0.80% 0,.81% 0.60% 0.99% 0.98%
(Gharkan & Abdulrahman, 2023)	CICDDoS2019	Naive Bayes random forests decision tree logistic regression	DDOS	90.04% 99.98% 99.98% 99.97%

2.1 Dataset analysis

Nowadays, the new technology and the digitalization with the fast movement of technology and high-speed internet raise several issues related to the IoT environment. To address and improve the problems that face the researchers several datasets have been demonstrated for this purpose.

The UNSW-NB15 and N-BaIoT datasets are network traffic datasets that were utilized by the authors in (Ahmad et al., 2022a) in the domains of network security and intrusion detection within the framework of the IoT. The authors of this paper evaluate the efficacy of different deep learning models in detecting IoT attacks across multiple classes. The classes include normal traffic, analysis, backdoor, DoS, exploitation, fuzzing, generic, reconnaissance, shellcode, and worm. The dataset exhibits several strengths, including the presence of realistic network traffic, diverse attacks, labeled data, and a large scale. However, it also possesses weaknesses. fixed dataset restricted IoT depiction, The datasets offer a range of characteristics that can be employed to train deep learning models for intrusion detection. They are suitable for evaluating the efficiency of intrusion detection systems because of their wide array of attack types and network protocols. The occurrence of certain attack types is significantly more frequent than others due to the extreme imbalance in the datasets.

The dataset in (Anthi et al., 2019) was created in controlled conditions; it collected 3 weeks of benign and 2 weeks of malicious data, allowing for large data collection. The dataset

covers DoS, DDOS/botnets, man-in-the-middle, spoofing, insecure firmware, and data leakage attacks on IoT devices and networks. This dataset enhances the proposed IDS. A drawback of the dataset is that the attacks used during data collection were off-the-shelf and may not represent real-world attacks that are more complex and harder to detect.

The UGR'16 (Larriva-Novo et al., 2020) is a dataset that includes attack types (e.g., denial of service, spam, network scan, blacklisted, botnet, and no attack). The dataset strikes a balance between network traffic and attack. It has a lot of information, is sensor-based, and includes network sensor traffic. This dataset is modern and large-scale, but it has limitations such as high instances and low attack instances, redundant features, and missing some types of attacks.

(Shafiq et al., 2020) used a bot-IoT dataset to choose an efficient machine learning (ML) algorithm for detecting bot-IoT traffic in the IoT network environment, as well as identifying abnormal and intrusion traffic patterns in an IoT network. The dataset comprises various categories of IoT traffic flows, including normal traffic, IoT traffic, and multiple variants of botnet attacks. The primary strength of the dataset for effective Bot-IoT malicious traffic identification lies in its focus on IoT security, labeled data, and realistic IoT scenarios. However, the dataset has several weaknesses, including limited representation, a static nature, a lack of information about network events, limited types of attacks, and a limited size.

Using real IoT scenarios (Chen et al., 2020), the authors launch DDOS attacks on eight smart poles and extract features of four types of DDoS attacks: sensor data floods, ICMP floods, SYN floods, and UDP floods. The dataset's strengths include realistic simulation, diverse attacks, and potential uniqueness due to its limited size, missing values, and specific environment.

The dataset utilized in (Tharewal et al., 2022) is the genuine dataset of the natural gas pipeline. The dataset includes various attack types, namely: “simple malicious response injection (NMRI)”, “complex malicious response injection (CMRI)”, “malicious status command injection (MSCI)”, “malicious parameter command injection (MPCI)”, “malicious function command injection (MFCI)”,(DDoS), and reconnaissance attack. Partition the dataset into three distinct sections for the purpose of conducting the experiment: The training set comprised 60% of the dataset, while the testing set comprised 20%. The dataset consists of standard network traffic data.

The CICIoT2023 dataset, which comprises 33 attacks against IoT devices categorized into 7 classes, is a novel and extensive benchmark for large-scale attacks in IoT environments used in this study (Neto et al., 2023) . With a thorough explanation of the testbed utilized and the framework for generating the dataset, the Edge-IIoTSet dataset is a realistic and comprehensive cybersecurity resource for IoT and IoT applications. The complexity and diversity of actual IoT environments may not be fully reflected in this dataset since it is based on simulated environments. The NSL-KDD dataset is a collection of network traffic data. The dataset comprises normal instances as well as different attack categories, including root-to-local (R2L), probing, user-to-root (U2R), and Denial of Service (DoS). The experiments in the studies (Hussien & Dhannoon, 2020b) (Ahmim et al., 2018) (Kevric et al., 2017) (Al-Yaseen et al., 2017)demonstrated that the proposed model can acquire expertise in real-time, regardless of whether the feature selection method is utilized or not.

2.2 IoT layers

There is no common architecture for IoT devices across different vendors, and different manufacturers have different designs for their IoT architecture layers. Since the three-layer architecture is the most popular among researchers, it is the focus of this study.

1. Physical layer: The initial layer of the IoT architecture handles node communication in addition to collecting data from sensors. Various technologies operating within this layer e.g., RFID, Zigbee, Bluetooth, etc(Tahsien et al., 2020).
2. Network layer: This layer serves as a mediator, facilitating data exchanges between various nodes and the application and physical layers. Among the protocols that function in this layer are TCP/UDP, IPv6, RPL, WIFI, IEEE 802.15.4 and 6lowpan which link devices to intelligent services (Singh et al., 2019), There are local clouds and servers in the network

layer. Functions as an intermediary between the network and the subsequent layer, storing and processing the data (Ali et al., 2023). Big data attracts an expanding economic market and is a crucial component at the network layer. Massive amounts of data are produced by physical layer sensors, which are then sent and processed by IoT devices. The data and information are then used for intelligent services in the network layer, such as deep learning (DL) and machine learning (ML). These days, a lot of people use it for analyzing data and make use of the best analysis methods for use on smart devices (Lv & Singh, 2021).

3. Application layer: This layer is in charge of data representation and is allowed to access data on IoT devices by other protocols like HTTP, COaP, and MQTT. That it the location where users and devices interact (Bhuiyan et al., 2021). Below [figure 1] demonstrates the IoT architecture design(Al-Hadhrami & Hussain, 2021a).

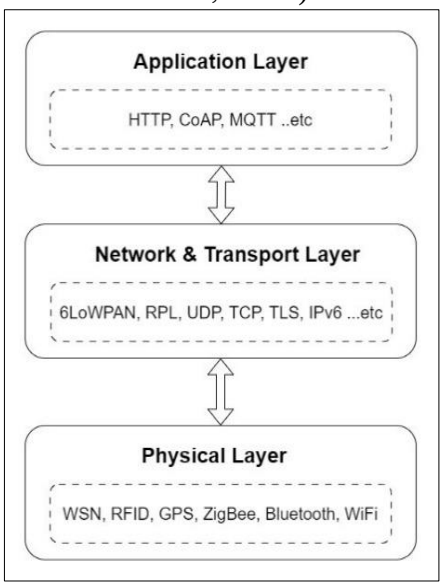

Fig. 1. IoT Three-layer architecture.

The above diagram shows the IOT stack architecture. Most studies focus on these layers, and this figure explains how each layer's protocol's function, how they integrate into the operation of existing networks, and how it provides a thorough overview and clarity of these layers(Tahsien et al., 2020).

2.3 Importance of security in IoT

IoT devices are available to everyone and are used for a variety of purposes via an open network. However, because of the potential for numerous threats and attacks, the IoT also works to improve user privacy while advancing human life through technological means (Alotaibi & Oracevic, 2023) . Certain IoT devices are accessible to anybody; anywhere. IoT device security may be compromised without user authorization. The question has grown urgent. To safeguard IoT devices, a variety of security measures should be put in place. However, the IoT infrastructure restricts the computational capabilities of these devices, which restricts applying in place an advanced security protocol (Butun et al., 2020).

2.4 Requirement of security and goals

In order to attain the desired outcome, the implementation of security protocols is necessary. The prevailing security and assurance paradigm, known as the Confidentiality, Integrity, and Availability (CIA) model, encompasses three main needs.

1. Confidentiality: The safeguarding of sensitive data against unauthorized access during data transfer or storage is of crucial significance. This is particularly critical in the context of IoT devices, which often contain sensitive applications such as healthcare systems. The potential compromise of these systems through hacking is a significant risk, as it may result in the loss of human lives (Ferrag et al., 2020).
2. Integrity: One of the most important factors in enabling the successful integration of IoT devices is the preservation of data integrity during transmission. In many different contexts

and applications, it is accepted as a basic security requirement. The alteration or manipulation of data during the process of transmission might give rise to potential risks, particularly in the healthcare industry where sensitive information, like blood pressure, heart rate, and oxygen saturation levels, is handled. Sensitive data is regarded as confidential information for both patients and doctors (Ferrag et al., 2020; Musleh et al., 2020).

3. Availability: Ensuring the continuous availability of the IoT within the network is a crucial security objective that necessitates attainment. The issue is typically discussed in the context of attacks on availability. Distributed Denial of Service (DDoS) assaults pose a significant threat within this context, restricting access to devices and services and resulting in substantial financial and operational failures. Therefore, it is essential to effectively mitigate such attacks (Al-Hadhrami & Hussain, 2021b).

2.5 IoT Device Constraints and Limitations

Like any other computer network, the Internet of Things involves several types of security mechanisms. However, IoT security measures must meet criteria that, due to their design and resource constraints, may not apply to different networks.

1. Resource Limitations: One of the challenges that IoT devices face is the lack of resources, such as CPU, memory, and energy(Al-Hadhrami & Hussain, 2021b). This makes it impossible to deploy security solutions that require a lot of processing power(Burhan et al., 2021). In other words, we cannot use traditional computer-based security solutions. From what has been mentioned so far, the IoT ecosystem has minimal network protocols and features that require little processing power, so developers and manufacturers face trouble developing a simple and effective security solution(Oudah & Maolood, 2023) .
2. Privacy and data confidentiality: On the IoT, privacy depends on the importance of the application. For example, the level of privacy of a healthcare application differs from the level of privacy of a weather application. This does not mean neglecting the privacy of some applications. On the contrary, we must harness security in the IoT environment to a greater extent when it comes to user information. Therefore, to preserve that, encryption of information is mandatory, provided that it does not affect how IoT devices work(Yang et al., 2020). The objective of the study conducted by(Taherdoost, 2023) was to propose and establish a specific and reliable data exchange system to ensure the confidentiality and integrity of data transmission.
3. Authentication: It is no secret to us. IoT devices generate a lot of data, so data must be sent securely across terminals by enabling the data authentication method. Unfortunately, because there is no standard technique for authentication, suppliers do not employ similar standards for authentication. As a result, integration across these platforms is poor, resulting in serious security issues(Conti et al., 2018).
4. Service Availability: IoT networking services are susceptible to several Denial of Service (DoS) attacks. Node penetration can occur through internal or external means, and such attacks can have detrimental effects on network functionality, leading to disruptions in all operations and services. These attacks often aim to exhaust the available resources of the device and given that s IoT devices often rely on battery power, they give rise to significant challenges. Ensuring the continuous availability of devices and services is of utmost significance due to the time and data sensitivity exhibited by numerous applications, particularly in the realm of healthcare.
5. Data Management Challenge: The functionality of the IoT depends upon the data derived from sensors. As the volume of data grows, data centers encounter a significant architectural challenge in the management and processing of this data. The significance of the IoT at the institutional level is quite pronounced. The current period is characterized by the generation of vast amounts of data that necessitate processing, analysis, and storage. However, this phenomenon also gives rise to intricate security challenges(Lee & Lee, 2015).

3. Machine learning Techniques

Machine learning is used to train machines in the optimization of data processing. In some cases, understanding the extracted information from data may prove challenging,

requiring the use of machine learning techniques. Machine learning is widely used across several industries for the purpose of extracting relevant information. The primary objective of machine learning is to acquire knowledge and insights from data through computational algorithms and statistical models. Numerous studies have been undertaken. Research has been conducted to explore the methodologies for enabling machines to acquire knowledge and skills autonomously, without the need for explicit programming (Mahesh, 2018).

3.1 Supervised learning

Supervised learning is the predominant approach in machine learning, wherein the output is categorized by utilizing a trained dataset and a learning algorithm, based on the input. Supervised learning is categorized into classification and regression learning (Abdulrahman & Ibrahem, 2021).

3.1.1 Support Vector Machine (SVM):

The SVM algorithm analyses regression and classification data. A hyperplane is created by SVM between two classes. The hyperplane maximizes distance from each class to distinguish each class with the least error at the maximum margin(Sriavstava et al., 2020). SVM's accuracy makes it ideal for IoT security applications like intrusion detection(Al-Garadi et al., 2020).

3.1.2Bayesian Theorem:

It uses statistical probability theory(Jalawkhan & Mustafa, 2021). Using supervised learning to generate new results from past data is a challenge that can be successfully implemented in the IoT. For network intrusion detection in IoT, naive Bayes is often used(Asharf et al., 2020).

3.1.3 K-nearest neighbor (KNN):

KNN refers to a statistical non-parametric method in supervised learning that commonly uses Euclidean distance. The Euclidean distance is determined in KNN The average value of the unknown node is the nearest neighbor. For example, if any node is lost, this can be inferred from the average value of the next neighbor. This value is not accurate But it helps to identify a probable missing node. The KNN approach is utilized in Intrusion detection, virus detection, and anomaly detection in the Internet of Things(Kadhm et al., 2021).

3.1.4 Decision Trees (DTs):

A decision tree is a widely used algorithm in machine learning for classification and prediction. The ID3 algorithm is a commonly used top-down approach for constructing decision trees. These algorithms produce decision rules that can be used to predict the outcome of new test cases. They offer high accuracy and improved interpretability. Additionally, decision trees can handle both continuous and discrete data(Alqahtani et al., 2020).

3.2 Unsupervised learning

This technique involves inputting variables without corresponding output data. Most of the data lacks labels, and the system seeks to determine similarities among these datasets. Partition them into separate clusters. Various types of unsupervised learning IoT devices' security has been enhanced through the implementation of advanced techniques that detect Denial of Service (DoS) attacks using multivariate correlation analysis(Khraisat et al., 2019)

3.2.1K-mean clustering:

In this technique, you create small groups in order to classify data samples into a group. This implementation technique relies on some rules to distinguish between the given data set into different groups, where each group contains a (k-centroid), where the main goal is to determine the k-centroid for each group. Then select the node for each group and connect it to the nearest central point. Continue doing this until each node is connected. After that, a recalculation is performed based on the average value of the node in each group. The method repeats its previous steps until they coincide to obtain a useful k-mean value(Liu et al., 2022).

K-means algorithms are additionally useful. In IoT systems, when labeled data is unnecessary due to its clarity.

3.2.2 Reinforcement learning (RL)

This technology allows the machine to learn to interact with the environment in which it is operating by implementing procedures to maximize total feedback. There are no pre-determined procedures for any particular task. The machine uses trial-and-error methods. Through the implementation of trials and errors, it can be determined and implemented to obtain the highest efficiency(Asharf et al., 2020).

4. Methodology

This study aimed to evaluate the effectiveness of various machine learning (ML) and deep learning (DL) models in detecting malicious activities within IoT networks. The research process consisted of several stages, including data collection, preprocessing, model selection, training, evaluation, and comparison with existing studies.

4.1 Data Collection and Preprocessing:

The datasets used in this study include UNSW-NB15, IoT-23, Bot-IoT, UGR'16, CICIOT2023, CICDDoS2019 and NSL-KDD. These datasets were selected due to their comprehensive representation of different types of attacks and normal network traffic. The data were preprocessed to handle missing values, remove duplicates, and normalize features to ensure consistency and improve model performance.

4.2 Participants and Characteristics:

The study focused on IoT devices operating in various environments, including residential, industrial, healthcare, and transportation sectors. The datasets encompassed different attack types, such as DDoS, spoofing, man-in-the-middle, reconnaissance, and more. The characteristics of the datasets varied in terms of the number of records, attack types, and network protocols.

4.3 Model Selection and Training:

We selected several ML and DL models, including Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Hybrid CNN-LSTM, and Generative Adversarial Network (GAN). These models were chosen based on their proven effectiveness in previous studies. Each model was trained using a portion of the datasets, with hyperparameters tuned to optimize performance.

4.4 Evaluation Metrics:

The models were evaluated using accuracy, precision, recall, F1-score, and confusion matrices. These metrics provided insights into the models' ability to correctly classify normal and attack instances. Cross-validation techniques were employed to ensure the robustness and reliability of the results.

4.5 Selection Method:

The literature reviewed for this study was selected based on relevance, recency, and contribution to the field of IoT security. We included articles published between 2019 and 2023 from reputable international journals. The selection process involved searching databases such as IEEE Xplore, SpringerLink, and ScienceDirect using keywords related to IoT security, ML, and DL.

Table 2 - Machine learning techniques criticism: strength and weakness.

Reference	Title	Strong	Weakness

(Liu et al., 2022)	"A Deep Learning Approach for Intrusion Detection on the IoT using Bi-Directional Long Short-Term Memory Recurrent Neural Network"	It is a novel approach for intrusion detection in IOT networks. the proposed model archives over 95% accuracy in attack detection. The model was trained using the benchmark dataset: UNSW-NB15 The model is efficient.	This proposed BLSTM RNN model should be examined using large datasets.
(Anthi et al., 2019)	"A Supervised Intrusion Detection System for Smart Home IoT Devices"	The results of this suggested architecture are good. This illustrates how the suggested architecture can recognize malicious devices and assaults automatically.	The authors state that because of limitations in processing power and heterogeneity in terms of hardware, software, and protocols, there is a major gap between the security requirements and the security capabilities of currently available IOT devices.
(Larriva-Novo et al., 2020)	"Efficient Distributed Preprocessing Model for Machine Learning-Based Anomaly Detection over Large-Scale Cybersecurity Datasets"	His proposal can aid in enhancing the scalability and accuracy of intrusion detection systems, and the distributed preprocessing design speeds up execution while reducing costs. High accuracy is also achieved by the suggested method in problems involving binary and multi-class categorization.	generate false negative when training MLP with a large dataset. The decision tree algorithm used the IP address as the primary variable, resulting in unsuccessful results due to the high susceptibility of IP addresses to modification during most attacks. The random split method is regarded as ineffective for intrusion detection systems due to its ability to yield a significant false-negative rate.
(Shafiq et al., 2020)	"Selection of effective machine learning algorithm and Bot-IoT attacks traffic identification for the IoT in smart city"	use of a new dataset and multiple machine learning algorithms to detect IoT network anomalies and intrusions.	The effectiveness of the proposed framework and algorithm is assessed using only one dataset, namely the Bot-IoT dataset. This may restrict the applicability of the findings to different IoT network settings.
(Sahu et al., 2021)	"Internet of Things attack detection using hybrid Deep Learning Model"	It detects malicious activity using a hybrid deep learning model that combines convolutional neural networks and long-term short-term memory. The model is trained on a large dataset of IoT device network traffic, improving accuracy and robustness. The model can detect both known and unknown attacks. It is scalable and can be applied to different types of IoT device.	Depending on the amount of data used to train the model, deep learning-evading attacks may be a weakness.
(Ahmad et al., 2022a)	"A comprehensive deep learning benchmark for IoT IDS"	This classifier provides quick convergence and gets optimal outcomes, thus addressing two crucial objectives: performance and accuracy.	In this research, poor-performing classifiers ("Autoencoder and BRNN") that took hours to train appeared.

(Tharewal et al., 2022)	"Intrusion Detection System for Industrial Internet of Things Based on Deep Reinforcement Learning"	The model utilizes GBM's feature selection algorithm to extract the most important feature set from industrial IoT data and merges it with the deep learning algorithm. The model was evaluated on a freely available dataset and demonstrated a 99 percent detection rate for various categories of network attacks. Compared to existing intrusion detection systems that utilize "deep learning models like CNN, RNN, and LSTM" as well as "deep reinforcement learning models" like "DDQN, DQN", the accuracy, precision, recall rate, F1 score, and overall performance are better. decreases the duration of training for intrusion detection models.	The suggested "Deep Reinforcement Learning-based Intrusion Detection System" is described without any flaws or restrictions in the text.
(Chen et al., 2022)	"A novel GAN-based predictive model, called HTA-GAN."	The system has a high level of accuracy in detecting anomalies in complex logs of multivariate time series. The model improves multivariate time series representation learning and BiGAN-based anomaly scoring using GAN's heterogeneous structure. The model can produce data on false anomalies, rendering it appropriate for detecting operational irregularities in of the highest quality data services.	The main disadvantage is that it takes much longer to handle.
(Neto et al., 2023)	"CICIoT2023: A Real-Time Dataset and Benchmark for Large-Scale Attacks in IoT Environment"	A new realistic IoT attack dataset is introduced. The large dataset includes 33 attacks in 7 classes. The proposal highlights the growing importance of IoT and the need for a large dataset to test and evaluate security measures.	The dataset in this paper is simulated and does not reflect real-world IOT complexity and diversity.
(Roy & Cheung, 2018)	Construct an Efficient DDoS Attack Detection System Based on RF-C4.5-GridSearchCV.	The article details the proposed DDoS attack detection system and dataset. Basic supervised classification algorithms were used to accurately classify the attack, and the results were compared to other algorithms. They reduced false positives and improved accuracy to boost production system uptime.	The proposed system's limitations and drawbacks, as well as its cost and feasibility in real network environments, are not discussed.

5. Fact-finding and discussion: Table3

Table 3 - Machine learning techniques: Finding and discussion.

Authors	Method/Technique	Findings	Discussion

(Roy & Cheung, 2018)	"Bi-directional long short-term memory Recurrent Neural network " (BlSTM RNN)	A multi-layer Deep Learning Neural The network is trained using a novel benchmark data set: UNSW- NB15. And focuses on the binary classification of normal and attack.	The model exclusively deals with binary classification, and the dataset includes a restricted range of attacks.
(Anthi et al., 2019)	Naïve Bayes Bayesian Network J48 Decision Tree Zero R OneR Simple Logistic Support V M MLP Random Forest	Feature selection Correlation attributes and a layer Intrusion detection system (IDS) to address Limitations of IoT Security. The model handles both binary and multiclass classification.	A better classifier is the J48 decision tree with a high accuracy 98%. The dataset does not include large-scale attacks or all typical types of attacks.
(Larriva-Novo et al., 2020)	Classifier Perceptron MPL, Decision tree	·Preprocessing for huge datasets using a unique distributed computing architecture. Results are faster and more reliable than neural networks.	The UGR'16 large-scale dataset serves as the main dataset, specifically designed to detect attacks within a real ISP network. Multiclass classification and a better classifier decision tree provide high accuracy when detecting DDOS attacks.
(Shafiq et al., 2020)	Bayes Net , C4.5 decision tree , Naïve Bayes Random Forest Random Tree	Multiple types of data are in this dataset. IoT traffic flows comprise regular and some IoT traffic, Botnet IoT traffic varies. For good ML performance, divide the dataset into categories and subcategories for multiclass classification. Five well-known machine learning classifiers were chosen, and Weka was used to implement the experiment and use bijective soft to choose and decide the good classifier.	While all machine learning algorithms are efficient and fast in building models, Naïve Bayes and Random Tree excel in performance. The most significant and difficult aspect in machine learning is model creation time. the model Naïve Bayes ML technique outperforms Random Tree ML in terms of model processing time.
(Chen et al., 2020)	Decision tree	A multi-layer IoT DDoS attack detection system using machine learning is proposed. he bandwidth of a normal packet on a mirror port is 10 Mbps, while a DDoS attack results in 80–100 Mbps of dataflow. is a multiclass classification and has good accuracy with the chosen classifier(DT). -The system distinguishes DDoS attacks from normal packets. -When suspicious packets are identified, SDN switches block	Only DDoS attacks are detected by this technology, not other IoT attacks. -can provide more classifiers to evaluate this result and use a large-scale dataset to include more types of attacks.

(Alissa et al., 2022)	SVM	DT Naive Bayes K-NN	This study involved the development of a machine learning model to identify robots or malicious traffic behavior, The decision tree outperformed the other classifiers.	Validating the results requires analyzing the entire UNSW-NB15 dataset using the same model. A different dataset and more classifiers were used in the experiment. test the result in real time. the dataset used is not IoT specific, old, and does not contain modern-day attacks.
(Anwer et al., 2021)		SVM RF GBDT	Create a framework that is developed to detect malicious network traffic using three well-known classification-based methods. supervised learning and feature selection.	a good accuracy in the fog layer of the cloud obtained by the RF classifier. The authors did not use IoT-specific datasets. unbalanced classification. detect specific types of attacks. binary and multiclass classification.
(Sahu et al., 2021)	The CNN model utilized in the initial stage of classification obtains the essential features from the network traffic and LSTM classifier (required a little computational process)		The study proposed a novel model is structured as a hybrid deep learning. Model Architecture using Convolution Neural Network and Long. Short-term memory models. A CNN module can be easily added to a new IoT sub-network using the security model. However, it would require little computing because the cost of attack detection is low.	The dataset does not contain all types of common current attacks. The model has been considered a Limited dataset with limited attacks The dataset is not focused on detecting large-scale attacks.
(Tharewal et al., 2022)	“Deep Reinforcement Learning IDS intrusion detection model”. DRL-IDS		DRL GDS proposed in this study performs well in detecting various types of network attacks on the Industrial Internet of Things. It outperforms deep learning and deep reinforcement detection systems in accuracy and training time	suggest using a distributed architecture and machine learning algorithms to identify intrusions in IIoT systems.

(Chen et al., 2022)	"HTA-GAN"	leverages popular GAN-based generative models and end-to-end one-class classification to improve unsupervised anomaly detection BiGAN-based structure improves anomaly score computational efficiency. HTA-GAN outperforms and is more robust than its competitors.	Supervised machine learning cannot utilize enormous amounts of data without labeled data. best performance among four prominent unsupervised methods: KNN, ABOD, IF, and AE. The proposed HTA-GAN is best.
(Neto et al., 2023)	Logistic Regression perceptron Adobos Random Forest DeepNeuraNetwork (DNN)	This study performs 34 class and 8 class and binary class classification. Dataset large-scale attacks and contains all common types of attacks in IoT environments.	The classification was performed on five classifiers in 34 classes and 8 random forest and deep neural network classes, and the result was good. In binary classification, all methods obtained high accuracy in identifying attacks

6. results analysis

This part explains the results obtained from the experimental and comparison studies, which are provided in Table 4. Due to the unavailability of complete setting details in research articles, the impact on classifier performance may deviate slightly when reproducing certain results using the same classifier, dataset, and available settings as stated in the original paper. For instance, while utilizing LSTM, the accuracy was 96.24%. After replicating LSTM on the same dataset, the accuracy increased to 96.99% (an increase of 0.78%). The analysis reported in Table 4 uncovers unexpected findings that further support the research question and hypothesis stated in the introduction section Most research papers have a built-in bias when they compare the performance of their classifiers to a separate benchmark dataset.

Table 4. Comparison of the results of the classifiers using different datasets of previous research.

Ref	Classifier	Dataset	Accuracy
(Chakrabarti & Saha, 2019)	1dCNN	CICIDS2017	95.14
	MLP		86.34
	Lstm		96.24
	CNN+LSTM		97.16
	SVM		95.5
	Bayes		95.19
	RF		94.64
(Ahmad et al., 2022a)	(MLP).	CICIDS2017	99.90%
	(CNN)		100.00%
	(LSTM).		99.90
	Autoencoder + LSTM.		99.90%
	CNN + LSTM.		99.90%
(Costa et al., 2020)	LSTM	UNSW-NB15	99.9

(Ahmad et al., 2022b)	Mlp	Bot-IoT	35.95
	CNN		4.89
	LSTM		38.7
	CNN-LSTM		20.66
(Shafiq et al., 2020)	Bayes Net	Bot-IoT	(99.77%)
	Decision tree		(99.79%)
	Naive Bayes		(99.79%)
	Random forest		(99.99%)
	Random tree		(99.99%)
	DT	Bot-IoT	99.87%
	Naive Bayes	UNSW	85.91%
	K-NN		82.14%
	SVM		99.25%
(Sahu et al., 2021)	CNN-LSTM	IOT-23	96.0
(Anthi et al., 2019)	Naïve Bayes	IOT-23	(79%)
	Bayesian Network		(96%)
	J48 Decision Tree		(98.8%)
	Zero R		(17%)
	OneR		(79.0%)
	Simple Logistic		(96.0%)
	Support V M		(89.0%)
	MLP		(N/A)
	Random Forest		(96.0%)
(Anwer et al., 2021)	SVM	NSL KDD	(32.38)
	GBDT		(32.38)
	RF		(85.34)
(Sadaf & Sultana, 2020)	Autoencoder	NSL KDD	88.98
(Moussa & Alazzawi, 2020)	Stacked Autoencoder	NSL KDD	90.40

7. Discussion

The comparative study provides a thorough empirical examination. Latest scientific investigation By analyzing the dataset and how it affects the classifier's performance, the aim is to apply study findings and identify any bias in the findings. The results indicated a difference in network attack traffic through the variable dataset, which significantly affects the results and causes them to decrease. Anomaly-based Intrusion Detection Systems (IDS) focus primarily on data to make crucial decisions when identifying and responding to attacks within a given system. False-negative outcomes carry significant effects as they pose a risk of targeting the actual environment, however the system cannot identify them. An IoT device can be incorporated into a botnet loop, resulting in Distributed Denial of Service (DDoS) assaults. False positives place an unwarranted load on the system by producing false alarms and dealing with benign traffic that is improperly labeled as harmful. There exist models that have been trained on a particular dataset, which may contain inherent biases that the researchers were not aware of(Hinnefeld et al., 2018). Our study demonstrates that models yield discriminatory outcomes when they are not trained on a dataset that encompasses a wide range of diversity. Anomaly detection A reliable dataset is needed for IDS to produce unbiased training and testing results. Selection of the correct dataset for the system is critical(Al-Hadhrami & Hussain, 2021a).

8. Conclusion

The increasing importance of the Internet of Things requires the development of robust security solutions for efficient, secure communications and reliable operations. This review comprehensively analyzed the different security threats and attack vectors within the complex IoT landscape, where diverse technologies such as hardware, software, protocols, and communications require coordinated security. Studies have evaluated the potential of machine learning (ML) and deep learning (DL) methods for anomaly detection and explored the advantages and limitations of existing approaches to address vulnerabilities and classification attacks. Furthermore, the review addressed the impact of datasets on building robust ML/DL models for IoT security. By identifying key challenges, and limitations and identifying

promising future directions, this review aims to provide a valuable resource that encourages researchers to push the boundaries of IoT security beyond secure communications and advance a comprehensive and sustainable security model.

References

Abdulrahman, A. A., & Ibrahem, M. K. (2021). Intrusion detection system using data stream classification. *Iraqi Journal of Science*, *62*(1), 319–328. https://doi.org/10.24996/ijs.2021.62.1.30

Ahmad, R., Alsmadi, I., Alhamdani, W., & Tawalbeh, L. (2022a). A comprehensive deep learning benchmark for IoT IDS. *Computers and Security*, *114*. https://doi.org/10.1016/j.cose.2021.102588

Ahmad, R., Alsmadi, I., Alhamdani, W., & Tawalbeh, L. (2022b). Towards building data analytics benchmarks for IoT intrusion detection. *Cluster Computing*, *25*(3), 2125–2141. https://doi.org/10.1007/s10586-021-03388-z

Ahmim, A., Maglaras, L., Ferrag, M. A., Derdour, M., & Janicke, H. (2018). *A Novel Hierarchical Intrusion Detection System based on Decision Tree and Rules-based Models*. http://arxiv.org/abs/1812.09059

Al-Garadi, M. A., Mohamed, A., Al-Ali, A. K., Du, X., Ali, I., & Guizani, M. (2020). A Survey of Machine and Deep Learning Methods for Internet of Things (IoT) Security. *IEEE Communications Surveys and Tutorials*, *22*(3), 1646–1685. https://doi.org/10.1109/COMST.2020.2988293

Al-Hadhrami, Y., & Hussain, F. K. (2021a). DDoS attacks in IoT networks: a comprehensive systematic literature review. *World Wide Web*, *24*(3), 971–1001. https://doi.org/10.1007/s11280-020-00855-2

Al-Hadhrami, Y., & Hussain, F. K. (2021b). DDoS attacks in IoT networks: a comprehensive systematic literature review. *World Wide Web*, *24*(3), 971–1001. https://doi.org/10.1007/s11280-020-00855-2

Ali, Z., Mahmood, A., Khatoon, S., Alhakami, W., Ullah, S. S., Iqbal, J., & Hussain, S. (2023). A Generic Internet of Things (IoT) Middleware for Smart City Applications. *Sustainability (Switzerland)*, *15*(1). https://doi.org/10.3390/su15010743

Alissa, K., Alyas, T., Zafar, K., Abbas, Q., Tabassum, N., & Sakib, S. (2022). Botnet Attack Detection in IoT Using Machine Learning. *Computational Intelligence and Neuroscience*, *2022*. https://doi.org/10.1155/2022/4515642

Alotaibi, A. I., & Oracevic, A. (2023). Context-Aware Security in the Internet of Things: What We Know and Where We are Going. *2023 International Symposium on Networks, Computers and Communications (ISNCC)*, 1–8. https://doi.org/10.1109/ISNCC58260.2023.10323735

Alqahtani, H., Sarker, I. H., Kalim, A., Minhaz Hossain, S. M., Ikhlaq, S., & Hossain, S. (2020). Cyber intrusion detection using machine learning classification techniques. *Communications in Computer and Information Science*, *1235 CCIS*, 121–131. https://doi.org/10.1007/978-981-15-6648-6_10

Al-Yaseen, W. L., Othman, Z. A., & Nazri, M. Z. A. (2017). Multi-level hybrid support vector machine and extreme learning machine based on modified K-means for intrusion detection system. *Expert Systems with Applications*, *67*, 296–303. https://doi.org/10.1016/j.eswa.2016.09.041

Anthi, E., Williams, L., Slowinska, M., Theodorakopoulos, G., & Burnap, P. (2019). A Supervised Intrusion Detection System for Smart Home IoT Devices. *IEEE Internet of Things Journal*, *6*(5), 9042–9053. https://doi.org/10.1109/JIOT.2019.2926365

Anwer, M., Umer Farooq, M., & Mahmood Khan, S. (2021). Attack Detection in IoT using Machine Learning. *Engineering, Technology & Applied Science Research*, *11*(3), 7273-7278.

Asharf, J., Moustafa, N., Khurshid, H., Debie, E., Haider, W., & Wahab, A. (2020). A Review of Intrusion Detection Systems Using Machine and Deep Learning in Internet of Things: Challenges, Solutions and Future Directions. *Electronics*, *9*(7), 1177. https://doi.org/10.3390/electronics9071177

Bhuiyan, M. N., Rahman, M. M., Billah, M. M., & Saha, D. (2021). Internet of Things (IoT): A Review of Its Enabling Technologies in Healthcare Applications, Standards Protocols, Security, and Market Opportunities. *IEEE Internet of Things Journal*, *8*(13), 10474–10498. https://doi.org/10.1109/JIOT.2021.3062630

Burhan, H. M., Attea, B. A., Abbood, A. D., Abbas, M. N., & Al-Ani, M. (2021). Evolutionary multi-objective set cover problem for task allocation in the Internet of Things. *Applied Soft Computing*, *102*. https://doi.org/10.1016/j.asoc.2021.107097

Butun, I., Osterberg, P., & Song, H. (2020). Security of the Internet of Things: Vulnerabilities, Attacks, and Countermeasures. *IEEE Communications Surveys and Tutorials*, *22*(1), 616–644. https://doi.org/10.1109/COMST.2019.2953364

Chakrabarti, S., Saha, H. N. (2019). University of Nevada, Institute of Electrical and Electronics Engineers. Region 1, Institute of Electrical and Electronics Engineers. Region 6, IEEE-USA, & Institute of Electrical and Electronics Engineers. (n.d.-a). *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC) : 7th-9th January, 2019, University of Nevada, Las Vegas, NV, USA*.

Chen, P., Liu, H., Xin, R., Carval, T., Zhao, J., Xia, Y., & Zhao, Z. (2022). Effectively Detecting Operational Anomalies In Large-Scale IoT Data Infrastructures By Using A GAN-Based Predictive Model. *Computer Journal*, *65*(11), 2909–2925. https://doi.org/10.1093/comjnl/bxac085

Chen, Y. W., Sheu, J. P., Kuo, Y. C., & Van Cuong, N. (2020, June). Design and implementation of IoT DDoS attacks detection system based on machine learning. In *2020 European Conference on Networks and Communications (EuCNC)* (pp. 122-127). IEEE.

Conti, M., Dehghantanha, A., Franke, K., & Watson, S. (2018). Internet of Things security and forensics: Challenges and opportunities. In *Future Generation Computer Systems* (Vol. 78, pp. 544–546). Elsevier B.V. https://doi.org/10.1016/j.future.2017.07.060

Costa, J., Dessai, N., Gaonkar, S., Aswale, S., & Shetgaonkar, P. (2020). Iot-botnet detection using long short-term memory recurrent neural network. *Int. J. Eng. Res*, *9*(8), 531-536.

Fadhil, M. S., Farhan, A. K., & Fadhil, M. N. (2021). A lightweight aes algorithm implementation for secure iot environment. *Iraqi Journal of Science*, *62*(8), 2759–2770. https://doi.org/10.24996/ijs.2021.62.8.29

Ferrag, M. A., Maglaras, L., Moschoyiannis, S., & Janicke, H. (2020). Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. *Journal of Information Security and Applications*, *50*. https://doi.org/10.1016/j.jisa.2019.102419

Gatea, M. J., & Hameed, S. M. (2022). An Internet of Things Botnet Detection Model Using Regression Analysis and Linear Discrimination Analysis. *Iraqi Journal of Science*, *63*(10), 4534–4546. https://doi.org/10.24996/ijs.2022.63.10.36

Gharkan, D. K., & Abdulrahman, A. A. (2023). Construct an efficient distributed denial of service attack detection system based on data mining techniques. *Indonesian Journal of Electrical Engineering and Computer Science*, *29*(1), 591–597. https://doi.org/10.11591/ijeecs.v29.i1.pp591-597

Hinnefeld, J. H., Cooman, P., Mammo, N., & Deese, R. (2018). *Evaluating Fairness Metrics in the Presence of Dataset Bias*. http://arxiv.org/abs/1809.09245

Hussien, Z. K., & Dhannoon, B. N. (2020a). Anomaly detection approach based on deep neural network and dropout. *Baghdad Science Journal*, *17*(2), 701–709. https://doi.org/10.21123/bsj.2020.17.2(SI).0701

Jalawkhan, M. S., & Mustafa, T. K. (2021). Anomaly Detection in Flight Data Using the Naïve Bayes Classifier. *2021 7th International Conference on Contemporary Information Technology and Mathematics (ICCITM)*, 26–30. https://doi.org/10.1109/ICCITM53167.2021.9677655

Kadhm, M. S., Ayad, H., & Mohammed, M. J. (2021). Palmprint recognition system based on proposed features extraction and (c5. 0) decision tree, k-nearest neighbour (knn) classification approaches. *J. Eng. Sci. Technol*, *16*(1), 816-831.

Kevric, J., Jukic, S., & Subasi, A. (2017). An effective combining classifier approach using tree algorithms for network intrusion detection. *Neural Computing and Applications*, *28*, 1051–1058. https://doi.org/10.1007/s00521-016-2418-1

Khraisat, A., Gondal, I., Vamplew, P., & Kamruzzaman, J. (2019). Survey of intrusion detection systems: techniques, datasets and challenges. *Cybersecurity*, *2*(1). https://doi.org/10.1186/s42400-019-0038-7

Larriva-Novo, X., Vega-Barbas, M., Villagrá, V. A., Rivera, D., Álvarez-Campana, M., & Berrocal, J. (2020). Efficient distributed preprocessing model for machine learning-based anomaly detection over large-scale cybersecurity datasets. *Applied Sciences (Switzerland)*, *10*(10). https://doi.org/10.3390/app10103430

Lee, I., & Lee, K. (2015). The Internet of Things (IoT): Applications, investments, and challenges for enterprises. *Business Horizons*, *58*(4), 431–440. https://doi.org/10.1016/j.bushor.2015.03.008

Liu, W., Zou, P., Jiang, D., Quan, X., & Dai, H. (2022). Zoning of reservoir water temperature field based on K-means clustering algorithm. *Journal of Hydrology: Regional Studies*, *44*. https://doi.org/10.1016/j.ejrh.2022.101239

Lv, Z., & Singh, A. K. (2021). Big Data Analysis of Internet of Things System. *ACM Transactions on Internet Technology*, *21*(2). https://doi.org/10.1145/3389250

Mahesh, B. (2018). Machine Learning Algorithms-A Review. *International Journal of Science and Research (IJSR)*, *9*(1), 381-386. https://doi.org/10.21275/ART20203995

Mishra, N., & Pandya, S. (2021). Internet of Things Applications, Security Challenges, Attacks, Intrusion Detection, and Future Visions: A Systematic Review. *IEEE Access*, *9*, 59353-59377. https://doi.org/10.1109/ACCESS.2021.3073408

Moussa, M. M., & Alazzawi, L. (2020). Cyber Attacks Detection based on Deep Learning for Cloud-Dew Computing in Automotive IoT Applications. *Proceedings - 2020 IEEE International Conference on Smart Cloud, SmartCloud 2020*, 55–61. https://doi.org/10.1109/SmartCloud49737.2020.00019

Musleh, A. S., Chen, G., & Dong, Z. Y. (2020). A Survey on the Detection Algorithms for False Data Injection Attacks in Smart Grids. *IEEE Transactions on Smart Grid*, *11*(3), 2218–2234. https://doi.org/10.1109/TSG.2019.2949998

Neto, E. C. P., Dadkhah, S., Ferreira, R., Zohourian, A., Lu, R., & Ghorbani, A. A. (2023). CICIoT2023: A Real-Time Dataset and Benchmark for Large-Scale Attacks in IoT Environment. *Sensors*, *23*(13). https://doi.org/10.3390/s23135941

Oudah, M. S., & Maolood, A. T. (2023). IoT-Key Agreement Protocol Based on The Lowest Work-Load Versions of The Elliptic Curve Diffie-Hellman. *Iraqi Journal of Science*, *64*(8), 4198–4207. https://doi.org/10.24996/ijs.2023.64.8.39

Roy, B., & Cheung, H. (2018). A Deep Learning Approach for Intrusion Detection in Internet of Things using Bi-Directional Long Short-Term Memory Recurrent Neural Network; A Deep Learning Approach for Intrusion Detection in Internet of Things using Bi-Directional Long Short-Term Memory Recurrent Neural Network. In *2018 28th International Telecommunication Networks and Applications Conference (ITNAC)*. https://www.unsw.adfa.edu.au/australian-centre-for-cyber-

Sadaf, K., & Sultana, J. (2020). Intrusion detection based on autoencoder and isolation forest in fog computing. *IEEE Access*, *8*, 167059–167068. https://doi.org/10.1109/ACCESS.2020.3022855

Sahu, A. K., Sharma, S., Tanveer, M., & Raja, R. (2021). Internet of Things attack detection using hybrid Deep Learning Model. *Computer Communications*, *176*, 146–154. https://doi.org/10.1016/j.comcom.2021.05.024

Shafiq, M., Tian, Z., Sun, Y., Du, X., & Guizani, M. (2020). Selection of effective machine learning algorithm and Bot-IoT attacks traffic identification for internet of things in smart city. *Future Generation Computer Systems*, *107*, 433–442. https://doi.org/10.1016/j.future.2020.02.017

Singh, A., Payal, A., & Bharti, S. (2019). A walkthrough of the emerging IoT paradigm: Visualizing inside functionalities, key features, and open issues. *Journal of Network and Computer Applications, 143*, 111-151. https://doi.org/10.1016/j.jnca.2019.06.013

Sriavstava, R., Singh, P., & Chhabra, H. (2020). Review on cyber security intrusion detection: Using methods of machine learning and data mining. *Internet of Things and Big Data Applications: Recent Advances and Challenges,* 121-132. https://doi.org/10.1007/978-3-030-39119-5_8

Taherdoost, H. (2023). Security and internet of things: benefits, challenges, and future perspectives. *Electronics*, *12*(8), 1901. https://doi.org/10.3390/electronics12081901

Tahsien, S. M., Karimipour, H., & Spachos, P. (2020). Machine learning based solutions for security of Internet of Things (IoT): A survey. *Journal of Network and Computer Applications*, *161*. https://doi.org/10.1016/j.jnca.2020.102630

Tharewal, S., Ashfaque, M. W., Banu, S. S., Uma, P., Hassen, S. M., & Shabaz, M. (2022). Intrusion Detection System for Industrial Internet of Things Based on Deep Reinforcement Learning. *Wireless Communications and Mobile Computing*, *2022*(1), 9023719. https://doi.org/10.1155/2022/9023719

Yang, W., Johnstone, M. N., Sikos, L. F., & Wang, S. (2020). Security and Forensics in the Internet of Things: Research Advances and Challenges. *Proceedings - 2020 Workshop on Emerging Technologies for Security in IoT, ETSecIoT 2020*, 12–17. https://doi.org/10.1109/ETSecIoT50046.2020.00007