2023

A Survey of Using Machine Learning in IoT Security and the Challenges Faced by Researchers

Khawlah M. Harahsheh
Old Dominion University, khara001@odu.edu

Chung-Hao Chen
Old Dominion University, cxchen@odu.edu

Original Publication Citation

Harahsheh, K., & Chen, C.-H. (2023). A survey of using machine learning in IoT security and the challenges faced by researchers. *Informatica*, *47*(6), 1-54. https://doi.org/10.31449/inf.v47i6.4635

A Survey of Using Machine Learning in IoT Security and the Challenges Faced by Researchers

Khawlah Harahsheh[1], Chung-Hao Chen[2]
[1] Ph.D. student, Department of Electrical & Computer Engineering, Old Dominion University, Norfolk, VA, USA
[2] Associate Professor, Department of Electrical & Computer Engineering, Old Dominion University, Norfolk, VA, USA
E-mail: khara001@odu.edu[1], cxchen@odu.edu[2]

Keywords: Internet-of-Things (IoT), machine learning, deep learning, cyber-security, intrusion detection, malware, big data

Received: January 27, 2023

The Internet of Things (IoT) has become more popular in the last 15 years as it has significantly improved and gained control in multiple fields. We are nowadays surrounded by billions of IoT devices that directly integrate with our lives, some of them are at the center of our homes, and others control sensitive data such as military fields, healthcare, and datacenters, among others. This popularity makes factories and companies compete to produce and develop many types of those devices without caring about how secure they are. On the other hand, IoT is considered a good insecure environment for cyber thefts. Machine Learning (ML) and Deep Learning (DL) also gained more importance in the last 15 years; they achieved success in the networking security field too. IoT has some similar security requirements such as traditional networks, but with some differences according to its characteristics, some specific security features, and environmental limitations, some differences are made such as low energy resources, limited computational capability, and small memory. These limitations inspire some researchers to search for the perfect and lightweight security ways which strike a balance between performance and security. This survey provides a comprehensive discussion about using machine learning and deep learning in IoT devices within the last five years. It also lists the challenges faced by each model and algorithm. In addition, this survey shows some of the current solutions and other future directions and suggestions. It also focuses on the research that took the IoT environment limitations into consideration.

Povzetek: Podan je pregled uporabe strojnega in globokega učenja v IoT napravah ter izzivov, rešitev in prihodnjih smeri raziskav.

1 Introduction

Pervasive growth and use of the Internet and mobile applications have expanded cyberspace [1]. The huge distribution of smart sensors and devices around us as an important part related to our lives makes researchers focus on the security and performance of the Internet of Things (henceforth IoT). IoT refers to a type of network that allows any object to be connected to each other using communication protocols [2]. The term IoT was invented by Kevin Ashton in 1999 while he was developing supply chain optimization at Proctor & Gamble, and according to a recent statistical study released in 2019, there were a total of 22 billion IoT devices connected worldwide in 2018. It also projects that the number will be increased to 38.6 billion in 2025 and 50 billion in 2030 [3]. Smart objects are called 'smart' because these objects are intelligent, and they can communicate with each other and with human beings. These objects became powerful as they have embedded chips with small processors, equipped with power sources, sensors, and data transmitters and receivers [4].

The IoT shares some security needs with traditional networks, but also has some unique security measures based on its own characteristics and limitations which make some differences between it and traditional networks. Peoples and individuals daily store huge data in the cloud, which makes it a challenge to secure this data and the back-and-forth connections, especially sensitive and private information. All information should be encrypted before transfers over the connections; on the other side, the authorized users will have the key to decrypt the data when arrived.

IoT technologies have been employed broadly in many sectors, such as telecommunications, transportation, manufacturing, water and power management, healthcare, education, finance, government, and even entertainment [5]. IoT is not an innovation: it is an evolution. IoT is the combination of technologies, including sensors, advanced automation systems, networking, data collection, data analysis, and small processing devices embedded into objects [4]. Most of the IoT and cyber–physical system (CPS) devices are comprised of physical objects, such as smart vehicles, drones, smart appliances, and other machines/machinery, which are embedded with sensors for either a single specific application or multiple applications [6]. The wide variety of IoT devices comes with security and privacy problems [3]. It is not only privacy as people rely more on technology for different activities such as shopping, banking, doing business, and

online studying. The proliferation of IoT was expected to reach 29 billion connected devices by 2022, and the IoT market size was anticipated to reach U.S. $54 billion by 2022 [7]. We believe those numbers increased because of Covid-19 where more people started using online shopping and online studying more than any time before. Some IoT devices are embedded in public areas and use shared networks, and this makes them vulnerable and easy to attack.

The IoT facilitates integration between the physical world and computer communication networks and applications (apps) such as infrastructure management and environmental monitoring make privacy and security techniques critical for future IoT systems [8]. The IoT ecosystem is likely to be confronted with nonconventional security challenges. Besides, the security vulnerabilities that the IoT faces due to the heterogeneity and resource limitations of the IoT devices, the interactions among the IoT, and fog and cloud layers, make room for additional vulnerabilities [9].

There exist several ways IoT nodes connect to the Internet, and this includes communication protocols such as the Transmission Control Protocol and the Internet Protocol (TCP/IP) using Message Queue Telemetry Transport (MQTT), Modbus TCP, Cellular, and Long-Range Radio Wide Area Network (LoRaWAN), among others [10]. Theft of sensitive data or network disruptions, such as Brute Force, Port Scanning, Denial of Service (DoS), Distributed denial of service (DDoS), Man in the middle (MITM), Remote to Local (R2L), Probing (Probe), User to Root (U2R) and operating system attacks are all examples of IoT attacks [11].

The volume of audit data surges rapidly when the network size is enlarged. This makes manual detection difficult or even impossible [5], due to the increasing quantities of data transmitted over the Internet which led to the introduction of new networking paradigms (e.g., the Internet of Things (IoT), cloud computing, and fog/edge computing, and complex inference models (e.g., deep learning (DL)) [12]. The concept of machine learning emerged in the middle of the 20th century; nevertheless, it was not until the 1990s that the application took off [2].

IoT devices are generally limited in computational capability and so are often unable to incorporate or employ the various security mechanisms and protocols used by more powerful systems [13]. Intrusion detection systems (IDSs) first collect and process data, and then apply a detection mechanism to raise alarms which are sent to a human network analyst for further screening [12]. This survey focuses on ML and DL techniques used in the last five years to secure the IoT environment. It also considers the devices' limitations and lists the most important open challenges and future works for hundreds of studies that help other researchers to improve IoT security.

1.1 About the survey goal

A lot of previous surveys tackled security using machine learning and deep learning for regular computers and servers, many of them took IoT security into count because IoT has special characteristics that make securing it different from regular systems. ML and DL techniques have transformed security in IoT systems in recent years. Several researchers have conducted surveys on security methods integrating machine learning on IoT networks to give a practical guide to existing solutions [2]. The aim of this survey is to list the largest possible number of challenges faced by researchers in the field of securing IoT devices using ML and DL. Listing the challenges may help us and other researchers to directly find problems, try to solve them, and draw a roadmap for future work. The survey also helps compare several ML and DL techniques and algorithms to finally get the best performance depending on its accuracy, precision, recall, and other results.

This survey mainly covers everything related to IoT security using machine learning methods and takes into account the environment limitations of such devices. This marks the difference between the current survey and other existing survives, where we cover ML techniques and IoT security, get lightweight in the count, list challenges, and list a group of the current solutions. The main contributions of this work include:

- Comparing existing surveys and the current one as this comparison provides a detailed explanation of ML-based security solutions for IoT environments and domains.
- Highlighting lightweight ML techniques and IOT environment limitations which makes us stand out from other surveys.
- Filling the gap between IoT limitations and characteristics and the strength of DL and ML to cover the security challenges caused by IoT limitations.
- Focusing on IoT security using ML in the last five years from 2018 to 2023. We chose only the recent five years because ML and IoT are changing and developing very quickly like other fields of technology.
- Focusing on detection systems such as malware detection, intrusion detection, and attack detection because IoT security has different steps and methods, and we exclude prevention and response systems.
- Presenting an in-depth review of different research challenges related to the application of ML and DL techniques in IoT that need to be addressed [22], in addition to listing all future directions that may be possible to solve the challenges.

This survey analyzed the most important machine learning techniques used in cyber security and identified the growing trend of applying these approaches to secure IoT environments. In this survey, we have given a brief deep overview of machine learning and deep learning techniques and how they may be used to identify and categorize threats. By presenting literature on ML

techniques for cyber security, including intrusion detection, spam detection, and malware detection on IoT systems in the last five years, our survey provided a comprehensive overview of the challenges that ML and DL techniques used in protecting cyberspace against attacks in IoT environments must overcome. Additionally, it offered concise explanations of each ML technique, frequently used security datasets, necessary ML tools, and evaluation metrics for model evaluation. The difficulties of using AI approaches to IoT security are also covered. This article presents the most recent comprehensive bibliography as well as the most recent developments in ML and DL in IoT security.

1.2 Paper structure

Figure 1 below shows the survey organization and structure. Section II discusses the related work and provides a clear comparison with existing surveys. It also highlights the major differences between our survey and the other ones which tackled IoT security using ML solutions. A quick overview of cybersecurity in general and vulnerabilities is provided in section III, followed by a detailed IoT discussion from history to the present found in Section IV. Moreover, Section V discusses IoT security and clears the most common vulnerabilities and threads in such areas. In Section VI, the top IoT security datasets are listed, focusing on the steps and pre-processing steps used to deal with datasets. Section VII provides the most work of our survey where we deeply discuss ML, DL, and how they were used to secure IoT systems. This section also contains different comparison tables and graphs with a brief description of ML techniques, and how those techniques are used to detect different types of cyberattacks in many IoT domains. We also prepared ten tables that list and compare 152 papers in the same field. Finally, all IoT security challenges were listed which were found in the papers written in the last five years, in addition to some of the current solutions and future directions in Section VIII. Finally, Section IX draws the conclusions from this survey.

Figure 2 views a visual representation of the process used to select papers for this survey. Our entire search was done in IEEE Xplore for the years 2018 to 2023. When we used the terms “machine learning” and “cyber security” as keywords to search for in the five years, we found 1066 different papers about using machine learning in cybersecurity, but when we specified our search for IoT systems, the number of papers reduced to 643. In this survey, we focus on detection methods and ignore anything related to prevention or response, so when we added “detection” to the previous search terms, the number of papers became 321, and after deleting any duplicated ideas, we finally selected 235 papers for this survey, and they are divided as follows:

- Previous surveys= 22
- Malware detection= 16
- Anomaly detection= 16
- Attack detection= 14
- Intrusion detection= 30
- DOS/ DDOS detection= 16
- BOTNET detection= 32
- Lightweight in account= 6
- General papers about IoT security using ML= 58

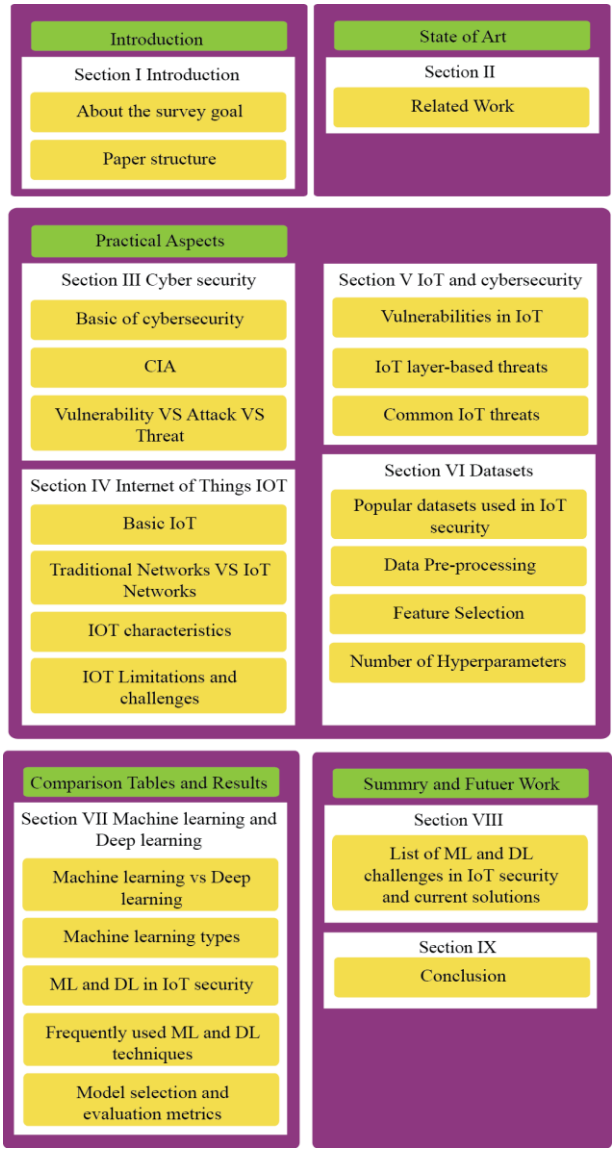


Figure 1: Survey organization and structure.

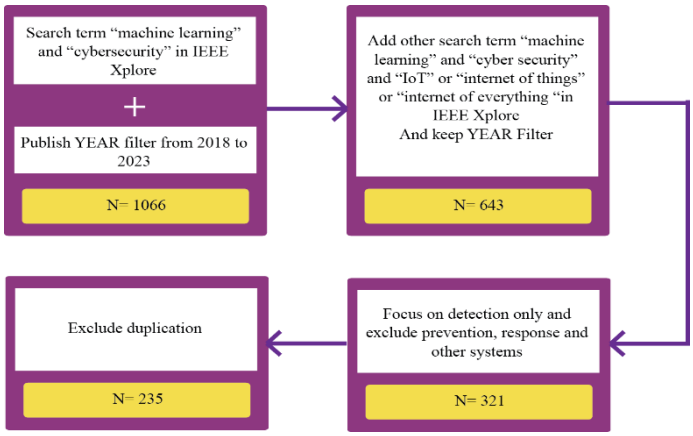


Figure 2: Visual representation views the process for paper selection.

2 Related work

Different surveys about using ML and DL in cybersecurity have been published to provide a deep discussion about using ML in cybersecurity, especially in IoT environments. In this section, a summary of many existing surveys published in the last five years on IEEE Xplore is presented. This summary covers different ML and DL algorithms and techniques applied to IoT systems as a security solution. The main comparison between our survey and others is about the two ideas of this survey, mainly challenges, the IoT limitations, and the solutions such as the works conducted by R. Zhao et al. [14], S. Zaman et al. [23], and F. Hussain et al. [15]. IoT devices usually have different limitations which give them special characteristics and differentiate them from regular devices. Those limitations inspire researchers to search more for lightweight method when they deal with IoT devices, and some of those limitations are low computing capabilities, low power sources, limited storage, among others. Table II

Table 1: List of acronyms

Acronym	Definition	Acronym	Definition	Acronym	Definition
AI	Artificial intelligence	GSOM	Growing self-organizing map	R2L	Remote to Local
AMQP	Advanced Message Queuing Protocol	GUI	Graphical User Interface	RAE	Relative Absolute Error
ANN	Artificial Neural Network	HPCs	Hardware Performance Counters	RAM	Random-Access Memory
API	Application Programming Interface	IDS	Intrusion Detection System	RaNN	Random Neural Network
ASR	Automated speech recognition	IIoT	Industrial Internet of Things	REP Tree	Reduced Error Pruning Tree
AUC	Area Under Curve	IoMT	The Internet of Medical Things	RF	Random Forest
BEC	Business email compromise	IoT	Internet of Things	RFID	Radio Frequency Identification
CIA	Confidentiality, Integrity and Availability	IoT PoT	Phone of Things	RISS	Resilient Information Security System
CNN	Convolutional Neural Network	IP	Internet Protocol	RL	Reinforcement Learning
CoAP	Constrained Application Protocol	IPv6	Internet Protocol version 6	RMSE	Root Mean Square Error
CPS	Cyber–Physical System	ISP	Internet Service Provider	RMSLE	Root Mean Squared Logarithmic Error
CPU	Central Processing Unit	KNN	K- Nearest Neighbor	RNN	Recurrent Neural Network
DBM	Deep Boltzmann Machine	LDA	Latent Dirichlet Allocation	ROC	Receiver Operating Characteristics
DBN	Deep Belief Network	LPWAN	Low Power Wide Area Networks	RRMSE	Relative Root Mean Squared Error
DBScan	Density-Based Spatial Clustering of Applications with Noise	LR	Linear Regression	RRSE	Root Relative Squared Error
DDOS	Distributed Denial-of-Service	LSTM	Long Short-Term Memory	RSE	Relative Squared Error
DDS	Data Distribution Service	M2M	machine-to machine	SARD	Software Assurance Reference Dataset
DGCNN	Deep Graph Convolutional Neural Network	MAE	Mean Absolute Error	SARD	Software Assurance Reference Dataset
DL	Deep Learning	MAPE	Mean Absolute Prediction Error	SDN	Software-Defined Networking
DOS	Dental of Service	MAPE	Mean Absolute Percentage Error	SE	Social Engineering
DT	Decision tree	MBE	Mean Bias Error	SGD	Stochastic Gradient Descent
DTLS	Datagram Transport Layer Security	MCC	Matthews Correlation Coefficient	SMBs	Small and Medium Business Solutions
ETA	Electronic Travel Authorization	MITM	Man-In-The-Middle attack	SNN	Spiking Neural Network
ETC	Extra Trees Classifier	ML	Machine Learning	SSR	Some Squared Regression
FBI	Federal Bureau of Investigation	MLP	Multilayer Perceptrons	SST	Sum Squared Total
FDR	False Discovery Rate	MQTT	Message Queuing Telemetry Transport	SVM	Support Vector Machine
FN	False Negative	MSE	Mean Square Error	SVR	Support Vector Regression
FNR	False Negative Rate	NB	Naive Bayes	TN	True Negative
FOR	False Omission Rate	NFC	Near Field Communication	TNR	True Negative Rate
FP	False Positive	NFV	Network functions virtualization	TP	True Positive
FPR	False Positive Rate	NIDS	Network Intrusion Detection System	TPR	True Positive Rate
FS	Feature Selection	NLP	Natural Language Process	U2R	User to Root
GBDT	Gradient Boosted Decision Trees	NN	Natural Network	VAR	Vector Autoregression
GDA	Gaussian Discriminant Analysis	OneM2M	A global standards initiative for Machine-to-Machine communications	VM	Virtual Machine
GNB	Gaussian Naive Bayes	OS	Operating Systems	VPN	Virtual Private Network
GPU	Graphics Processing Unit	PLA	Perceptron Learning Algorithm	WSN	Wireless Sensor Network

represents the difference between the present survey and the other existing surveys from the last five years.

Moreover, we include some useful surveys that talk about ML security solutions in general, not IoT only. K. Shaukat et al. [1] and other references such as [12], [18], and [30] provided an extensive review of different ML and DL applications and systems for securing regular devices and networks.

Those surveys provide a comprehensive overview of the challenges that ML techniques face in protecting cyberspace against attacks by presenting literature on ML techniques for cybersecurity including intrusion detection, spam detection, and malware detection on computer networks and mobile networks in the last decade [1]. In parallel, other surveys provide a comprehensive review of using ML methods in IoT security with or without taking care of environment limitations. However, most existing security solutions generate a heavy computation and communication load for IoT devices and outdoor IoT devices such as cheap sensors with lightweight security protections are usually more vulnerable to attacks than computer systems [8]. This survey includes full and deep details about ML in IoT security with respect to its characteristics. It also lists all challenges faced by the researchers from 2018 to 2023, in addition to all current solutions and future directions. This helps the researchers to field the best solutions for those challenges, and it draws a road map for future work. Topics in F. Hussain et al. [15] are the most likely to what we cover in our survey, but we enhanced our work when we covered all challenges we found in the last five years, in addition to all possible solutions, where [15] covers a part of them as shown in TABLE II.

Several surveys covering various facets of IoT security have been released. We outline the research on IoT network threats, machine learning algorithms used to counter them, and more specifically, ML-based security solutions in this section. Table 2 provides a summary of all these surveys. K. Shaukat et al. [1] provide a comprehensive overview of the challenges that ML techniques face in protecting cyberspace against attacks by presenting literature on ML techniques for cyber security including intrusion detection, spam detection, and malware detection on the computer and mobile networks in the last decade. Surveys [1] and [19] focus more on regular mobile and wireless networks and devices. On the other hand, the IoT environment is more challenging to secure, but machine learning techniques, especially the lightweight ones, help more in such environmental limitations as mentioned in surveys [5], [14], and the present survey. A comprehensive overview of ML approaches to enable more effective and less detectable attacks is discussed in survey [2].

Moreover, several criteria for the role of AI in wireless networking for CPS and IoT are discussed. For example, they are discussed briefly in the comprehensive survey conducted by B. Salau et al. [6] where authors focus on ML paradigms, such as transfer learning (TL), distributed learning, and federated learning that have evolved as building blocks for the utilization of large data for learning, adaptation, and predictions in CPS and IoT systems that leverage wireless networking. Furthermore, they also highlight challenges faced by current and future wireless networks pertaining to CPS/IoT.

Some open challenges and possible solutions to security problems in IoT environments have been proposed in surveys [1], [2], [14], [15], [16], [17], [19], [21], [22], [23], [25], and [28]. E. Rodríguez et al. [16] discuss the challenges of using DL methods in each cybersecurity threat or attack, and for each contribution, we review the implementation details and the performance of the solution. F. Hussain et al. [15] discuss thoroughly the existing ML and DL solutions for addressing different security problems in IoT networks. We also discuss several future research directions for ML- and DL-based IoT security.

Due to the specific characteristics of each layer of the IoT system, IoT security threats that are related to inherent or newly introduced threats are presented, and various potential IoT system attack surfaces and the possible threats related to each surface were mentioned in survey [17]. Several authors have proposed a classification of possible anomaly attacks. These anomalies can be identified using the techniques of anomaly detection (AD). There are many ways to detect anomalies such as classification, nearest neighbor, clustering, statistical, spectral, information-theoretic, and graph, survey [18] provides an overview of such different Anomaly Detection Techniques (ADT).

Several existing surveys, as indicated in the table above, either exhibit applications in a particular domain or failure to provide the fundamental knowledge that a new researcher needs to enter or comprehend. However, the majority of survey articles primarily cover specific network dangers and assaults. As we searched on IEEE Xplore, we discovered that the majority of studies were conducted on common networks and devices, including servers, laptops, and computers. These devices have solid infrastructures that provide them with the resources they require for ML and DL processing. IoT, on the other hand, consists of unique devices with numerous resource limits, which present a significant security concern. We have concentrated on important aspects of IoT cyber security, including spam classification, malware detection, and intrusion detection on networked computers and mobile devices. We are also one of the few that discuss the use of simple ML and DL approaches to protect the IoT while taking resource constraints into consideration.

The dataset is essential for developing and testing ML models. In Table 7, we have provided a description of frequently used security datasets. Finally, in comparison to other surveys that have been published in the field, our survey is comprehensive and distinctive in that it offers the following elements: popular ML and DL tools, evaluation metrics, a focus on IoT security, a list of recent datasets used in IoT security, and current challenges and recent solutions.

Table 2: Comparison between existing surveys and our survey (Y means yes it covers, N means not cover, and % means partially covers).

Serial #	Reference #	Year	Citations	# Of References	Use ML in security	IoT security	Lightweight into account	Challenges	Solutions
1	[1]	2020	53	668	Y	N	N	Y	N
2	[2]	2022	3	198	Y	Y	N	Y	N
3	[6]	2022	1	125	Y	Y	N	N	N
4	[12]	2021	16	142	Y	N	N	N	N
5	[14]	2022	2	49	Y	Y	Y	N	N
6	[15]	2020	128	229	Y	Y	%	Y	%
7	[16]	2021	3	261	Y	Y	N	Y	N
8	[17]	2020	218	291	Y	Y	N	Y	%
9	[18]	2020	3	94	Y	N	N	N	N
10	[19]	2022	-	189	Y	Y	N	Y	N
11	[20]	2022	-	15	Y	Y	N	N	N
12	[21]	2022	-	192	Y	Y	N	Y	%
13	[22]	2020	41	100	Y	Y	N	Y	%
14	[23]	2021	3	167	Y	Y	%	Y	%
15	[24]	2021	-	42	Y	Y	N	N	N
16	[25]	2019	42	120	Y	Y	N	Y	%
17	[26]	2018	41	18	Y	Y	N	N	N
18	[27]	2021	-	56	Y	Y	N	%	N
19	[28]	2021	14	119	Y	Y	N	Y	N
20	[29]	2020	51	128	Y	Y	N	N	N
21	[30]	2018	344	78	Y	N	N	%	%
22	[31]	2020	6	96	Y	Y	N	N	N
23	Our Survey	2023	-	235	Y	Y	Y	Y	Y

1 Cyber security

3.1 Basic of cybersecurity

Nowadays, everything in our life depends on technology, such as work, online studying, online banking, online shopping, smart homes, and smart cities. These old and new technologies help us make life easier, but on the other hand, it surrounds us with millions of threats and vulnerabilities which present a real danger and raise cybercrimes. A cyberattack is a planned attack between computers that interrupts, incapacitates, destroys, or seizes control of a computer system, and damages or steals the data it houses. Cyberattacks can be carried out in a variety of ways, such as by infecting networks and computers with harmful codes (such as viruses, worms, Trojan horses, and other malware), exploiting spyware to find security holes or steal data, or scrolling down the window in the left of the MS Word Formatting toolbar [227]. The simple meaning of cybersecurity is how to protect devices, networks, and data from electronic and cyberattacks which may be done by hackers, attackers, spammers, and cyber theft. A cybercrime refers to all the unauthorized activities in systems, devices, networks, and data that cause harm to others. According to the Federal Bureau of Investigation (FBI), there are tens of online cybercrimes, and here are the most common five cybercrimes in the United States [208]:

- Business email compromise (BEC) scams exploit the fact that so many of us rely on email to conduct business—both personal and professional—and it is one of the most financially damaging online crimes.
- Identity theft happens when someone steals your personal information, e.g., your Social Security number, and uses it to commit theft or fraud.
- Ransomware is a type of malicious software or malware that prevents you from accessing your computer files, systems, or networks and demands you pay a ransom for their return.
- Spoofing and phishing are schemes aimed at tricking you into providing the sensitive information to scammers.
- Online predators are a growing threat to young people.
- According to the Kaspersky Lab report, cybercrime will cost the business more when it takes longer time to notice; Small and Medium Business Solutions SMBs estimate a cost to their business of $28k, rising to $105k if undetected for more than a week. For enterprises, where a detection system is in place, the estimated financial damage is still $393k, increasing to over $1m if it remains undetected for over seven days [32].

Scams usually increase during natural disasters, and the Covid-19 pandemic had the same effect. Scam emails were sent to millions about government pay-outs and relief efforts, surveys about the virus, fake donations websites, and more, where all of them were full of malicious codes and links. Hackers take the advantage of hot news to make new crimes, such as the earthquakes in Japan and Ecuador in 2016, bush fires in Australia in 2020, and Michael Jackson's tragic death in 2009. Spam emails claiming to know the specifics of incidents were circulated online within a mere eight hours after his demise [12]. Table III below presents a summary of the worst cyberattacks in history that tend to the loss of billions of dollars, damages, and destruction of a huge number of computers. These cyberattacks are not limited to Melissa, ILOVEYOU, MyDoom, Zeus, Stuxnet, CryptoLocker, and Wannacry, but they are still counting [33].

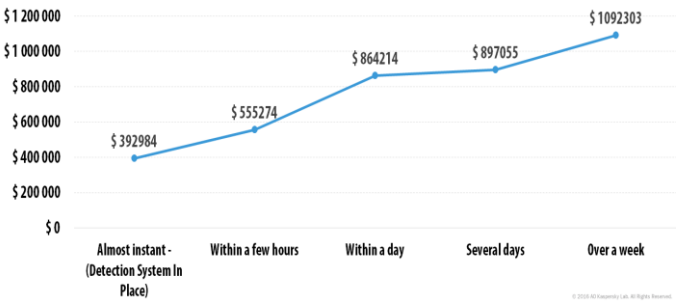


Figure 3: Cost of recovery vs. time needed to discover a security breach for enterprises (source: Kaspersky [38]).

Table 3: Summary of the worst cyberattacks in history [50].

Year	Attack	Infections	Loss
1999	Melissa virus	~ 100,000 computers	~ $1.1 billion
2000	ILOVEYOU worm	~ 45 million computers	~ $10 billion
2004	MyDoom worm	~ 1 million computers	~ $38 billion
2007	Zeus trojan	~ 2,500 companies	~ $100 million
2010	Stuxnet worm	~ 1,000 centrifuges ~ 60,000 computers	~ $50-60 billion
2013	CryptoLocker ransomware	~ 250,000 Windows systems	~ $3 million
2017	Wannacry ransomware	~ 200,000 computers	~ $4 billion

According to the Cisco annual report for 2022 [34], security ranked first in product revenue (see Table IV), and the End-to-End Security product category increased by 9% or $317 million in 2022, compared to 2021, and it also increased by 7% or $224 million in 2021, compared with 2020. This was primarily driven by the growth in the Zero Trust portfolio, Network Security, Unified Threat Management, and Security Endpoint offerings. The Covid-19 pandemic started in 2020 when stores, education, banking, and different areas continued their work online. This brings a huge of new cyber threats in various ways as it changed digital life and affected both companies and individuals.

Back to the Cisco annual report, the numbers show how companies spends many millions in the security section in the last year. The Secure, Agile Networks product category represents Cisco's core networking offerings related to switching, enterprise routing, wireless, and computing. Secure, Agile Networks revenue increased by 5% or $1.1 billion, with growth across the portfolio except enterprise routing in 2022, compared with 2021.

Table 4: Presents product revenue by category (in millions, except percentages) (source: cisco [34]).

	Years Ended			2022 vs. 2021		2021 vs. 202	
	July 30, 2022	July 31, 2021	July 25, 2020	Variance in Dollars	Variance in Percent	Variance in Dollars	Va in P
Product revenue:							
Secure, Agile Networks	$ 23,829	$ 22,722	$ 23,265	$ 1,107	5%	$ (543)	
Internet for the Future	5,278	4,514	4,180	764	17%	334	
Collaboration	4,472	4,727	4,823	(255)	(5)%	(96)	
End-to-End Security	3,699	3,382	3,158	317	9%	224	
Optimized Application Experiences	729	654	524	75	11%	130	
Other Products	11	15	28	(4)	(29)%	(13)	
Total	$ 38,018	$ 36,014	$ 35,978	$ 2,004	6%	$ 36	

Amounts may not sum and percentages may not recalculate due to rounding.

Social engineering (SE) is one of the most common security threats that emerged during Covid-19. It is the act of tricking someone to gain private information, access, or valuables through human interactions without breaking the law. SE did not need to have good experience in coding or technology, but it is just about how to gain others' trust by speaking elegantly, wearing elegant clothes, and being confident. According to a CyberEdge report, "the number of organizations hit with at least one successful social engineering attack per year is around 79%." Similarly, 99% of cyber threats were observed and executed through human interactions and done with the assistance of social engineering approach [6].

An IoT botnet can be utilized to launch DDoS attacks, send spam, mine cryptocurrency, and exploit other weakly configured devices [7]. IoT systems need to protect data privacy and address security issues such as spoofing attacks, intrusions, DoS attacks, distributed DoS (DDoS) attacks, jamming, eavesdropping, and malware [8]. Therefore, there is a pressing need to come up with effective intrusion detection and prevention solutions that help protect the IoT ecosystem from these increasing attacks and threats [9].

3.2 CIA

CIA is a fundamental security model designed to protect data and develop information security. It consists of three parts: Confidentiality, Integrity, and Availability. CIA combines and integrates three means for interacting with data security.

Figure 4: CIA components.

The first one, i.e., Confidentiality, refers to the authorized persons who can access the data within the system. The data will be private unless the person is authorized to see it, and this is important not only for military or government sensitive data but also in every other system since users' privacy is crucial in both sensitive and insensitive data. The second one, i.e., Integrity, is the way to keep data clean. This means while uploading, downloading, and storing data, authorized users only can modify the data. Accidentally, altered data cost companies too much, not only in time and money but also in lacking customers' confidence. The last one is Availability which means that data must be available and accessible whenever and wherever the user needs it.

Sometimes, people are confused between Confidentiality and Availability. While Confidentiality is to make sure only authorized users can access data, Availability refers to making sure authorized users can access the data anytime and from everywhere. This includes checking the availability of the network and hardware which host the data, in addition to checking the applications and security protocols that are running correctly. The combination of these three words creates the main guidelines for protecting and securing information.

Table 5: IOT areas

S#	Area / Environment	Subdomain	Reference In
1	Android		[96], [102], [108]
2	Cloud/ Cluster/ Big data		[100], [101], [117], [123], [134], [143], [164], [171], [197]
3	Industry		[130], [52], [148], [171], [206], [227]
4	Autonomous Vehicle		[103]
5	Smart Cities		[125], [169], [172]
6	Healthcare		[176]
7	IoT devices in general	Hardware	[94], [95], [112], [114], [163], [168], [179], [180]
		Software	[97], [98], [113], [122], [141], [152], [155], [157], [160], [181], [193]
8	IoT Networks in general	Network traffic	[105], [107], [109], [110], [111], [115], [116], [118], [119], [120], [126], [127], [129], [132], [136], [137], [139], [140], [142], [145], [146], [147], [149], [151], [153], [154], [158], [159], [162], [165], [166], [175], [178], [182], [183], [185], [186], [187], [190], [192], [194], [195], [196], [199], [200], [201], [202], [204], [205]
		Protocols	[106], [131], [135], [174]
		ISP	[104]
		Wireless /sensors network	[121], [128], [133], [138], [144], [156], [188], [189]
9	Other	NLP/ Images/ Surveillance systems	[33], [99], [124], [150], [161], [167], [173], [177], [184], [191], [198], [203], [207], [228]

The right balance between the three CIA components should be demonstrated, and this depends on the business's needs to work properly and the level of data sensitivity. For example, imagine that we are talking about a data center that has many servers that host military information. Admins and managers are the only ones who can access this server (Confidentiality). Sometimes, managers need to access the data on the weekend from their homes using the Virtual private network VPN (Availability). The admins added a policy to each user, where some of them can see data – read-only – and others can modify it (Integrity).

3.3 Vulnerability VS Attack VS Threat

A threat is any potential occurrence, whether it is harmful or not, that may badly affect the system, network, application, data, and devices, among others. Threats can be either intentional or unintentional. They could be also natural threats, which are called natural hazards. This includes fires, floods, and earthquakes. Vulnerability is a gap or weakness in any part of the system that makes a threat possible to badly affect, for example, open ports and off firewalls. The Vulnerability occurs due to different reasons such as unprotected designs, users' mistakes, employees' misunderstanding, mistakes in configuration, insecure coding, and any other reasons that make the organization open to cyberattacks. Vulnerability can be either intentional or unintentional. Finally, an attack is when a bad guise that uses vulnerability enacts a threat, so it refers to the cybercrime itself and the action to do this crime, and it is always intentional.

4 Internet of Things IOT

In simple words, the Internet of things (IoT) is related to any physical devices connected wirelessly to each other via the Internet or any other communication networks. Different IoT applications, such as smart grids, healthcare, transportation system, city, supply chain, farming, retail, wearable, environment, manufacturing, home, security, and emergencies are generally referred to as the IoT system. The IoT ecosystem aims at referring to all IoT applications as mentioned [35].

IoT devices are controlled remotely using a center device which usually communicates useful data from the surrounding environment. It is similar to an umbrella of devices "things" that exchange data using the Internet. IoT devices became more popular in the last 15 years and are increasingly used in our lives and other areas such as industry, healthcare, military, smart homes, and smart cities. According to Cisco, by 2030, the number of connected devices is expected to exceed 500 billion [2]. Table V has a list of IoT areas found in the papers studied in this survey.

4.1 Basic IoT

IoT devices are most likely similar in the way of processing and have the same main standard of work, but they vary in their functionality. The main idea of IoT devices is to sense and record data from the physical world and shares it with connected devices through a communication network. In terms of hardware, they are similar too as they have a build-in CPU and network adapter to connect to a Dynamic Host Configuration Protocol (DHCP) which supports the devices with IP addresses to work. The whole devices and data are managed through a web service or software application in a central device.

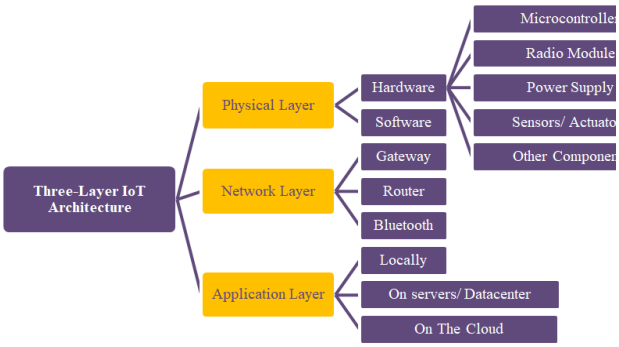


Figure 5: IoT Architecture.

IoT devices store and analyze data which sometimes can be done locally, but most of those devices use the cloud or data centers for this purpose. IoT devices can be monitored and controlled, and they can also communicate, exchange data, and interact over the internet. IoT devices now include many devices such as mobile phones, home security, smart television, and vehicles. They are often called connected or smart devices as they can communicate through a process known as machine-to-machine (M2M) communication [36].

Machine learning became popular in processing IoT data which may be integrated into the same IoT device or in the cloud, and this also depends on the data amount which increases day by day. Below is a list of IoT smart environments:

- Smart Object: They are also called intelligent objects, and they are devices that have the ability to communicate with other devices through a network. Smart objects can collect, store, and process data with other devices to do a specific automated decision, and they are also known as IoT devices.
- Smart environment: It refers to the situation where there are many IoT devices and applications that control or do the main jobs in the environment.
- Smart home: The capacity to regulate domestic appliances using the electronic control. Internet-connected devices are known as smart home automation. Complex heating and lighting systems, alarms, and home security controls may all be programmed in advance, connected to a central hub, and operated remotely by a smartphone app.
- Smart car/ Autonomous vehicle: Those vehicles use a completely automated driving system. This type of vehicle requires little human input to move safely and sense its environment. Self-driving cars use a range of sensors to gather information about their environment. Different sensors are essential to self-driving cars. IoT is essential to the operation of self-driving cars as IoT enables all kinds of devices to be connected to the Internet for information sharing and value-added functions.
- Smart transportation: By giving a precise Electronic Travel Authorization (ETA) for trains and buses as well as by leveraging traffic data to optimize bus transit routes, IoT technology can enhance public transportation. IoT also makes traffic management more effective. Smart traffic lights and sensors powered by IoT can detect heavy traffic volumes automatically and change the duration of the traffic lights as necessary to relieve congestion. Moreover, IoT sensors can be used by cities to monitor which parking spaces are unoccupied around the city. The tracking of busy and unoccupied locations using this data will allow parks to be optimized for maximum effectiveness.
- Smart buildings: By integrating building operations with the IoT, functions like managing the temperature of a building, security, fire detection, water monitoring, maintenance, and more can be made simpler and smarter using computers and mobile devices.
- Smart healthcare: A health service system known as "smart healthcare" uses technology such as wearables IoT devices which use mobile internet to dynamically access information and connect individuals, resources, and institutions involved in healthcare, and then actively manage and intelligently respond to the needs of the medical ecosystem.
- Smart cities: To provide linked solutions for the public, smart cities combine the IoT with a range of software, user interfaces, and communication networks. Smart city sensors for sound and air quality monitoring, water and waste management, and parking management are made possible by IoT.
- Smart metering and smart grids: The most common application of smart metering is smart grids, where the electricity consumption is measured and monitored. Smart metering may also be used to address the problem of electricity theft [37].
- Smart security and emergencies: IoT technologies enable organizations and individual consumers to remotely control and monitor their home security. If the doors have smart locks, these systems can control the monitoring inside and outside the house as well as who has access to them.
- Smart retail: The usage of IoT in the retail sector is closely linked to GPS and RFID technology, which assists firms in tracking products along the supply chain. It provides retailers with the visibility they need to keep an eye on goods movement, conditions, and location while also being able to pinpoint a delivery date.
- Smart agriculture and animal farming: Monitoring soil moisture and condition can also be done with IoT technologies. It is possible to monitor the growth of plants plot by plot using technology like drones. With all of this knowledge, decisions regarding irrigation or fertilization are made fully informed. Grain and vegetable production can be avoided by controlling the temperature and humidity levels during different processes. Increasing the yield and quality of vegetables and crops can also benefit from climate control. Similar to crop monitoring, there are IoT applications that use sensors attached to farm animals to track their movements or stolen from farms and general health.

Just like having many different areas to use IoT, there are many connectivity options, and this produces many attacks and different new vulnerabilities in this environment. Let us first consider the communicative ways which have many protocols such as MQTT, CoAP,

DDS, AMQP, and DTLS, and other wireless protocols such as Zigbee, LPWAN, Bluetooth Low Energy, IPv6, NFC, RFID, and Z-Wave. Some other networks used in different scenarios include regular Wi-Fi networks, satellite, Cellular, and Ethernet. Which one to choose? The answer depends on the scenario and the function of this environment. Each of the previous networks and protocols has its characteristics which were affected by range, bandwidth, power consumptions, and the object of the project.

Choosing the best communication way or connectivity protocol depends on the IoT application, and below are some popular ones:

• Consumer IoT applications: This type is commonly used in smart homes or smart personal devices.
• Commercial IoT applications: They are mainly used in businesses such as smart organizations, smart markets, and healthcare.
• Military Things (IoMT) applications: They support and advance technologies in the military such as drones, robots, and every monitoring system.
• Industrial Internet of Things (IIoT) applications: They are used in the manufacturing and industrial fields.
• Infrastructure IoT applications: They are primarily used in smart cities.

4.2 Traditional networks VS IoT networks

The devices used in traditional networks mostly are more complicated than the ones used in IoT networks. IoT systems depend mainly on small sensor nodes which collect data and store it somewhere in the cloud. IoT nodes have some limitations which will be discussed below such as low power, small memory, less computational process, and few capacities. On the other hand, the devices in traditional networks are connected to a fixed infrastructure that supports them with stable power, alternate power supply, servers, and storage devices, among others.

The second difference concerns security as IoT devices use less secure wireless protocols, such as ZigBee, 802.15.4e, SigFox, LoRa, and 802.11x, which are used by IoT devices to connect with the gateway or the Internet, which result in data leakage and privacy issues [35], while the traditional network administrators use firewalls and other strong encryption protocols for both wired and wireless networks. In addition, the operation systems OS used in network devices are more stable and secure than the OS used in IoT devices.

4.3 IOT characteristics

IoT devices have different characteristics which give them a kind of distinction and difference from the usual systems and regular devices. IoT has some security requirements similar to traditional networks, but because of its characteristics, it also has some unique security features. IoT has speeded devices connected remotely using the Internet where organizations and individuals daily collect records through them and store huge amounts of data in the cloud. This makes it a challenge to secure sensitive and private information through back-and-forth connections. IoT devices, in general, lack strong security measures to protect themselves from security attacks. Furthermore, security patches might not be updated regularly due to the irregular software release or the lack of awareness and expertise of the IoT device users [38]. The ways in which those devices work together give the network and IoT systems special characteristics as follows:

- Scalability: Every day, more and more objects are being connected to IoT. It is speculated that billions of devices are connected with each other and through the Internet will likely surpass the capabilities of the current Internet [15]. Scalability should be considered while designing your system because of the Internet of Things rising ubiquity. The ability of a system to expand without impacting its performance is frequently used to define scalability. This can be done by enhancing an existing system with greater hardware resources or by introducing new software layers. In other words, the system can accommodate more users and data without suffering from performance degradation. IoT is a new technology that is completely changing the way we live and work. IoT can be used in numerous ways. Scaling it to fit your business is one approach.
- Connectivity: IoT has made possible the interconnectivity of Physical and Virtual things with the help of the Internet and global communication infrastructure that is built using wired and wireless technologies [39]. Connectivity is the capacity of two or more devices to exchange information and communicate with one another. In other words, it enables the communication between devices. Businesses have a ton of opportunities to develop new goods and services as a result of this connectedness. Everything is now connected to IoT which creates endless opportunities for the future.
- Safety and security: They are the main issues everyone searching for, especially when personal information is gathered and shared without permission. Data privacy is a major problem with IoT devices. IoT expanding exponentially and more and more products and appliances are being connected to it, and this increases the number of cyber-attacks on these appliances and devices. Some IoT devices are embedded in public areas and use shared networks, and this makes them vulnerable and easy to attack.
- Self-Adapting: Because of the IoT's fast growth over the past several years and because it is an essential aspect of many systems, self-adapting, and dynamic techniques are required for IoT designs to better handle these changes. A self-adaptive system adapts its behavior while it is used in response to changes in the system or its surroundings. IoT devices and systems may be able to dynamically adapt to changing contexts and take appropriate action.
- Self-organization/ self-healing: Self-healing methods allow a system to operate on its own and resolve problems. These are necessary for urgent and

modern IoT communication, including emergency or disaster scenarios. Because relying on the network infrastructure in these critical systems is not an option, self-organizing networks should be implemented.

- Intelligence: The intelligence that goes into the IoT determines how useful it is. The ability of IoT devices to sense data, communicates with one another, and gathers enormous amounts of data for analysis constitutes their intelligence. To link IoT devices to networks and process the data from millions of data nodes, sophisticated software, algorithms, and protocols are employed. They should be always updated with the newest software and firmware if you want your IOT to be intelligent.
- Sensing: The main idea of IoT systems is sensing without sensors. IoT would not be able to detect or measure environmental changes to produce data that may be used to report on their condition, decision making, or even interact with the environment.
- Heterogeneity: One of the major aspects of IoT is heterogeneity. IoT devices can communicate with other devices or service platforms across various networks and are built on different hardware platforms and networks. Direct network connectivity between heterogeneous networks should be supported by IoT networks.
- Communication: When there are so many IoT devices in our lives, it is critical to be able to connect with them in order to ensure effective operation. With these devices, you can connect in a few different ways. The first way is called cloud service, which is a kind of software that enables the connection of the device to the Internet. Another method involves using a gateway that is linked to other devices and enables the communication between them. One IoT device can connect separately to the Internet even though the second device is not connected to a network.
- Low power/ Low cost: IoT devices need very low power and low-cost solutions to work properly. ML algorithms focus most of the time on the effectiveness of the attack which has been configured in a specific way. However, it could be interesting to evaluate other parameters such as energy consumption from an attacker's point of view or the optimal distance from its victim to carry out an attack [2].
- Data: IoT is made up of linked devices that monitor, sense, gather, record, and exchange data. IoT devices and system performance and efficiency can be enhanced with the help of the data they collect. The data is useless without analyzing the collected data using software and tools that transform the records into useful reports or help in decision-making.
- Architecture: Many manufacturers and companies in the IoT industry are utilizing the architecture to power their devices. The architecture is primarily in charge of ensuring that the devices cooperate and interact with one another. It also plays a crucial role in preventing cross-interference between the devices.

4.4 IOT limitations and challenges

IoT devices have power, processor, and memory restrictions such as low-computing capabilities, low power sources, limited storage, or limited memory capacity. Those limitations cause the IoT devices to not be always handled with the sophisticated security protocols they require; they are more likely to be attacked or experience flaws. Because of this, the hardware's design must be expandable in order to provide higher security.

IoT limitation and challenges have different types such as hardware challenges or limitations, software, network, and security. All those limitations and challenges should be taken in account before developing any systems and security solutions. The most critical weakness of IoT would most likely be security. Applications for the IoT may encounter many of the security restrictions and difficulties which recently become the most popular research in this field. These limitations inspire the researchers to implement special steps in securing those devices, such as lightweight cryptography algorithms for encryption, lightweight feature selection algorithms, and lightweight machine learning security framework as in [40], where the authors produced lightweight machine learning based security framework for the detection of malicious phishing URLs. IoT offers many advantages and benefits in various areas and addresses a variety of problems in different sectors. Below is a list of some typical issues and challenges in IoT solutions:

- Data storage: In IoT systems, there are many difficulties for the developers of IoT applications because of the huge recorded and stored data by the heterogeneous IoT sensors. The number of IoT devices is exponentially increasing and its recorded data is increasing as well. As reported, there will be more than 50 billion terminal devices worldwide, and the annual data generated will reach 847 Zebytes by 2021. "Big data" hereby becomes common in IoT applications, such as industrial manufacturing, smart cities, energy Internet, and wireless sensor network (WSN), among others [41]. Huge storage areas are made available by cloud computing, which also provides a platform for IoT-connected devices to communicate. Multiple sensors through IoT store data in the cloud and communicate using cloud systems, instead of using local servers or storage.
- Data format: The variation in IoT systems causes different data formats. Thousands of IoT systems are currently running all around the world, and billions of those devices are used in those systems. Most IoT environments are for sensing purposes, and sensor data is often presented as a tiny tuple of structured data such as Boolean, numeric, continuous data, binary, and more. The different ways that IoT data can be represented make a challenge for developers and data analyzers.
- Architecture challenge: The IoT include numerous linked devices, where each one communicates using a unique set of protocols and standards that

differentiate them from other devices' protocols, so in IoT, there are no clear criteria or guidelines for linking devices together. This causes difficulty for developers in the system architecture.

- Integration with other devices: IoT systems must collaborate with other forms of technical infrastructure in order to generate useful outputs if they are to be even more advantageous. For example, the cloud must be used to store recorded information. Then the data is transferred and distributed using hubs and routers which may also use block-chains to add more security. Additionally, the data collected must be analyzed using big data analytics and other techniques. The problem becomes more complicated when all these technologies are combined with IoT. So, a solution is needed to help in gathering, transferring, and analyzing data quickly and securely without hurting the performance.
- Security, privacy, and trust: Attackers become more interested in using the data from the huge unorganized IoT devices which are scattered around us. For security professionals, integrating security features is difficult since sensors have limited computing and storage capacity. Peoples and individuals store huge data in the cloud daily, which makes it a challenge to secure this data as the back-and-forth connections include sensitive and private information. By creating gaps, attackers might take control of the sensors and break the system.
- Need skills and experience: It is important to consider certain capabilities and experience while designing, deploying, improving, and maintaining security. Any of these components that are disrupted could harm the security of the IoT environment. Because of the daily development of new IoT devices, there is a lake of experts in this field. There are extremely few skilled individuals that can manage IoT technology effectively.
- legal and ethical issues: IoT is connected everywhere around us and inside our homes, and it monitors and records some sensitive data about us. It does not only maintain privacy but also people rely more on technology for different activities such as shopping, banking, doing business, and online studying, especially after the quarantine imposed during Covid-19. Currently, many companies are requesting patients and people to upload their symptoms and vitals to online portals for further prediction and analysis of Covid-19 outbreaks [42]. Users cannot feel confident enough to give personal health information because there are no clear legal operations or instructions offered by the government to protect people's privacy in IoT until now. IoT and other technical systems should be managed and guided by the government to cover the whole process of using such technologies to share information with respecting their legal rights.
- Technical complexity: IoT devices may appear to be doing straightforward activities, but their development requires a great deal of complicated technology. In addition, by giving crucial data to another system, they risk having a harmful impact on every system to which it is linked, and fixing the issue is really not simple. Behind these devices, a wide range of complex processes operates to execute the task. The amount of code and machine connection between the many devices makes it challenging. The main part of the IoT complexity comes from the failure that affects the whole network if one device defects.
- Connectivity dependency to the power and Internet: Internet service is important for IoT devices, where the devices are unable to function, and tasks cannot be completed without a stable internet connection. The proper operation of IoT also requires constant electricity. Both the equipment and everything attached to it stop working when either one fails. IoT devices are so embedded in today's organizations and when it goes down, anything else could come down.
- Higher costs (time and money): IoT device deployment requires a significant time and budget. There are numerous devices that must be ordered and set up, as well as expert members to install the devices and other members to link them to the network and support their teams, and this entire step is costly.
- Forensics challenges: Digital forensics, a branch of forensic science, focuses on recovering and investigating digital materials, such as document and image files, often in relation to computer crimes. IoT forensics can be defined as a branch of digital forensics [43]. IoT forensics is more complex than regular network forensics because of the huge number of devices, the big data collected by those devices, and the diversity of protocols. A variety of devices is another challenge for IoT forensics where it is not only for computers but also for sensors, phones, and all smart devices in the IoT environment. This causes difficulty in identifying the huge number of devices with a variety of hardware and the nature of each. Not only devices but also data privacy added another challenge to IoT forensics.
- Other challenges based on the network: IoT requires a variety of protocols in order to connect to other networks, some of which use IP and others do not. This results in a variety of features and ineffective security measures for devices. Another issue with IoT devices in a single network is their variety, as it is challenging to locate a single network that can support all of the different IoT types. Additionally, IoT devices have the ability to join or leave the network from anywhere at any moment, which results in a dynamic network topology. As a result, the IoT smart devices and their security are not compatible with this architecture, and the

network's current security cannot handle this kind of unexpected topological changes. One of the most important characteristics of IoT devices is mobility, which refers to the ability of these devices to join other networks without the need for pre-configuration.

- Other challenges based on software: Operating systems of IoT that are embedded in IoT devices have thin network protocols, and some vulnerabilities of the IoT devices cannot reprogram because it depends on the embedded protocol.
- Other challenges based on resources: Memory capacity, power capacity, and processing capacity are the main resource limitations in IoT which make it more challenging to secure the comparison with regular network devices. IoT uses a small RAM that stores a few kilobytes of sensing data, and when this RAM is full, some data is dropped and ignored, which may be important sometimes. Most IoT devices use low-bandwidth network connections because they do not consume much power. Many IoT devices have limited power sources that need to be replaced continuously; at the same time, we have some non-rechargeable IoT devices with a large power capacity.

5 IOT and cybersecurity

Nowadays, IoT devices can be found in every corner of our lives. They store and transmit sensitive and costly data, starting from web cameras in our houses which may destroy our privacy with unsecured devices. They also became a main part of the military field such as drones, robots, and datacenters, among others. IoT can be viewed as "A global infrastructure for the information society, enabling advanced services by interconnecting (physical and virtual) things based on existing and evolving interoperable information and communication technologies [44]". IoT also covers large cities, helps autonomous cars to take quick and hard decisions on the road, and it became a main part of everything surrounding us. Thanks to IoT for helping make life easier, and this popularity makes vendors fight to produce and improve more devices. IoT can bring between $3.9 trillion to $11.1 trillion in income by 2025 [2]. At the same time, cybercrime damages are also on the rise [45]. On the other hand, the device itself and its communication network became a cybersecurity challenge due to the large number of devices, the variety of vendors, the characteristics of each layer, and communication options. Some of the most extensive and destructive cyber-attacks deployed on the Internet have been Distributed Denial of Service (DDoS) attacks [46].

The IoT system layers divided into three main parts: physical layer, network layer, and application layer. Different layers characteristics present very critical vulnerabilities. These vulnerabilities are primarily related to poor physical security, resource constraints, insufficient authentication and encryption, insecure access controls, and inadequate update management as mentioned in survey [2].

5.1 Vulnerabilities in IoT

Because of the big range of IoT covers in different industries, academic environments, vendors, manufacturers, and competitors to produce new devices daily, all compete to be the top developer in the IoT world, little of them take into account to improve devices' security, which makes IoT devices a target to many attackers. Threats are not only in the device itself, but they are in all parts of the IoT system. The steps are as follows:

1) Hardware threats in the same device.
2) Network threats.
3) Clouds threats.
4) Web and application threats.
5) Other threats.

Each IoT layer has different kinds of attacks which could be active or passive. "An active attack disturbs the operation of running services, whereas a passive attack enumerates IoT network information without disturbing the live service" [12] page 1649. The most common attack in all IoT systems, services, and layers is the denial of service (DOS) which will be discussed in detail in the next sections.

5.2 IoT layer-based threats

This section discusses in-depth IoT threats and vulnerabilities for the different layers as in Figure (7) and Table (11) in the appendix which compares 152 papers in the last five years and covers different kinds of detections in the IoT environment.

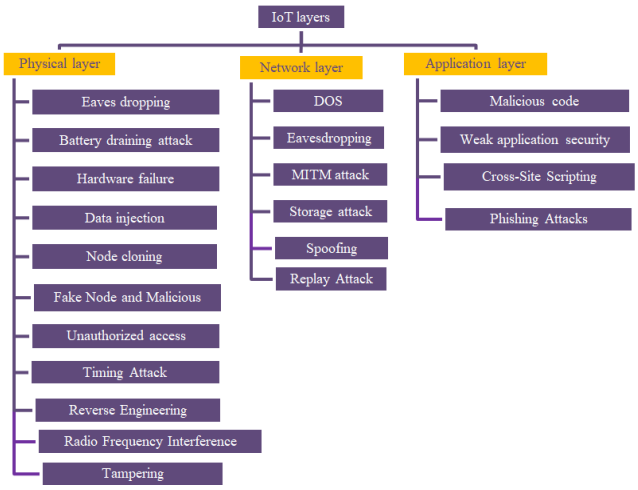


Figure 6: IoT layer-based threats.

1) Physical layer, also known as perception layer: It contains the hardware especially the sensors which are the main part of the IoT environment. The main idea of the physical layer is sensing and gathering data. Machine learning, data encryption, and secure authentication are three ways to secure the physical layer [3]. The security

of this layer is very important because it can be attacked physically and from cyberspace. Below is a list of the most popular attacks in the IoT physical layer.

- Eavesdropping: This happened by connecting a device in the traffic path to get useful information by making passive snigging.
- Battery draining attack: a huge number of requests done by malicious devices controlled by attackers. These multiple requests make power loss in IoT devices, as discussed above in section IV under "IOT Limitations and challenges". Such a kind of those devices have multiple limitations, and power is one of them.
- Hardware failure: when the device itself is physically damaged.
- Data injection: it is the process where the hacker changes the meaning of the original data before sending it to the application; it may happen by changing one binary digit.

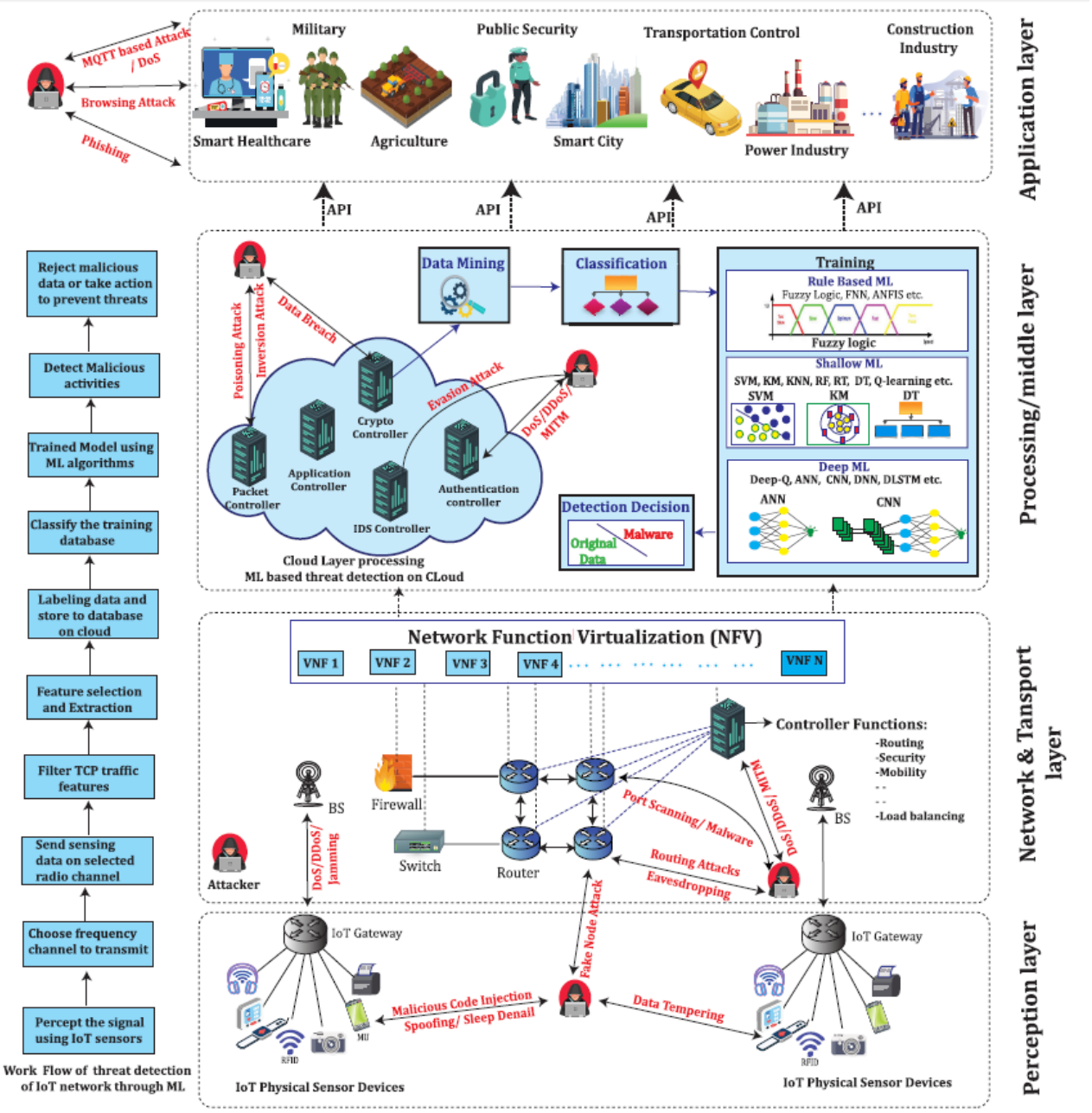


Figure 7: IoT architecture. Data captured by sensors in the perception layer can be sent Network and Transport layer for reliable communication. The processing layer is responsible to secure big data in a cloud server where AI-based security mechanisms are implemented to provide security services to the Application layer users against frequent threats on IoT networks [13].

- Node cloning: Cloned IoT devices are dangerous behavior, where an attacker can physically steal the devices, take critical information, make an exact copy of the devices, and deploy the duplicates in the systems to harm them.
- Fake Node and Malicious: A malicious node is one that tries to prevent other IoT nodes in the network from receiving services. It also makes changes to data before, after, and while transmission.
- Reverse Engineering: occurs when the attacker disassembles the IoT system into small pieces to identify the system weaknesses and then uses these vulnerabilities to attack similar devices.

- Radio Frequency Interference: involves an attacker using a device to hinder the connectivity of IoT devices. Jamming and RF interference occur when the attacker is usually in the vicinity of the device's location [47].
- Tampering: In most cases, tampering is the first step of a cyber-attack. It happens when the attacker can physically alter the IoT device and get the encrypted access credential.

2) Network layer: is used for transmitting the data collected in the physical layer to the application and storage places. Below is a list of the most popular attacks in the IoT network layer

- DOS: is an attack that causes different problems such as collision, channel congestion, and battery exhaustion.
- Eavesdropping: In eavesdropping attacks, hackers listen to network communication passing through IoT devices. When a link between two endpoints (such as an IoT device and a server) is insecure, hackers take advantage of this weakness for network sniffing or snooping.
- Man-In-The-Middle MITM attack: this attack happens when the hacker intercepts, alters, and sends data as the original sender to the receiver.
- Storage attack: Data is stored in storage where people and apps can interact with data. The storage should be secured in order to prevent unwanted access to data and underlying storage systems and to ensure that authorized users and apps are only the ones allowed to access.
- Spoofing: To create spoofing attacks in IoT networks, an attacker can generate routing nodes, transmission paths, and fake error messages [48].
- Replay attack: it is a kind of network attack that happens when criminals track down and identify data transmission after which they delay or repeat.

3) Application layer: this layer is varying from one IoT device to another, and there are no common standers to follow to secure this layer. An application layer is the interface between the IoT physical structure and the user. Most of the time, it uses a graphical user interface (GUI) to interact with the device. Below is a list of the most popular attacks in the IoT application layer:

- Malicious code: It is any code that may harm a system.
- Weak application security: it is the weakness in the software code, cryptographic, or access control that can be exploited by a malicious hacker and potentially cause security risks.
- Cross-Site Scripting: When a hacker inserts malicious scripts into a trusted website, the user will run the script regardless of the trusted website, and then the malicious script will access the cookies and any data stored by the browser.
- Phishing Attacks: Phishing is the practice of attackers sending malicious emails meant to lead recipients to fall for a scam. Attackers trick the user into doing things like downloading harmful software by visiting malicious links.

4) Middleware: is a summary layer used in IoT to eliminate the layer between network and application Middleware layer, and it includes brokers, persistent data stores, queuing systems, and machine learning, among others [37]. Middleware has different kinds of attacks such as:

- Man-in-the-Middle Attack.
- SQL Injection Attack.
- Signature Wrapping Attack.
- Cloud Malware Injection.
- Flooding Attack in Cloud.

Finally, Ml and DL may be used on both sides: attack and defense. Hackers may use ML to find holes in the systems, and they are possible to be employed in malicious things.

5.3 Common IoT threats

This section highlights the most popular threats found in most IoT environments. Threats lists are daily updated and add new dangers to the systems. Several approaches are currently in use: traffic analysis, content analysis, application, and user behavior analysis [49]. Malware compromises and challenges the integrity, confidentiality, and availability of the victim's information on hardware or software. Malware is a combination of `mal' from `malicious' and `ware' from `software'. Viruses, Worms, Trojan Horses, Spyware, and Adware are commonly taken examples of it [1].

1) Dental of service DOS/DDOS: are very common in IoT networks than regular networks due to the IoT environment limitations such as low power, low computation, and low capacity which are discussed in section IV under "IOT Limitations and Challenges". Those limitations make the network resources unavailable and disrupt services. Due to the heterogeneity characteristics of the IoT environment and interconnected networks, the malware can easily spread across the network and propagated to the adjacent network through the gateway [50]. DDoS attacks are one of the most severe and frequent attacks in IoT networks. This attack can occur at multiple tiers of the architecture, which makes its detection and resolution increasingly complex [51].

Table 6: Overview of different kind of detection systems on the IoT environment

S#	Detection type	Reference In
1	Malware detection	33-108, 193
2	Intrusion detection	138-140, 142, 143-155, 158-165, 205
3	Attack detection	52, 166, 168-178
4	BOTNET detection	179-192, 194-203, 206, 207
5	Anomaly detection	123- 137, 141, 204
6	DOS DDOS detection	109-122, 156, 157
7	Spam detection	167

2) Hardware and software vulnerability: not all threats are in cyberspace, as physical threats in the device itself are also very important to consider. Sometimes an open port in the device is used remotely by attackers. Universal passwords and weak embedded codes are examples of this kind of threat.

3) Social engineering: it is when malicious activities are done through human interaction. Social engineers trick organizations and individuals to break security or get sensitive data. IoT devices are important for social engineers because it gives them a brief about someone's behavior which is one of the main steps for success to social engineers.

4) User weakness: many studies show that most companies' attacks were because of employees. Social engineering, mail phishing, and other security problems are caused by the lack of security knowledge and training.

6 Datasets

The dataset is a collection of data of different types such as text, image, audio, video, and numeric data that is used to train the model to learn and predict outputs depending on the dataset pattern. So, it is a file that contains many records, where the record is the main unit of information stored in the dataset as a row. Dataset sources on the Internet are from public and private institutions, or from individuals and researchers who collect data themselves. Preparing and choosing the right dataset is a very important part of the machine learning training module. For using the dataset, it must slice into three parts: training data, testing data, and validation data. Testing data is the biggest part of the dataset around, where 60% of the main dataset is used to train the ML model. The testing part uses 20% of the dataset to evaluate the accuracy of the ML model, whereas the validation part also around 20% of the dataset to evaluate the model's parameters after training and testing.

6.1 Popular datasets used in IoT security

IoT system is a collection of devices linked with sensors, applications, and other parts to collect and send data over the Internet. Those systems collect large amounts of data from its environment as records in the dataset. There are many free datasets related to IoT systems on the Internet. In this survey, I focused on the datasets related to IoT security such as in Table VII which contains the IoT security dataset used in all references in this survey for the last five years and all the properties of those datasets.

6.2 Data Pre-processing

Data pre-processing is a combination of data mining and data analysis. It converts raw data to another format that can be understood by the system. IoT data are huge and called big data as it consists of thousand, or maybe millions, of records collected when IoT senses the environment as texts, images, or videos. Raw data are messy data because it is collected from different sources, and it not only had errors but, most of the time, is incomplete or has no uniform design, which is also called unstructured data. Unstructured data should first be cleaned and reformatted before analysis. When the ML model trains bad data, it will produce bad analysis results. Machine learning research requires good and comprehensive data analysis. The first step is to arrange data in such a configuration that it will be compatible with the input of any ML algorithm [52]. Cleaning data is the process of ignoring the missing record, filling missing values manually, filling in using computed values, and correcting errors in data. Getting clear and useful data will directly affect the learning model.

The next step of data pre-processing is data integration which is used to combine data from several sources into a single, larger data storage, such as a server. After that, we do data transformation which means converting the useful data into new forms by changing its structure or format. The last step is data reduction and compression to minimize the size of the dataset to make it easier to handle by data analysis and data mining algorithms. The last step, which is the most important task, is to identify the type of features.

Data cleaning is the process of removing duplicate information, correcting existing errors, and providing data consistency. It is estimated that the anomaly and impurity in the data generally account for about 5% of the total data, which may be even worse for IoT. The data types that need to be cleaned are Incomplete Data, Incorrect Data, and Duplicate Data [22].

6.3 Feature Selection

Feature selection (FS) is one of the optimization techniques that is used as a pre-processing step in machine learning problems to improve or at least maintain the classification accuracy and simplify the complexity of the used classifier. It is considered one of the most critical steps in the process of building an intrusion detection system [53]. Most research in IoT dataset feature selections works on producing a lightweight technique that is compatible with this environment's limitations which is discussed in Section 5 under the title of IoT challenges and limitations. The main idea of FS is to reduce the amount of data by selecting the most important features which affect the

model accuracy and ignoring other less important features. Feature selection is an important process for building a Network intrusion detection system (NIDS) where it helps to remove the noisy features and keep only the features that are relevant to system output [54]. Reducing the number of data samples affect positively in the model performance, reducing the computational processes time, and increasing the testing and training speed.

6.4 Number of hyperparameters

In general, an increasing number of hyperparameters is associated with additional work for the user. Either appropriate values have to be defined through user experience or reasonable parameters have to be found via research or optimization on a set of possible values. A high number of hyperparameters can yield better model performance through extensive parameter search [55].

7 Machine learning and Deep learning

Machine learning is frequently mistakenly with artificial intelligence; however, it is a subfield or form of AI. Predictive analytics and predictive modeling are other names for machine learning. Deep learning also is a domain of AI, and it is a subset of machine learning at the same time. The word "deep" relays to the depth of the layers in the neural network, where the neural network is the base of DL, ML, and AI.

Neural Network (NN) is a set of algorithms nodes used to recognize the relationship between input data through a process similar to human brain operations. Several NNs make a layer, more layers add to the depth, and more in-depth generate Deep Learning which is discussed further in this section.

Table 7: An overview of the IoT systems datasets used in the most of the references

S #	Dataset Name	Dataset purpose	# of records	# of features	Ref. In	Accessible link
1	Bot-IoT	Botnet Detection	72.000.000	43	[114], [142], [148], [154], [160], [165], [166], [183], [198], [199], [200], [201], [202], [206], [231]	[76]
2	NSL-KDD	Intrusion detection	4,898,431 training 311,027 testing	42	[110], [138], [139], [146], [147], [158], [164], [166], [171], [173], [176], [230]	[77]
3	UNSW-NB15	Intrusion Detection	175,341 training 82,332 testing	49	[125], [132], [143], [154], [157], [161], [162], [171], [172], [202], [205], [225], [234]	[78]
4	CICIDS2017	Intrusion Detection	30,540	80	[110], [115], [117], [122], [127], [153] [159], [166], [186]	[79]
5	N-BaIoT	Botnet Detection	7062606	115	[126], [152], [177], [179], [183], [184], [192], [204]	[80]
6	IoT-23	Malicious and benign IoT network traffic	15M biflows	20	[105], [118], [175], [178], [185], [186], [188]	[80]
7	KDDCup99	Classification	4000000	42	[110], [141], [152], [166], [169], [171]	[82]
8	ISCXIDS2012	7 days of network activity which includes normal and malicious traffic for intrusion detection	571,698	-	[109], [110], [113], [117], [157]	[83]
9	CICDDoS 2019	Network traffic classification	50,063,112	80	[117], [120], [194]	[84]
10	CTU-13	Botnet, Normal and Background traffic.	-	13	[186], [194], [201]	[85]
11	TON_IoT	Network traffic	22,339,021 malicious and benign records	45	[139], [140], [145], [155]	[86]
12	CSE-CIC-IDS 2018	Anormal detection	16,000,000	80	[117], [139]	[87]
13	IoTPoT	Malware detection	36,078,737	-	[94], [97]	[88]
14	Virus Share	Malware detection	280	10	[97], [102]	[89]
15	DS2OS	Traces captured in the IoT environment	57,800	13	[137], [52]	[90]
16	Anubis	Malware Detection	9,458	-	[33]	[91]
17	SARD	Program weakness in various languages such as (C, C++, Java, PHP)	Almost 1 million rows	-	[99]	[92]
18	DT Dase Malware	MD5 hashes and behaviors	-	54	[100]	[93]

7.1 Machine learning vs deep learning

Making machines think and behave like humans is the goal of artificial intelligence (AI). Artificial intelligence includes the field of machine learning. Neural networks are the foundation of deep learning algorithms, which are a branch of machine learning as in Figure (8). Machine Learning (ML) refers to intelligent methods used to optimize performance criteria using example data or past experience(s) through learning, ML algorithms build models of behaviors using mathematical techniques on huge data sets, ML also enables based the ability to learn

without being explicitly programmed, these models are used as a basis for making future predictions on the new input data [15]. In actuality, the depth of a neural network—the number of node layers—is what defines a deep learning method, which needs more than three layers. The discipline of Machine Learning is a subset of Artificial

Intelligence is concerned with the capability of computer systems or machines to improve their performance automatically throughout their experience [56]. Deep artificial neural networks, such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs), are a specific family of machine learning algorithms that convert input data into output

ML	DL
Need labeled data	Can handle unstructured data
Work with small dataset	Work with large dataset
A part of AI	Apart of ML
Short training time	Long training time
Lower accuracy	Higher accuracy
needed More human interaction	learns independently from its environment and previous errors without the need for human interaction
Less number of layers in NN	More than three layers in NN
linear correlations	Non- linear correlations
Use CPU	Use GPU
Less levels of algorithms	Many more levels of algorithms
Solve simple problems	Solve complex problems
Machine learning usually breaks down the problem into multiple sub-problems and solves the sub-problems, ultimately obtaining the final result [26].	DL does end-to-end problem solving.
Feature extraction is manual	Feature extraction is automated rather than manual
Examples of ML algorithms: KNN, SVM, Decision Tree, and Bayes.	Examples of DL algorithms: DBM, CNN, and LSTM.

Table 8: Differences between ML and DL

through multiple layers of non-linear transformation. Table 8: Multiple deep-learning solutions are already utilized in IoT forensics [43].

A branch of artificial intelligence called machine learning focuses on making computers capable of completing tasks without user intervention. Data that is organized into rows and columns is sent to the computers. A computer can continuously accept fresh data after it has been programmed, sort it, and take action on it without additional human input. Even if you quit labeling your data over time, the computer could eventually be able to recognize each sample. The core of machine learning is what is known as "self-reliance."

Only a small portion of machine learning includes deep learning. In terms of how each algorithm learns and how much data each type of algorithm consumes, the following presents the main areas where they differ. The concept of DL was proposed by Hinton based on the deep belief network (DBN), in which an unsupervised greedy layer-by-layer training algorithm is proposed that provides hope for solving the optimization problem of deep structure. Then, the deep structure of a multi-layer automatic encoder is proposed [57]. Much of the feature extraction portion of the process is automated via deep learning, which reduces the need for manual human interaction. Large data sets can also be used due to it.

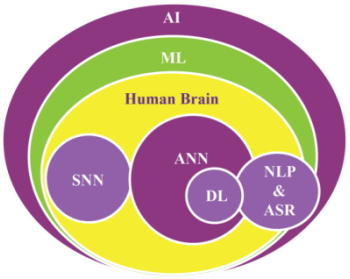


Figure 8: AI taxonomy

The majority of data in an organization is thought to be unstructured, and deep learning can handle unstructured data correctly. Traditional, or "non-deep," machine learning is more reliant on human input. To grasp the distinctions between different data sources, human specialists create a hierarchy of attributes, often learning from more structured data. A labeled dataset is not always necessary for "deep" machine learning. It can take in unstructured data and automatically identify the features that set one sample apart from the others.

As a result, Both ML and DL are types of AI. Machine learning, in essence, is AI that can autonomously adapt with little assistance from humans. While deep learning is a type of machine learning to simulate the way the human brain learns. And Table VIII lists the differences between ML and DL in simple words.

Some of the security-related real-world applications of ML are as follows [15]:

- Face recognition for forensics: pose, lighting, occlusion (glasses, beard), make-up, hairstyle, etc.
- Character recognition for security encryption: different handwriting styles.
- Malicious code identification: identifying malicious code in applications and software.
- Distributed Denial of Service (DDoS) detection: detecting DDoS attacks on infrastructure through behavior analysis.

7.2 Machine learning types

There are four types of machine learning: supervised, unsupervised, Semi-supervised, and reinforcement learning. It is different types to handle all kinds of targets. The data are huge, and we use machine learning because it allows us to cover a large amount of data, learn and make predictions, find patterns, or classify data. In these highly dynamic times, a wide variety of machine learning algorithms have been developed to assist in

resolving challenging problems in the real world. The automatic, self-correcting ml algorithms will get better over time.

In its simplest form, machine learning relies on pre-programmed algorithms that take input data and analyse it to estimate output values that fall within a certain range. These algorithms learn from fresh data as it is given to them, optimizing their processes to increase performance and gaining "intelligence" over time. One of the most popular examples of ML is Covid-19. During the COVID-19 pandemic, many studies have been conducted to explore ML in fighting against the virus to save many lives [58].

1. Supervised learning: Supervised learning in machine learning use a labeled dataset in training the ML model with labeled input to expect the output. Supervised learning keeps comparing the correct output with its output, and the process is still repeated until the model gets the best accuracy. Based on whether target labels are discrete or numeric, the learning process is defined as classification and regression, respectively [59].
2. Unsupervised learning: Unsupervised learning used unlabeled datasets to find the missing dataset and unknown relationships by grouping similar data into clusters. In unsupervised learning the machine will figure out the output without telling the pattern, the output will be classified depending on its similar features using different algorithms.
3. Reinforcement learning: Reinforcement learning is between supervised and unsupervised learning; it has no labeled information as input, but it works with reward values. It involves Learning by observation of the environment to self-train continually using trial and error. Examples of RL are Q-Learning and Deep QLearning [60]. They are trial-and-error learning algorithms; in which training is done through data collected from the environment [61]. We have two main methods for reinforcement learning:

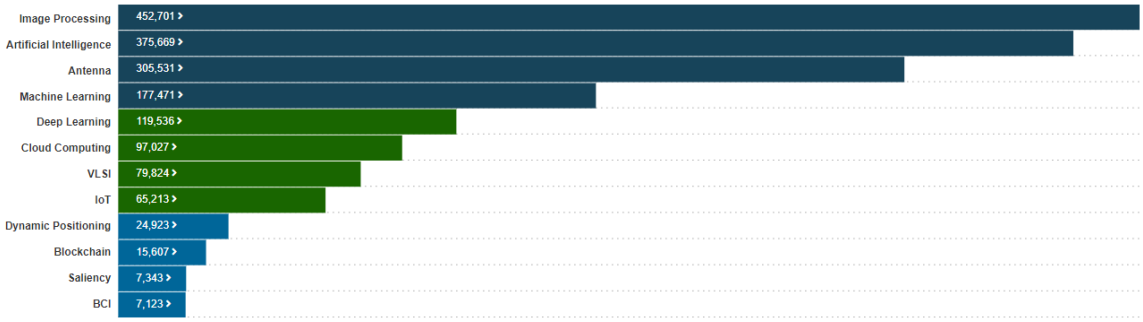


Figure 9: Top Search Terms from IEEE Xplore [65].

- Policy search: This is the search for an optimal policy using gradient-based or gradient-free methods. For example, Google's Alpha Go is based on policy search and can learn without any human intervention or interaction and still achieves superiority [59].

- Value function approximation: This method estimates the expected rewards of actions and attempts to reach an optimized learning process and results. The key component of the value function is the state-action value function, known as the quality function [62].

7.3 ML and DL in IoT security

After the big growth in IoT applications and devices and after being surrounded by varies vulnerabilities, it becomes harder to secure those systems with humans, ML techniques play an important role in regular and IoT networks security. Considering the various vulnerabilities in the IoT domain, ML algorithms are widely being used to tackle the potential Issues [226]. The requirements for securing IoT devices have become complex because several technologies, from physical devices and wireless transmission to mobile and cloud architectures, need to be secured and combined with other technologies. The advancement in ML and DL has allowed for the development of various powerful analytical methods that can be used to enhance IoT security [6]. IoT devices depend on big data, and the most prevalent methods for dealing with big data analytics are machine learning and deep learning methods. Machine learning refers to the deployment of artificial intelligence (AI) to teach a machine (a computer system) by exploring patterns and discovering inferences among unclassified training data without the use of explicitly programmed instructions [63].

ML algorithms build behavioral models using mathematical expression techniques on enormous data sets. Without explicit programming, ML can empower smart devices to learn. Based on new input data, these models serve as a source for future predictions [35]. Not only ML but also DL and AI, in general, are getting expanded rapidly in cybersecurity fields, especially in the early detection and prediction in different domains such as malware detection in references [33] to [180], DOD/ DDOS detection in [109] to [122], intrusion detection as in references [158] to [165], spam detection as in [167], and BOTNET detection in references [179] to [192]. More details are in Table 5 which overviews different kinds of detection systems in the IoT environment.

The Intrusion Detection System (IDS) is a typical system designed to monitor protected networks and systems for malicious activities and is an important approach to protecting cyber infrastructures and enforcing system security [59], Figure (10) from [64] proposed intrusion detection and response methodology. IDSs can also be categorized into active and passive detection systems. Passive detection systems send an alarm to the network administrator when an attack is detected. Then, action is required by the administrator to look at and decide the appropriate decision. On the other hand, active detection systems are responsible for detecting attacks and taking automatic and immediate action to stop or mitigate the impact of the attacks by executing a predefined script [54]. IoT devices can apply supervised learning techniques to evaluate the runtime behaviors of the apps in malware detection. In the malware detection scheme as developed, an IoT device uses K-NNs and random forest classifiers to build the malware-detection model [8].

Yet no unified IoT security standards have been developed. Various organizations, such as IEEE and ETSI, attempt to create IoT slandered for security [17]. ML techniques can address the scarcity available of required personnel with expertise in these niche cybercrime detection technologies. Moreover, vigorous approaches are needed to detect and react against the cyberattacks of the new generation (automated and evolutionary) [1].

AL, ML, and DL are quick solutions against multiple kinds of attacks, they can learn from labeled datasets or from experiences and environments for new attacks which are not listed before in the dataset. Technology is daily updated but at the same time attacks and new vulnerabilities are daily generated, and this is the advantage to use ML in such solutions to predict new attacks. Figure (9) shows how much AI, ML, and Dl trends in the cybersecurity field in the IEEE Xplore search. We can notice how they become more popular in the last decades.

Currently, ML and DL can be used and achieve success in all cybersecurity areas, especially in Big Data which is mostly created from the IoT sensing environments. DL works better with huge datasets with no need for human decisions. AI-based systems give better performance than traditional AI detection systems and reduce the investigation time. Machine learning algorithms have improved and solved many open challenges and problems such as resource control and location in IoT networks [2]. Traditional cybersecurity systems are weak in automation, and they mostly depend on humans and some static rules. This makes ML and DL solutions an interesting area for people who search for automotive cybersecurity systems which take decisions automatically through experience.

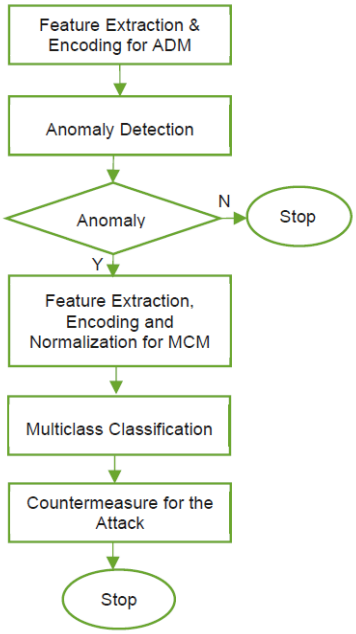


Figure 10: Proposed intrusion detection and response methodology [64].

- Linear Regression (LR): it is a supervised learning method for regression problems and to find the relationship between variables. LR predicts a dependent (y) based on independent (x).

Learning algorithms improve the prediction through learning from training, for example classifying new traffic to be normal or abnormal from training previous behaviors similar to references [123] to [137], [141], and [204].

7.4 Frequently used ML and DL techniques

Under this title, we discuss different types of ML algorithms including supervised, unsupervised, semi-supervised, and reinforcement learning as listed in Figure (11). ML algorithms help solve problems as they are automated and develop themselves over time to get smarter. Machine learning algorithms functions can discover hidden patterns in data, predict results, and enhance performance based on past performance. Those algorithms are various depending on the tasks, such as prediction and classification problems.

- Supervised learning / classification problems:
  - Support Vector Regression (SVR): SVM is a supervised machine learning algorithm that can be used for both classifications of data and regression analysis, but mostly in classification problems [66]. SVR is similar to SVM but with a few differences. SVR works with continuous values in regression problems while SVM works properly in classification problems.

- Decision Tree (DT) is a supervised ML algorithm for classification problems. It makes decisions based on simple decision rules learned from training data. The pros of using this algorithm are it is simple to interpret and can handle the missing value in data well as in Figure (13). The cons are it is prone to overfitting and unstable [67].

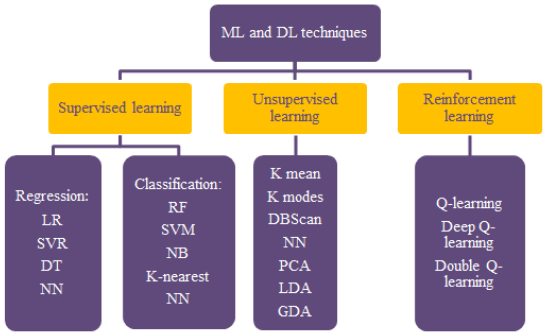


Figure 11: ML and DL techniques.

- Random Forest (RF) is a supervised ML classification algorithm that consists of a large number of DTs discussed above, and because of that, it is called Forest. Each tree in the RF consists of a class, and the class which gets the highest number of votes becomes the model prediction for the problem, as shown in Figure (14).
- Support Vector Machine (SVM): The optimization objective of Support Vector Machine (SVM) is to maximize the distance between adjacent margins between the separating hyperplane (decision boundary) and the training samples that are closest to this hyperplane [68]. SVM is a supervised ML model for solving classification problems using labeled data. The line in the model is the decision boundary where anything that falls on the first side belongs to category 1, and the other side is category 2 as in Figure (15).
- Naïve Bayes (NB): It is a supervised learning algorithm, not a single algorithm. It is a family of algorithms based on Bayes' theorem, all the Bayes algorithms have a common principle, and their results are classified independently of each other.

$$P(A\backslash B) = \frac{P(B\backslash A)P(A)}{P(B)} \quad (1)$$

- K- Nearest Neighbor (KNN): The simplest ML algorithm for supervised classification problems. KNN categorization process works in the similarity between the new data and the existing dataset by calculating the Euclidean distance between the new data point and the categories, where the nearest is the category that belongs to it, as in Figure (16). The k-nearest-neighbor (KNN) anomaly detection is one of the most commonly used distance-based anomaly detection methods. It is a simple technique that works out of the box in most cases and detects global anomalies precisely [69].
- Euclidean distance between

$$A\ (X_1, Y_1) \text{ and } B\ (X_2, Y_2) = \sqrt{(x_2 - x_1)^2 + (Y_2 - Y_1)^2} \quad (2)$$

- Neural Network (NN): An artificial intelligence technique called a neural network allows computers to analyze data in a manner similar to the human brain. NN is built by many neurons or nodes gathered in a layered framework. It has different branches depending on the task we want to solve, the NN types are:

1. Artificial Neural Network (ANN): it is a branch of neural network where the data trained forward pass and vice versa cycles. ANN is easy to use where the activation value is counted in all nodes within all layers including the hidden and the output layer; this is why the activation value affects the classifier performance.
2. Convolutional Neural Network (CNN): it is a network architecture for DL which learns from data, it is commonly used for image recognition, processing, and classification as in Figure (17).
3. Recurrent Neural Network (RNN): it is a type of artificial neural network that has different hidden layers and is used in speech recognition and natural language processing. Some layers in RNN are used as memory locations to store results during the process in a loop.
4. Deep Belief Network (DBN): It is a class of deep neural network, it has multiple hidden layers with different units, and all units interact together through connections.

- Unsupervised learning problems:

- K-means: it is an unsupervised ML algorithm for solving clustering problems for unlabeled datasets. K-means groups the unlabeled data into k different clusters where all the data in each cluster are similar to each other as in Figure (18).
- K modes: k-modes unsupervised learning used clustering to categorize variables. It creates clusters depending on the matching and similarity between data points.
- Density-Based Spatial Clustering of Applications with Noise (DBScan): it is an unsupervised learning method used to separate data points into clusters different in size and shape.
- Principal Component Analysis (PCA): it is an unsupervised learning algorithm used to decrease the number of features in the dataset without affecting the useful information.
- Latent Dirichlet Allocation (LDA): it is an unsupervised learning algorithm but not a clustering algorithm because it does not generate new clusters but distribution groups from the big data.

- Gaussian Discriminant Analysis (GDA): it is a learning algorithm used to find the distribution

for the classes. It is commonly used for the data that can be distributed and fit to Gaussian distribution.

- Reinforcement learning problems:
- Q-Learning: it is a Reinforcement learning algorithm based on values learning. Q-Learning has an agent used to deal with the environment "as input" and take actions "as output." It also uses a Q-table containing rows and columns to help the agent in the next movement.
- Deep Q-Learning: it is similar to Q-learning but the main difference is that Deep Q-learning uses a neural network instead of a Q-table.
- Double Q-Learning: it is a reinforcement learning double faster than Q-learning by reducing overestimation problems with traditional Q-learning.

Most reliable ML systems in intrusion detection use several techniques and vote for the most usable one as in Figure (21) to get the best output.

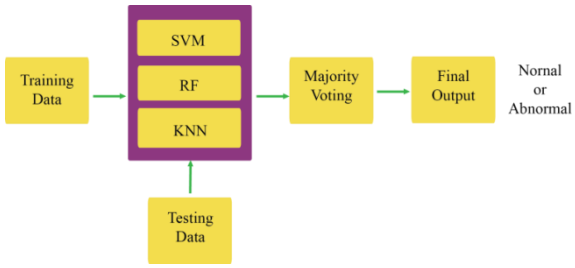


Figure 12: Intrusion detection system process.

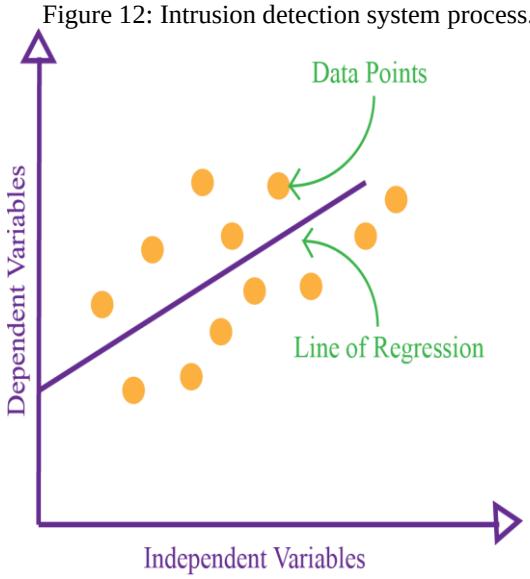


Figure 13: Linear Regression (LR).

7.5 Model selection and evaluation metrics

Model Selection and Evaluation is a very important step in the machine learning process to analyze the model. Model evaluation is the way to estimate the correctness of the model over the test data which has not been applied in the model before, while model selection is the method to select the best model for the data after comparing multiple models and checking their performance. We do model evaluation and selection to get the best predicts results and increase the accuracy levels to a higher percentage and this is done depending on different matrices and Scoring.

Before discussing matrices and equations we must know the meanings of some terms related to the topic. The Model is the thing learned and saved after applying machine learning algorithms in training data to find patterns or make predictions. The Learning algorithms are a collection of instructions and mathematical operations used in machine learning to make the computer work like a human mind in solving problems, finding patterns, and classifying. The hyperparameters are the effective parameters whose values really can control the model, the hyperparameter is set before the learning process begins and they are very important because they influence the behavior learning algorithm and the performance of the model. We should choose the most suitable hyperparameter to get a successful score.

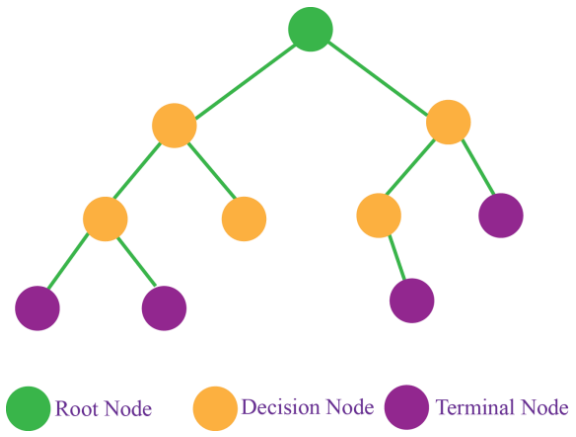


Figure 14: Decision Tree (DT).

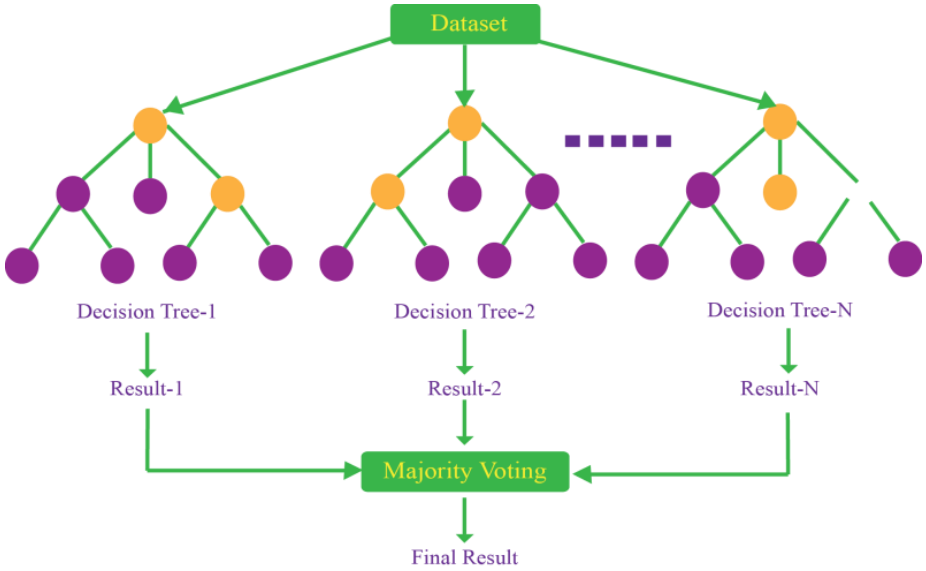


Figure 15: Random Forest (RF).

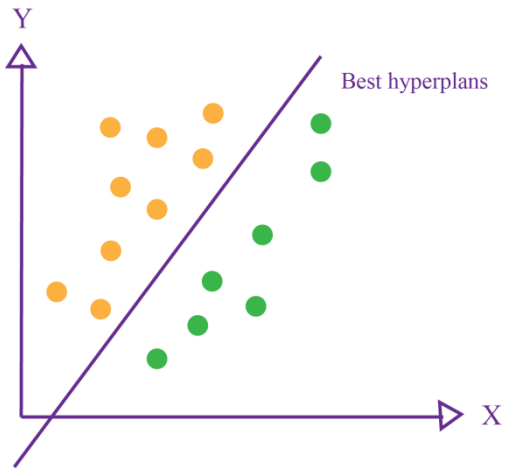


Figure 16: Support Vector Machine (SVM).

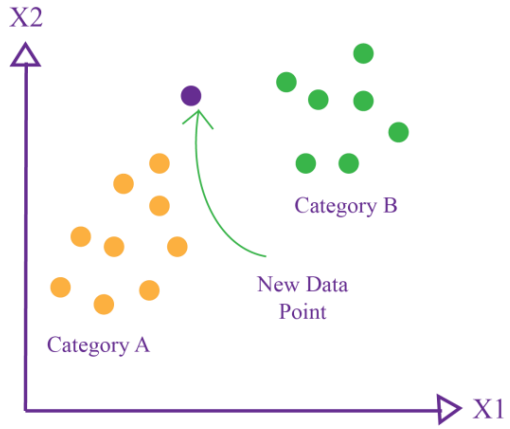


Figure 17: K- Nearest Neighbor (KNN).

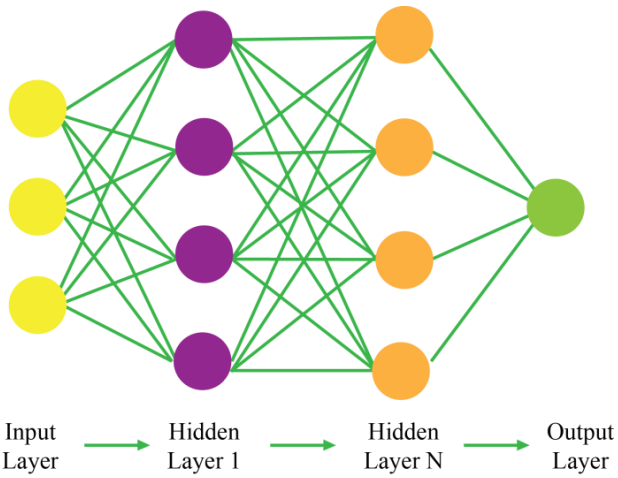


Figure 18: Convolutional Neural Network (CNN).

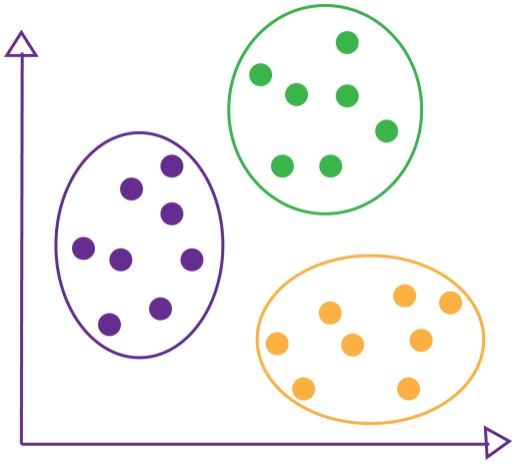


Figure 19: K-mean

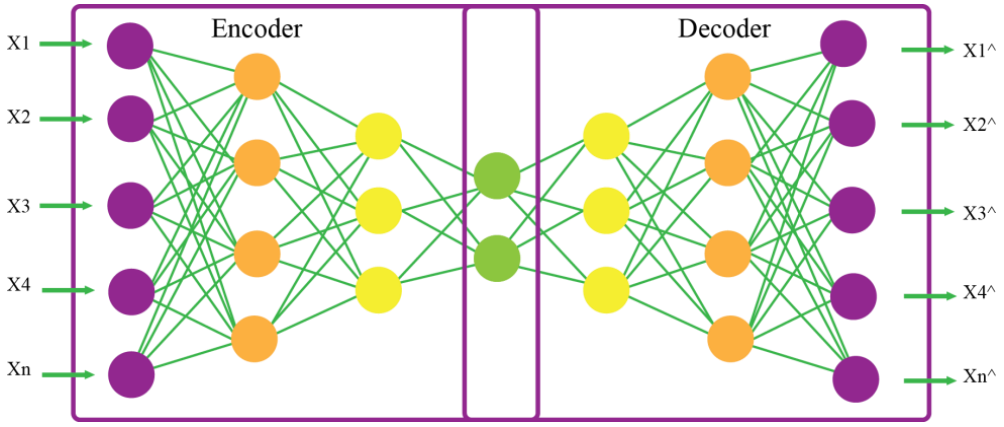


Figure 20: Autoencoder

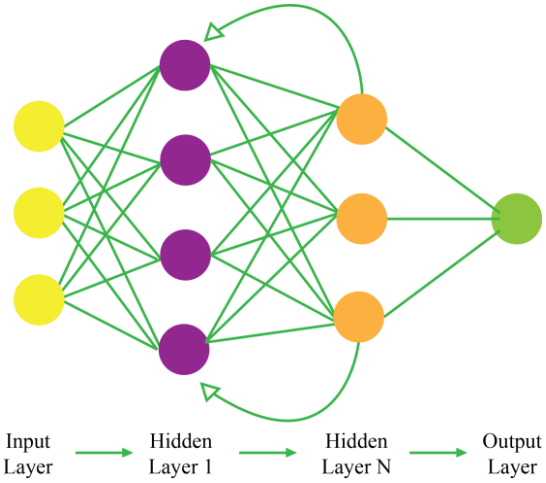


Figure 21: Recurrent neural network (RNN).

Table 9: An overview of top Ml techniques and algorithms used in the references

Ml tech./ algorithms	Paradigms	References In	Advantages	Disadvantages
RF	Supervised	33, 96, 97, 98, 102, 103, 104, 106, 109, 113, 120, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 127, 132, 133, 134, 140, 143, 145, 149, 151, 152, 153, 155, 160, 162, 164, 169, 170, 175, 176, 178, 179, 181, 182, 183, 184, 186, 189, 192, 193, 196, 197, 199, 201, 205, 206, 207	- Reduces overfitting. - Used for classification and regression. - Slower for large number of trees. - Works with categorical and continuous values. - Avoids data-overfitting problem. - Powerful and high accurate. - Can handle multiple features at once. - No need for normalizing.	- Need much computational power. - Need much time for training. - More accurate need more trees, which cause slower process. - Difficult to interpret. - Not easy to understand predictions. - Cannot use for linear problems. - Cannot handle large datasets.
SVM	Supervised	95, 98, 102, 103, 106, 109, 111, 113, 115, 117, 119, 121, 122, 129, 131, 133, 134, 135, 138, 140, 141, 143, 145, 146, 148, 149, 150, 151, 157, 164, 165, 168, 170,	- Works well with high dimensional spaces. - Highly accurate. - Handle many features. - Works with structured and unstructured data.	- Not acceptable for large data sets. - Low speed. - Need more time for large dataset. - Need labeling for input data.

		171, 173, 182, 183, 185, 187, 189, 192, 194, 198, 199, 202,203, 205, 206, 207		
DT	Supervised	95, 100, 102, 103, 106, 109, 117, 118, 119, 123, 124, 140, 142, 143, 153, 155, 162, 166, 168, 170, 171, 173, 174, 175, 176, 179, 185, 190, 192, 195, 196, 197, 198, 201, 202, 205, 207	- Easy to understand. - Easy to implement. - Works with categorical and numerical values. - High speed. - Better results for larger data.	- Data over-fitting problem. -Too simple and cannot handle complex data.
KNN	Supervised	33, 95, 98, 103, 104, 109, 111, 112, 113, 117, 119, 127, 130, 140, 143, 145, 147, 149, 153, 157, 166, 170, 173, 174, 176, 178, 179, 181, 182, 183, 186, 193, 195, 202, 203, 206, 207	- Easy to understand. - Easy to implement. - Single hyperparameter. - Used for classification and regression. - Can handle multi-classes. - Higher accuracy compared with other supervised learning models.	-Sensitive for noisy dataset. - Sensitive to outliers. - Need much computational power. - Need much time for training. - Need larger space.

Creating and choosing a model that makes accurate predictions on the dataset is the first step of building a model. It is followed by a feedback mechanism, and we can get the feedback from metrics, and depending on the metrics results, we make adjustments and keep going until the results we want are achieved. In regression problems, the model results are continuous values, and we aim to be as close to these values as possible. In classification problems, we classify data into a finite number of classes. Figure (22) shows the evaluation models for both supervised learning types: classification and regression.

The Confusion matrix, which is also called the error matrix, is the matrix which is the formal way to represent the results of the classification model. The confusion matrix presented in Table 9 shows the results of a binary classification in form of True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN). Binary classification is the task to classify the data into two groups, such as classifying emails into spam or not spam, breast test as cancer detected or cancer not detected, and COVID-19 tests as Positive or negative.

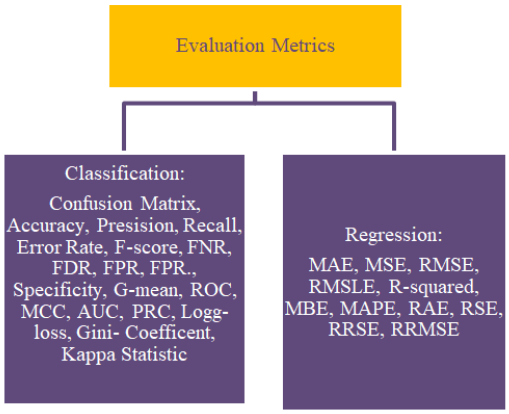


Figure 22: Evaluation metrics for classification and regression models.

Table 10: An overview of top Ml techniques and algorithms used in the references

Ml tech./ algorithms	**Paradigms**	**References In**	**Advantages**	**Disadvantages**
NB	Supervised	33, 95, 96, 97, 102, 115, 117, 120, 123, 124, 127, 133, 142, 143, 145, 151, 152, 155, 164, 170, 174, 175, 183, 184, 185, 186, 194, 197, 199, 202	- Easy to implement. - Fast and flexible. - Handle large datasets. - Spends less time. - Can handle missing data.	- Precision decrease when the data decrease. - Good results need large data. - Lower performance compared with other classifiers.
LR	Supervised	95, 106, 117, 120, 123, 126, 129, 132, 137, 142, 145, 149, 151, 155, 162, 164, 166, 170, 171, 182, 183, 186, 190, 197, 204, 205	- Easy to implement. - Easy to interpret. - Easy to understand. - Works well with linear dataset. - Reduce data-overfitting problem.	- Sensitive to outliers. - Prone to noise and overfitting
ANN **NN**	Supervised Unsupervised	95, 103, 109, 114, 119, 121, 126, 137, 52, 148, 158, 161, 162, 172,	- Can handle complex task. - Best algorithm for image recognition.	- Slowest algorithms because of having many layers.

	Reinforcement	174, 182, 182, 184, 189, 197, 203, 204		- Impossible to understand predictions.
MLP	Supervised	112, 115, 117, 122, 123, 127, 134, 156, 157, 171, 190, 194	- Solving different kinds of tasks. - Needs fewer parameters. - Easy to design.	- Low speed. - Data over-fitting problem. - Need much computational power.
LSTM	Unsupervised	97, 112, 115, 121, 145, 154, 183, 200, 204	- Can handle noise. - Can handle continuous values. - High accuracy in prediction. - Needs few parameters.	- Too simple. - Cannot handle complex tasks. - Require a lot of resources. - Requires much time.
K-mean	Unsupervised	106, 132, 138, 169, 171, 177, 180	- Simple. - Easy to understand.	- Requires a number of clusters. - Cannot handle categorical values. - Sensitive to outliers.
RNN	Supervised	94, 112, 204	- Can handle any input length. - Remember all information. - Model size is fixed and does not increase if the input increases. -Shared weight in all steps and time.	- Slow computation. - Difficult training. Able to store only one layer of data.

The Confusion matrix values are used in specific equations to find other results; the following are the meaning of each value in the cybersecurity field which then follows with the equations and the idea of each one:

- True Positive (TP): In general TP is the number of outcomes where the model success to predicts the positive class. In the cyber security field, TP is the number of normal network traffic that is not attacked and is correctly classified by the model (Predicted True and True in reality).

Table 11: Confusion matrix

		Actual		
		0	1	
Predicted	0	True Positive (TP)	False Positive (FP)	R1
	1	False Negative (FN)	True Negative (TN)	R2
		C1	C2	

- True Negative (TN): in general, TN is the number of correct results for the negative class. In the cyber security field, TN is the count of results when the network flow is classified as an attack, and it is an attack (Predicted False and False in reality).
- False Positive (FP): in general, FP is the number of outcomes where the model incorrectly predicts the positive class. In the cyber security field, FP is also called a False alarm where the network flow predicts to be an attack, but it is not an attack but a normal traffic (Predicted True and False in reality).
- False Negative (FN): in general, FN is the number of outcomes where the model incorrectly predicts the negative class. In the cyber security field, FN is the number of network flows that predicts to be normal, but, unfortunately, it is abnormal (attack), and this is a vulnerability in the system to allow abnormal traffic to flow in the system, device, or network (Predicted False and True in reality).

The hold-out method is the simplest and easiest way to evaluate the model, this method split the data into two sets, one called the training set and the other one is the Test set. The flowchart in Figure (23) explains the process of Hold-Out. The hold-out process has four main steps, step 1 is to split the data randomly, and most of the time the split ratio is 2/3. We keep test data aside and work with the training part. After this, we go to step 2 in which we chose a learning algorithm with specific hyperparameter values which we think may be appropriate for the problem and then run the model. In step 3 we use test data to evaluate the model, find out the performance and the model accuracy and bring the results to the final step. In step 4, which is the last step, we check if the model scores are good enough to solve the problem. If yes, we keep it as the final model, and if not, we change the hyperparameter values, select another learning algorithm, and run the model again, and so on until we find the best values.

The neural networks can successfully classify multiple class classification problems by automatically learning their features via neural nodes at each hidden layer and classifying the input at the output layer [70]. Model evaluation and selecting the best model in Hold-out and any other methods depend on different scores which are the results of doing some calculations on confusion matrix terms as in the following evaluation equations:

1. Accuracy is the most popular evaluation metric, and it refers to how the model correctly classifies the data; it is the ratio of correct predictions to all data in the dataset. If we go back to the network traffic example, the model will work well if it can classify normal traffic (positive class) and abnormal traffic (negative class) correctly.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

2) Error Rate is also called Misclassification, and it is the opposite of accuracy. While accuracy is the correctly classified data, the error rate is the incorrect classified data, and it is the ratio of incorrect predictions to all data in the dataset. In the cyber security network traffic example, the error rate is the number of normal traffic that is classified as abnormal and vice versa.

$$\text{Error Rate} = \frac{FP + FN}{TP + TN + FP + FN} \quad (4)$$

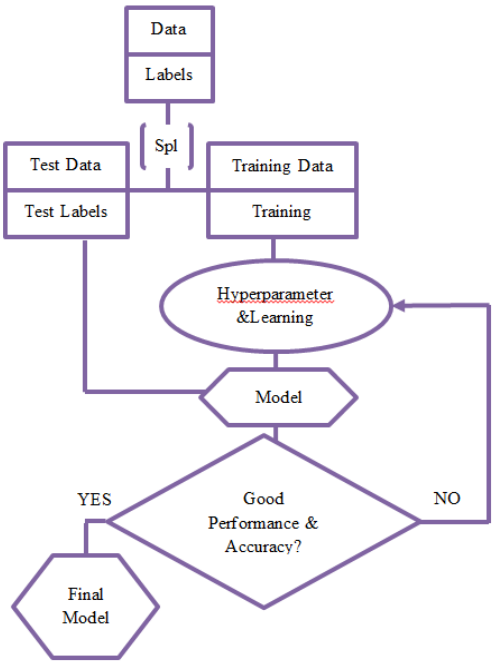


Figure 23: Hold-out processes.

3) it is also called positive predictive value; it is all about YES (positive), it is the ratio between data correctly classified positive (normal traffic) to the total number of data classified positive, even if it is correct or not (normal traffic + abnormal classified as normal).

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

The precision increase when the nominator is huge compared with the denominator, which means it is

8) False Omission Rate (FOR): it is the ratio between incorrectly classified positive data to the total of positive samples in the whole dataset. The lower result for FOR describes better performance.

$$\text{FOR} = \frac{FN}{FN + TN} \quad (11)$$

9) F1 Score: it is a harmonic combination of precision and recall into a single metric to measure the model accuracy. This measure will be helpful if the user seeks a balance between recall and precision, and the sample distribution is an uneven class distribution. A higher value of the F1-score shows the ML model is performing better than other models [1].

increase when the model results in a lot of correct positive classified data, and when the model makes the least amount of incorrect positive classification. Because of this precision is more accurate than accuracy to decide the model performance. Precision reflects the reliability of the model in classifying positive data.

4) Recall, or Sensitivity, is also called True Positive Rate (TPR), and it refers to the ratio of positive classified data to all positive data including missed positive predictions (FN). While precision care about correct positive classified data out of positive predictions only, recall cares about the same plus the missed ones.

$$\text{Recall} = \frac{TP}{TP + FN} = \text{Recall}_{\text{Positive}} \quad (6)$$

When to use precision or recall? This depends on the kind of problem you want to solve. In cybersecurity problems, which are more sensitive than other problem, recall is used. It is OK if you miss one of the cat photos when the model puts it with raccoons' groups, but it is a big problem to miss one attack traffic, add it to the normal traffic group, and let it flow in your network. So, in my opinion, I prefer to select a model with higher recall scores in most of the cybersecurity problems.

5) False Negative Rate (FNR): Also called the (Miss Rate). It is the proportion of incorrectly classified positive samples to all positive samples.

$$\text{FNR} = \frac{FN}{TP + FN} \quad (7)$$

6) False Positive Rate (FPR): Also called the (Fall Out). It is the proportion of incorrectly classified negative samples to all negative samples

$$\text{FPR} = \frac{FP}{FP + TN} \quad (8)$$

In addition, we can write the FPR equations as:

$$\text{FPR} = \frac{FP}{FP + TN} = \frac{FP+TN-TN}{FP + TN} = 1 - \frac{TN}{FP + TN} = 1 - \text{Recall}_{\text{Positive}} \quad (9)$$

7) False Discovery Rate (FDR): it is the ratio between incorrect classified data to the total of negative samples in the whole dataset. The lower result for FDR describes better performance.

$$\text{FDR} = \frac{FP}{FP + TP} \quad (10)$$

$$\text{F1 Score} = 2 * \frac{\text{precision} * \text{Recall}}{\text{precision} + \text{Recall}} \quad (12)$$

10) Specificity: it is also called True Negative Rate (TNR), it is the opposite of recall, while recall is the ratio of positive classified data to all positive data including missed positive predictions (FN), Specificity is the ratio of negative classified data to all negative data include missed negative predictions (FP).

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (13)$$

11) G-mean: it is also called the geometric mean, it is calculated using the correct classified data, G-mean is more powerful and helpful in accuracy when the number of negative data is huge compared with positive data, and

in this case, G-mean is more correct to describe the model performance than the accuracy.

$$\text{G-mean} = \sqrt{Recall * Specificity} \quad (14)$$

$$\text{Or, G-mean} = \sqrt{TPR * \frac{TN}{FP+TN}} \text{ where } \frac{TN}{FP+TN} = (1 - \frac{FP}{FP+TN})$$

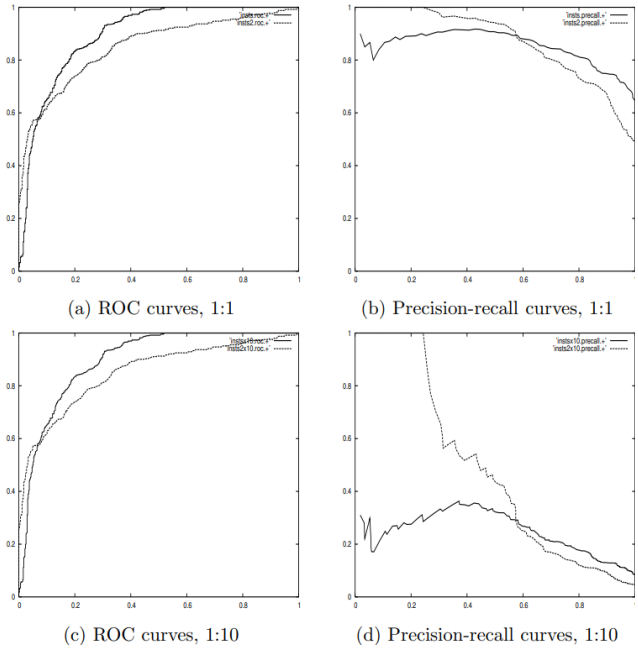
$$\text{So, G-mean} = \sqrt{TPR * (1 - FPR)} \quad (15)$$

12) Matthews Correlation Coefficient (MCC): It is the relationship between the predicted classes and the fundamental truth. MCC is commonly recognized as a balanced measure that can be applied even when the classes are in various sizes.

$$MCC = \frac{TP}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (16)$$

13) Receiver Operating Characteristics (ROC) Curve: The TPR and FPR trade-off is represented graphically by the ROC curve as in Figure (24). We compute TPR and FPR for each threshold and plot them on a single graph. The better, the greater TPR and the lower FPR for each threshold. The AUC score, which is found in the region below the ROC curve, expresses how effective the ROC curve is. Both ROC and AUC scores demonstrate how well the model ranks predictions.

14) PRC Area (Precision-Recall Curve Area): Positive Predictive Value (precision) and TPR are combined to create the PRC curve as in Figure (24). Positive Predictive Value and TPR are computed for each threshold, and the appropriate graph point is presented. A higher AUC indicates a good PRC curve.



(a) ROC curves, 1:1 (b) Precision-recall curves, 1:1

(c) ROC curves, 1:10 (d) Precision-recall curves, 1:10

Figure 24: ROC and precision-recall curves under class skew. In (a) and (b), the test set has a balanced 1:1 class distribution. Graphs (c) and (d) show the same two classifiers on the same domain, but the number of negative instances has increased tenfold [23].

15) Area Under Curve (AUC): The area under ROC is measured by the AUC, which ranges from 0.5 to 1.0. A higher AUC value indicates that a classifier is performing better.

16) Log Loss: it is also called Cross-entropy. How closely the prediction value matches the true value is indicated by log-loss. The log-loss result is 0 or 1 in case of binary classifications. Lower the log-loss result shows better predicted.

$$\text{Log-loss} = -\frac{1}{n}\sum_{i=0}^{n}(y_i \log(\hat{y}_i) + (1 - y_i) . \log(1 - \hat{y}_i) \quad (17)$$

Where n is the number of observations, is the given record, y is the true value, and ŷ is the prediction probability for this formula and all other ones.

Gini Coefficient: it is calculating the probability of a specific value which misclassified when it is randomly selected. It is similar to the range of log-loss, Gini values between 0 and 1. Where 0 shows the purity of the classification model. Gini is the sum of squared probabilities for each class.

$$\text{Gini} = 1 - \sum_{i=1}^{n}(p_i)^2 \quad (18)$$

Where Pi is the probability of an element being classified for a distinct class.

18) Mean Square Error (MSE): It is one of the "Metrics for Regression" which is calculated by averaging the squared difference or error between the classifier's actual and predicted values. MSE is popular because it is simple, continuous, and separable. Taking into account one significant outlier may result in a massive value, which can result in significant errors. A lower MSE value is preferable and indicates better classifier performance.

$$\text{MSE} = \frac{Squared\ Error}{n} \quad (19)$$

Where n is the total number of samples, and Squared Error = $(y_i - \hat{y}_i)^2$

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (20)$$

19) Root MSE (RMSE): It is calculated by taking the square root of the MSE. A lower RMSE value is preferable and indicates better classifier performance.

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (21)$$

20) Mean Absolute Error (MAE): MAE calculates the mean of the absolute difference between prediction and actual data returned, where the outlier values in the train set have no effect on MAE. This metric can be calculated by taking the average of the absolute difference or error that occurred between the actual values and predicted values of the classifier. A lower value of MAE is desirable and shows better performance of a classifier [1].

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (22)$$

21) Mean Bias Error (MBE): bias in "Mean Bias Error is a tendency of a measuring technique to

overestimate or underestimate the value of a parameter. Bias only has one direction and can be in either a positive or negative direction, a positive one indicates an overestimation of the data error, whereas a negative bias indicates an underestimation of the data error. Taking the mean of the difference between the expected values and the true values is known as the mean bias error (MBE). With the help of this evaluation metric, the total bias is quantified, as well as the typical bias in the forecast. The only difference between it and MAE is that the MAE takes the absolute is considered. It is important to use caution when using this evaluation metric because positive and negative errors might cancel one another out.

$$MBE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i) \quad (23)$$

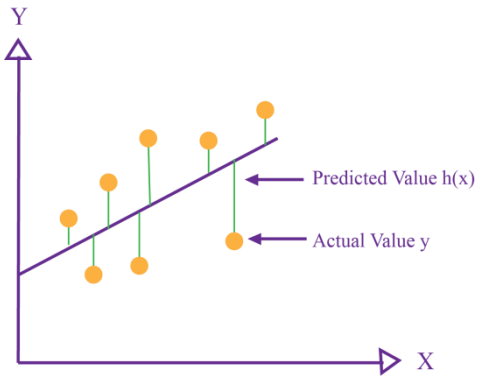


Figure 25: Mean Square Error (MSE).

22) Mean Absolute Percentage Error (MAPE): It is also called The Mean Absolute Percentage Deviation (MAPD), and it is just the difference between the observed and the real value. It is calculated by dividing the absolute error by the actual data. The relative error is then multiplied by 100 to produce the percentage error. The lower MAPE value shows better performance.

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (24)$$

23) Relative Absolute Error (RAE): It is calculated by dividing the total absolute error by the absolute difference between the mean and the actual value.

$$RAE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{\sum_{i=1}^{n}|y_i - \bar{y}|} \quad (25)$$

Where $\bar{y}$ is the mean of the n actual values, and its formula is:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad (26)$$

24) Relative Squared Error (RSE): it is calculated by dividing MSE over the square of the difference between the true value and the mean of the whole data.

$$RAE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (27)$$

25) Root Relative Squared Error (RRSE): It is calculated by taking the square root of the RSE.

$$RRSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (28)$$

26) Relative Root Mean Squared Error (RRMSE): is the RMSE normalized by the RMS. The difference between RRMSE and RMSE is that the RMSE is restricted with original measurements and each predicted value is scaled against the true value, but in RRMSE there are different measurement methods. The lower RRMSE is the better for model accuracy, and when RRMSE increases, it means your predictions are inaccurate.

$$RRMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(\hat{y}_i)^2}} \quad (29)$$

27) it is the root of the difference between the log for actual values and the log for predicted values. RMSLE is important because it is handled the outliers and treated them even though they are large outliers or small.

$$RMSLE = \sqrt{(\log(y_i + 1) - \log(\hat{y}_i + 1))^2} \quad (30)$$

28) R-squared (R2): it is an evaluation method for linear regression; used to calculate how close are the predicted data to the fitted line. R2 shows the ratio of the total difference between data points and the filling line, in addition to the total difference for the data from the mean.

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (31)$$

Where SSR is the Some Squared Regression error, SST is the Sum Squared Total Error as shown in.

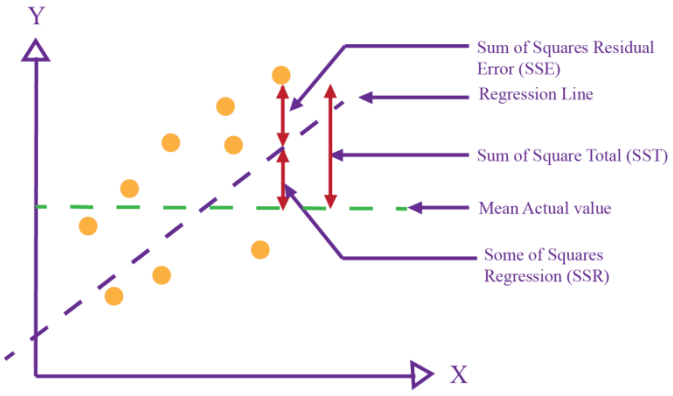


Figure 26: R-squared calculation.

29) Kappa statistic - Cohen's Kappa coefficient (k): It is applied to evaluate the model's predicted labels to the actual labels in the data, so it shows how many data or records classified by the ML model are matching the true one. K result ranges from -1 (worst performance) to 1 (best performance). In the problem of binary classification pe = pe1 + pe2; pe1 - the probability that the predictions agree randomly with the actual values of class 1 - "good"; pe2 - the probability that the predictions agree randomly with the actual values of class 2 - "accidentally". The assumption is that the two classifiers (model prediction and actual class value) are independent. In this case, the probabilities pe1 and pe2

are calculated by multiplying the share of things in the class and the share of the predicted class [72].

$$K = \frac{(p0 - pe)}{(1 - pe)}$$

Where:

The total samples is $N = TP + TN + FP + FN$

The Total of first raw is $R1 = TP + FP$

The Total of second raw is $R2 = FN + TN$

The Total of first column is $C1 = TP + FN$

The Total of first second is $C2 = FP + TN$

The total accuracy of the module $p0 = \frac{(TP+TN)}{N}$

The random accuracy of the model

$$pe = \left(\frac{pe\,(C1)}{N} * \frac{pe\,(R1)}{N}\right) + \left(\frac{pe\,(C2)}{N} * \frac{pe\,(R2)}{N}\right)$$

8 List of ML and DL challenges in IoT security and current solutions

In many IoT systems and applications, machine learning (ML) techniques are essential to recognize the cyber-attack in those systems. However, there remain major difficulties and challenges in guaranteeing the reliability of ML techniques in cyberspace, there are many ML weaknesses and vulnerabilities that may use by bad hackers to damage the systems. Understanding which methods are suitable for protecting IoT systems is a challenge because of the extensive variety of IoT applications and scenarios [17]. This survey has many targets but the most important one is the challenges that faced the researchers in the last 5 years when they use ML to improve IoT security, including intrusion detection, spam detection, and malware detection on IoT systems and networks.

Additionally, it offers concise explanations of each ML technique, frequently used security datasets, necessary ML tools, and assessment metrics for classification model evaluation. The difficulties of using ML approaches to cyber security are fully covered in this section with the most recent comprehensive citation as well as the most recent ML in IoT security developments. Currently, researchers are focusing on the urgent need of finding new automated security methods to cope with these security challenges. One of the best and the most effective considered practices is to use automated machine learning techniques to detect new and previously unseen cyber threats [1].

We can summarize the IoT security challenges we found in references in the main points which will be discussed below and followed with some future directions if found:

- ML and DL need enough data.
- Data format is different because it is collected from different resources.
- Big Data and huge datasets need more computational time and resources.
- Producing new data samples, especially in malware and zero-day attack.
- Most ML techniques need labeled and high-quality data.
- How ML work in data with respect to privacy, access, availability, and safety.
- The integration of cloud and IoT.
- The heterogeneity of IoT.
- Need for lightweight procedures and algorithms which make ML solutions compatible with the IoT limitations.
- Computational complexity and resource consumption.
- Algorithms challenges.
- Other researchers' solutions, and some future directions and suggestions.

One of the challenges of using ML techniques is that it needs enough amounts of data while training the module, Shaukat et al. 2020 [1] suggest using multiple GPUs, which is neither a power-efficient nor cost-effective solution and must have powerful and robust ML techniques that are specifically designed to deal with security attacks and handle adversarial inputs instead of having traditional ML techniques. On the other hand, One ML model cannot detect all attacks, so securing a full environment means having multiple ML models to handle all possible cyber-attack. These challenges include the generation of labeled data required for the effective training of the model because a network traffic dataset is multi-part and irregular [17].

Malware exists in a way that can be copied to 3 million new samples in an hour, and some new attacks are able to bypass end-point detection and can be launched at variable rates [209]. Toward solving some of the aforementioned issues, potential solutions include using high-dimensional data and incremental learning for non-stationary data. Using high-dimensional data can increase model complexity, accuracy, and diversity of the

features for malware fingerprinting [59]. High-dimensional data solve the huge growth in malware new records, but at the same time it may affect the training and testing speed, researchers try to solve this problem for example Liang et al. [210] investigated an online learning strategy that utilized slices of continuous data to update machine learning models to fit the new datasets dynamically. On the other hand, distributed learning ML methods can also handle this problem, like what the researchers improve in [211] after designing and deploying different distributed platforms and algorithms. Another research direction and future work are to optimize computing resource management as in reference [212].

Authors in [130] developed a solution to solve the challenge of anomaly detection which is done because of the limitation in anomaly data records. They developed an unsupervised the ANN (Artificial Neural Network) detection model based on the LSTM-based Auto-Encoder. The LSTM cell can capture temporal dependencies in multivariate sensor data. The Auto-Encoder network architecture is used to learn the normal behavior without a labeled dataset. Once the model is well trained, the online evaluation data is fed into our model as a set of sequences by a sliding window for real-time anomaly detection. Finally, an alert is triggered when a time interval is voted as an anomalous data point by the majority. Another way to solve this problem is using Bidirectional LSTMs (BiLSTMs) instead of the regular ones. BiLSTMs simply by running two different LSTMs - one going from start to end and the other in the opposite direction, once the LSTMs finish moving across the sequence, the outputs obtained during the forward and backward LSTMs are combined for a final prediction [228].

The openness of IoT and the expansion of attack scale. Openness is reflected in the various processes of IoT systems. IoT can obtain data from various fields, integrate various communication technologies and standards, and provide open services for users in various fields [22]. The massive data collected by sensors in CPS and IoT can use wireless communications to exchange the data as well as the data can be used with AI/ML to address the challenges associated with communications in CPS/IoT, including communication-related issues [6]. Noisy Data Is a Challenge for ML and DL Mechanisms: Most of the real-world data is embedded with noise that negatively affects the learning models used for classification. DL algorithms have better classification capabilities as compared to traditional ML algorithms; however, these algorithms are undermined by noisy data [15].

Preventing Zero-day attacks or “new attacks” is another challenge for ML in cyber-attack detection because ML techniques depend on training previous features on the datasets, while new attacks have different features and may not be detected in the same way. We may depend on the behavior to detect new attacks in the future, which means focusing more on unsupervised learning with unlabeled data, Bout et al. 2022 in [2] talked about deduction of the behavior of the IoT network to find the optimal strategy autonomously, but they found that only one attack responds to this motivation in the literature which is the jamming attacks. Data, in general, is a major challenge in ML. Most datasets are out of data with various numbers of features in each one and, sometimes, it has no complete records. The number of features and categories for each dataset is different. Also, dataset shift is another problem where the model was trained and tested with different datasets, which may avoid by removing the leaked data or changing the training data. The key challenge in malware data augmentation is to produce new samples that preserve adequate data distribution for each class. This will improve the classification accuracy of DL methods since improving the coverage of collected data translates to better detection capabilities of new, and existing, malware attacks [16].

However, the performance of DL-based methods strongly depends on the quantity and quality of the data available [12]. Machine learning models have strict requirements for the sizes, shapes, and types of input data, even though CPS collects massive data, the quality of such data may not be guaranteed, especially as the lifetime of newly created IoT hardware will be unverified [59]. To fill the gap of need to truth and high-quality datasets, [213] is the first to propose a benchmarking data set containing vulnerable source code collected from nine open-source software projects written in C programming language. Their data set offers labels at two levels of granularity i.e., the function level and the file level. However, their proposed data set is still at the preliminary stage since it only consists of around 1400 vulnerable functions and 1300 vulnerable files.

Another performance challenge is maintaining the module performance over time. Most existing ML IDS gives high accuracy, but they should be controlled over time because of adding new data continuously. Abdel Wahab [233] discuss an online outlier detection technique that identifies the outliers that diverge both from historical and temporally close data points. His study was about an online deep neural network (DNN) that dynamically adjusts the sizes of the hidden layers based on the Hedge weighting mechanism, thus enabling the model to steadily learn and adapt as new intrusion data come. the results suggest to his solution reduces the FP by approximately 6% and FN by approximately 4.5%, compared to the static DNN model.

The ML challenges in IoT security are not only the amount or the form of the data but also the way to use this data taking into account its privacy. Security and privacy are two of the main factors in the commercial realization of IoT services and applications [15]. Data security and privacy are challenges in processing and using data. In the process of data collection and transmission, you may face the risk of leakage of privacy [22]. Fernandes et al. [214] focused on similarities and differences in the security issues in IoT and traditional IT devices. Furthermore, they focused on privacy issues. The main driving factors to argue on the similarities and differences include software, hardware, network, and applications. The existing security trade-offs, such as that

between availability and safety, are another challenge to the achievement of a robust security scheme for IoT systems. Moreover, the importance of various security trade-offs differs from one IoT application to another. For example, an IoMT system should provide a security scheme, but it should also offer the flexibility of being accessible in emergency situations [17].

Mehedi et al. [229] solve the new challenges with a small amount of data, they proposed in their paper a dependable intrusion detection system for IoT. The unique contributions include effective attribute selection, which is best suited to identify normal and attack scenarios for a small amount of labeled data, designing a dependable deep transfer learning based ResNet model, and evaluating considering real-world data. Their model improved the IDS performance compared with previous research. The overall accuracy of the proposed detection model is 87%, ensuring dependability and low time complexity. Moreover, the model greatly improves the precision score of 88%, the recall score of 86%, and the f1-score of 86%, which are higher than the benchmark models.

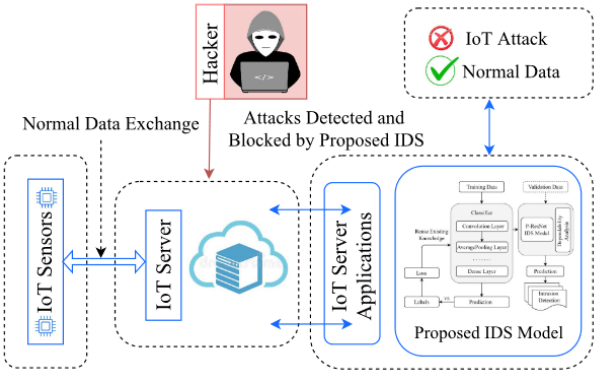


Figure 27: Overall scenario of the proposed IDS applied to the heterogeneous IoT network [229].

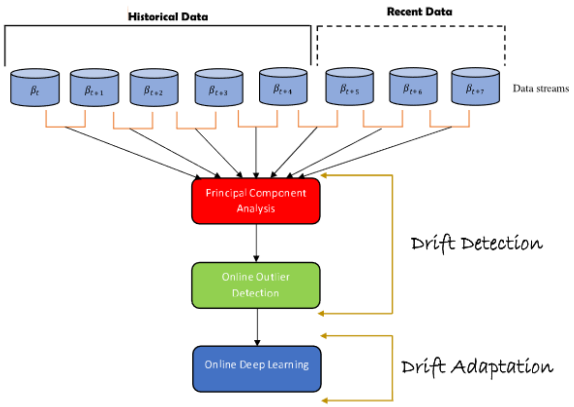


Figure 28: Overall architecture of Abdel Wahab solution. [233].

IoT has also an architecture challenge because of using the clouds. Because of the IoT limitations especially in capacity, it relays mostly on clouds, and this causes a new challenge in processing, storing, analyzing, and controlling the data. In the Cloud, a wide range of security challenges are observed due to two factors: misconfiguration in hardware equipment, and poorly secured Cloud databases. These caused the theft or leakage of 70 million records in 2018 [16]. It is worth mentioning that cloud can be an essential part of the IoT infrastructure due to several-fold reasons: handling big data, storing and processing huge amounts of data from IoT, and producing end results for the respective applications in the IoT environment [15]. To solve the cloud challenge, many researchers discuss an “Embedded edge AI chip into the IoT devices”.

Most traditional AI computing tasks are performed remotely on centralized core devices or platforms, but this is not the best solution for IoT. Edge AI chips make it possible to embed AI computations into IoT devices. An Edge AI chip [215] is a chip that can perform or accelerate machine learning tasks on edge devices. At present, Google, NVIDIA, Intel, Qualcomm, and Huawei are all rapidly developing edge AI chip technologies, such as Google Coral Edge TPU [216]. Edge AI chips can also protect data privacy and security. The calculations do not need to send original data back to the cloud, which can greatly protect the security and privacy of data and reduce the possibility of data leakage and interception or abuse of personal/corporate data [22].

The next challenge is the huge number of IoT resources that collect different forms of data. Heterogeneity of IoT data accords when the data is collected from numerous sources, which forms a new challenge for ML models. In addition, using multiple parameters for module evaluation is one of the ML difficulties Researchers in [2] suggest considering an agreed standard set of metrics for the model's comparison for further improvements. In the same paper, data privacy for sensitive data is considered a weak point during ML training and testing, and researchers suggest randomly dividing the confidential values into several parts so that the opponent cannot deduce any information on the private value with a single element. Machine learning algorithms themselves have vulnerabilities that can be exploited by a malicious user. These kinds of attacks, named adversary machine attacks, aim to target the ML algorithms of a system to alter these initial functions. It is worth noting that it is essential to consider this threat when setting up security solutions integrating ML systems. Such networks usually have heterogeneous devices, dense users, frequent communication requests, and ubiquitous channel interference, which impose significant challenges on improving the Quality of Service (QoS) [21].

Anastasakis et al. [230] take the huge IoT network log as an advantage to enhance the security in IoT systems using federal learning (FL). FL is a distributed machine learning approach that aims to build and train global models based on training datasets that are distributed across different remote devices while avoiding data leakage. The data is never processed on central servers, decoupling the machine learning process from the data sources. In practice, FL solutions train an initial, generic machine-learning model in a central server, which is a baseline to start with, but their application faced a data

privacy challenge. Venkatasubramanian et al. [232] did a comprehensive survey about FL-based systems in IoT malware analysis, where their survey provides an overview of different approaches that integrate FL with IoT.

Performance and accuracy are not fixed parameters we want to achieve, but they should depend on the IoT system itself. For example, if the mission is life-critical (e.g., autonomous car, Military, Healthcare), we should provide a very high-performance model considering speed and accuracy. On the other hand, if we want to save power or battery, the calculation should be different. Bout et al. [2] say that there is no comparative study of battery consumption, either from an attacker or a victim's point of view that was carried out in a real context during most of the previous experiments. It is essential to improve learning methods in order to reduce the cost of the necessary resources and training time while increasing their performance. Due to the rapid data generation speed and the complexity of data sources of IoT, maintaining high-quality data in real-time is a challenging task [217]. Therefore, it remains an open research issue to investigate whether the available features in known benchmark datasets are sufficient to achieve high detection rates even in the presence of changing attack patterns or whether it will be necessary to add new features to maintain a high level of detection accuracy [12].

One of the main IoT challenges is its characteristics which cause limitations in different ways like power, capacity, computational processes, and more, those limitations encourage researchers to find lightweight technologies and techniques to deal with IoT devices, and one of those is in ML and DL solutions, IoT characteristics and limitations discussed more in section IV. Latif et al. [52] discussed a lightweight random neural network stack detection schema in the industrial environment IIOT, and they talked deeply about the challenges they found and some research future directions to control those challenges. One of the directions was about the generation of new security-related datasets, where high-quality data is an important thing for model performance evaluation. The other future direction was the improvement in existing ML schemes for low-quality and noisy datasets, and they said the improvements in existing proposed schemes and the development of new algorithms are required to deal with low-quality and noisy data. They also list other directions such as the implementation of learning schemes at the edge, fog domain security, and the last solution was in using Blockchain-Based secure ML schemes for IoT security.

Machine learning decisions about IoT detection are another challenge, most of the time ML can detect the intrusion but cannot decide how to deal with it. Abou El Houda et al. [225] produce Explainable Artificial Intelligence (XAI) which is a full system that can detect and take decisions about IoT intrusions using ML and DL techniques. They first build an ML/DL-based IDS using a deep neural network (DNN) to detect and predict IoT attacks in real-time. Then they develop multiple XAI models (i.e., RuleFit and SHapley Additive exPlanations, SHAP) on top of our DNN architecture to enable more trust, transparency, and explanation of the decisions made by their ML/DL-based IDS to cyber security experts.

Abdelmoumin et al. [148] found a solution for costly training and testing time and computational overhead, they solve it using a distributed intelligent IDS architecture that can help reduce the computational overhead and; hence, reduced latency. Shukla [218] proposed XML-based lightweight IDS for low-power IoT networks running 6LoWPAN. They used the IDS mechanism to detect wormhole attacks in IoT networks. The proposed IDS mechanism uses three ML techniques, i.e., K-means clustering (unsupervised learning), decision tree (supervised learning), and a hybrid technique combining the aforementioned techniques.

M. Raza et al. [42] present the role of IoT in healthcare and cover suitable IoT-driven solutions, especially in Covid-19. They also present challenges in the adoption of IoT in healthcare such as patient monitoring from home because of the lockdown. Patients need technologies in their homes to keep their health monitored and this was one of the challenges. The other challenge is energy and power usage, which is a main limitation in the IoT environments, and researchers cover this gap with suggestions for empowering IoT solutions. Authors listed other IoT challenges and limitations in healthcare during Covid-19 such as security, privacy, integration with other technology, data format, and legal and ethical issues.

Alrashdi et al. [125] focus on two main IoT challenges, but in the smart city environment. The first challenge was in zero-day attack detection, and the second one was finding the best AI method in cyberattack detection. In this research, the authors faced three challenges: Limited resources, heterogeneity, and a high false positive rate. They introduced an approach based on NIDS, called the AD-IoT system, to detect various loT attacks in a distributed fog layer instead of a cloud layer to solve those challenges.

Bagaa et al. [141] provided a list of promising technologies and designed a security framework to integrate them comprehensively. The research challenge they found was in defining standardized interfaces to ease the interactions among the envisioned framework modules, including common languages to specify the IoT security policies needed to react according to the AI-based decisions. Secondly, as the IoT landscape is continuously evolving, the AI system will need to be autonomously reconfigured to deal with additional emerging (and potentially unknown) IoT cyber-attacks, which do not follow previous network/systems signatures and patterns. Thirdly, another challenge deals with machine learning methods and algorithms that can be used by the reaction agent to dynamically plan the best countermeasure(s) to enforce according to different contexts.

Other challenges faced by Liang et al. [59] were, first, the training process is time and computationally expensive, and that traditional machine learning cannot handle dynamic systems, such as intrusion detection systems are dynamic systems, in which new training data

are continuously generated. Second, applying one well-trained machine learning model to multiple scenarios is a challenging issue. Third, machine learning is a black-box process, and backtracking through specific training steps is difficult, if not impossible. Authors found some future directions which may help in reducing those challenges. For example, using some online distributing learning will improve training and testing time and speed, especially in dynamic systems in which those distributed platforms can do training on the new slice for data only, instead of entering and retraining the whole dataset. They also suggest a solution for the problem of deploying ML in different CPS scenarios, where using data normalization, discretization, and sampling could an a appropriate solution. However, these solutions are not satisfactory in computation-limited devices in collaboration with time-sensitive CPS requirements. Therefore, mechanisms to reduce the size of collected and stored data in constrained CPS devices are another possible research direction [59].

Salau et al. [6] did a survey that provided a review of the ML techniques that have been applied to solve some of the challenges of wireless networks for IoT and CPS, and some open challenges in wireless networking for CPS and IoT systems that AI techniques can be used to fill the gap, and the future work needed to create robust AI systems for IoT/CPS.

Gümüşbaş et al. [12] discuss the criteria for a reliable benchmark dataset, which concerns the diversity of the traffic data, the diversity of the protocols, the volume of collected data, the diversity of the attacks considered, the inclusion of novel attack types, the inclusion of full payloads without anonymization, the presence or absence of informative features, the updatability, the consideration of realistic traffic, the extent of labeling, and the size of the feature set. Finally, any discussion of dataset reliability should consider the ability of a dataset to adapt to changes over time, by mimicking statistically normal traffic in accordance with upcoming needs, for example. Similarly, researchers in [219] propose the following 11 criteria for assessing the reliability of a dataset for intrusion detection:

1) Attack diversity.
2) Anonymity.
3) Available protocols.
4) Complete capture (with payloads).
5) Complete interaction.
6) Complete network configuration.
7) Complete traffic.
8) Feature set.
9) Heterogeneity (all network traffic and system logs).
10) Correct labelling.
11) Metadata (full documentation of data collection).

Code Analysis and Neural Learning are other trends that can also reveal that, with the network becoming more complex, the effort required for code analysis efforts decrease [29]. From the reviewed literature, a trend can view the network models applied for vulnerability detection as becoming increasingly complex and more expressive for better learning code semantic indicative of vulnerable code snippets. From recent studies using the CNN [220], [221] or LSTM [222], until the very recent studies adopting memory networks [223], the evolving network structure has shown the research effort that has been put into exploring the potential of neural networks for reasoning about the code semantics and rich patterns for facilitating vulnerability discovery. Researchers from the ML and NLP communities have been motivated to adopt state-of the-art tools/approaches for code analysis for vulnerability detection [224].

We cannot ignore IoT connectivity challenges that are encountered in the deployment of IoT devices [17]:

• The first one is providing unique IPs to billions of devices connected to the Internet. This challenge can be mitigated by incorporating 6LoWPAN which uses IPv6.

• The second challenge is developing low-power communication for transmitting data generated by sensors.

• The third challenge is implementing effective routing protocols that consider the limited memory of sensors and support the flexibility and mobility of smart objects.

Finally, Hussain et al. [15] listed at the end of their research a number of future directions and challenges to be solved such as:

1) DL–One Size Does Not Fit All: DL techniques are very much application-specific where a model trained for solving one problem might not be able to perform well for another problem in a similar domain.

2) Neural Networks Are Black Boxes: Deep neural networks act like a Blackbox, as we do not know how any DL model reaches a conclusion by manipulating the input data using the neurons at the intricately interconnected layers.

3) Longer Convergence Time: Most of the RL algorithms have longer convergence times, and it may make them unsuitable for real-time applications.

4) Butterfly Effect of ML and DL: The Butterfly effect is a phenomenon where a minute change in the input of a system creates chaos in the output. In this regard, ML and DL are also susceptible to this effect where a slight change in the input data to the learning system will create an enormous change in the output which is the learned model.

5) Challenges for DL in the Edge: IoT will leverage the advantages of edge computing which will increase the IoT applications and services space. However, due to the sheer amount of data generated by IoT devices, it will be hard to implement DL techniques in edge devices.

6) Over-Fitting Requirements and Hyper Parameters: Training offline from the fixed data logs (specified with external behavior policy) and learning from limited samples on the real system greatly affects the credibility of the decision-making of DL models.

7) Real-Time Response Requirements: Real-time mission-critical IoT applications such as autonomous vehicles, e-health, online banking, etc. perform continuous sensing and information gathering from their surroundings. Therefore, model updates, interferences from surrounding knowledge sources, and predictions are based on live-streaming data.

9 Conclusion

To improve security measures to recognize and respond to assaults, cyber security has grown to be a concern on a global scale, especially when technology has become an integral part of our life. IoT gained more popularity in the last 15 years and became one of the top components around us. All companies and individuals search for smart things to make life easier, and IoT helps them to reach this. However, when we think about security and privacy, they are in danger with this enormous growth. The standard security systems that were previously in use are no longer appropriate since they are ineffective at identifying new threats. In a variety of applications, ML and DL techniques are essential in new cybersecurity systems and become more known in this area. Our survey has shown that there is a fast expansion in using ML in securing devices, especially IoT which became a main part of learning, shopping, business, banking, and government.

The survey also compared different ML techniques for developing systems security against attacks in IoT networks in the last five years. It has many comparison tables which list 152 papers from IEEE Xplore in IoT security using ML techniques. We have offered a full review of the application of ML and DL for IoT systems. The most popular technologies are also listed, and many titles in ML, cybersecurity, IoT, Models evaluations, datasets, and more are thoroughly discussed. The current paper also attempted to cover all possible challenges facing the researchers who use ML in security IoT environments. Those challenges help us and other researchers solve problems to improve the security model.

Our paper centers on IoT environment limitations and characteristics which make it different from regular networks and need special care in power, capacity, and computational limitations. The survey ends with some of the most recent solutions and future directions from different research from 2018 to 2023.

However, the last section includes some open challenges to investigate and offer a useful road map for academic and industrial researchers working in the field of ML and DL for securing IoT and all its layers.

We recommend our paper to all researchers who are interested in solving one of the biggest challenges in security which were caused by the IoT environment limitations and characteristics which make it different from regular networks and need special care in power, capacity, and computational limitations. Our survey provided open challenges to investigate and offer useful directions for academic and industrial researchers working in the field of ML and DL for securing IoT and all its layers. Listing the challenges may help us and other researchers to directly find problems, try to solve them, and draw a roadmap for future work. This survey also helps compare several ML and DL techniques and algorithms to finally get the best performance depending on its accuracy, precision, and recall, among other results. We also recommend this paper because it provides a list of the last updated datasets used in IoT security in Table 7. The table provides different useful information for researchers such as dataset purposes, number of records, number of features, and a direct and updated accessible link.

In conclusion, we think that our study may be helpful in shedding light on how machine learning techniques are used to create cyberattacks and can aid readers who are interested in creating fresh defenses against more sophisticated and potent attacks on IoT networks. We have also provided comprehensive literature in this field and briefly outlined some of the major difficulties associated with applying machine learning techniques to cyber security. Future studies are recommended to pay attention to the aforementioned difficulties.

With a thorough discussion of the evaluation matrix, datasets, IoT characteristics, and other helpful material, this paper aims to give readers a roadmap for understanding the potential of ML and DL methods for IoT security and detection systems.

APPENDIX

Table 12: Comparison between ML modules for all detection systems for papers published in the last 5 years in IEEE journal.

							Results			
Serial #	Ref #	year	Detection type	Dataset	Domain	ML technique	Accuracy	Precision	Recall	F1-score
1	[33]	2021	Malware detection	Anubis: Analyzing Unknown Binaries,	Visualization	KNN RF NB	98.17 97.58 96.78	98.48 97.48 96.34	98.17 97.45 96.52	98.1 97.45 96.52
2	[94]	2021	Malware detection	IoTPoT	CPU architecture	RNN	98.71	-	-	-
3	[95]	2020	Malware detection	HPCs values collected by authors	Hardware-based malware	NN SVM	81.15 73.95	79.6 67.9	82.16 77.24	- -

						KNN LR NB DT	76.98 54.45 52.12 79.9	74.15 57.18 13.18 78.3	77.24 54.31 59.6 80.88	- - - -
4	[96]	2017	Malware detection	3,258 Android applications collected by authors	Android applications	NB J48 RF	99.19 98.79 99.59	-. - -	- - -	- - -
5	[97]	2020	Malware detection	IoTPOT TWISC Virus Share	API calls	CNN LSMT RF NB	98.56 96.99 90.25 89.65	99.37 98.57 87.96 80.47	96.87 91.36 84.17 89.67	98.55 96.96 90.25 90.97
6	[98]	2020	Malware detection	122,504 malware ELF files collected by authors	Antivirus software reports	RF KNN SVM	98.80 94.01 94.35	- - -	- - -	98.79 93.96 94.15
7	[99]	2022	Malware detection	SARD	NLP	BERT Model codeBETR	95 95	91 96	88 93	88 94
8	[100]	2022	Malware detection	DT Dase Malware	VM and cloud	DT		96.8	97.8	97.5
9	[101]	2019	Malware detection	RISS	Clouds and edge computing	DL ML	99.7 98.25	- -	- -	- -
10	[102]	2019	Malware detection	VirusShare Malgenome Contagio Minidump	Android	NB DT RF SVM	90.9 86.4 87.5 92.5	- - - -	92.3 93 78.9 80	- - - -
11	[103]	2019	Malware detection	1790 application collected by authers	Autonomous driving	DNN DT SVM KNN ADA GBDT GNB RF ET xgboost	93 89 93.5 92 89.3 90.2 81 91 91 95	- - - - - - - - - -	- - - - - - - - - -	
12	[104]	2019	Malware detection		IoT (ISP)	RF KNN GNB	88.8 94.44 77.78	86 92 75	100 100 100	92 96 86
13	[105]	2022	Malware detection	IoT-23 VARIoT	Network traffic	Task 1 model Task 2 model Multitask1 Multitask2	92.63 88.45 95.38 89.10	92.05 91.91 95.38 95.62	96.14 89.89 99.88 94.54	96.90 92.66 99.16 95.74
14	[106]	2021	Malware detection	MQTTset	MQTT	DT RF LR K-means GB SVM	49.97 50.87 98.45 49.86 49.95 87.53	- - - - - -	- - - - - -	33.32 35.25 98.45 33.27 33.31 87.44
15	[107]	2020	Malware detection	1000 Binvis images collected by authors	Network traffic	CNN	94.50	95.78	94.02	94.90
16	[108]	2022	Malware detection	CICMalDroid 2020 Drebin	Android	CNN	97.86 98.43	98.76 92.92	98.46 91.01	98.61 91.96
17	[109]	2021	DDoS Detection	Collected by authors ISCX from UNB CAIDA DDoS 2007	IoT network traffic	SVM KNN RF DT NN	93.4 92.5 92.3 99.9 97.2	- - - - -	- - - - -	- - - - -
18	[110]	2020	DDoS Detection	CICIDS2017 NSL-KDD KDDCup99 ISCX from UNB	loT networks	CNN	85.55 91.50 95.55 97.00 97.27	- - - - -	- - - - -	- - - - -
19	[111]	2018	DDoS Detection	Data collected by authors	loT networks	SVM KNN	95 90	- -	- -	- -
20	[112]	2022	DDoS Detection	C1DDoS C2DDoS C3DDoS C4DDoS	IoT devices	LMT/author MLP KNN RT Bagging AdaBoostM1 DL RNN SGD LSTM	99.99 87.47 96.26 98.64 97.63 97.41 91.10 50.46 90.50 99.06	99.99 90.2 96.3 98.7 97.7 97.4 91.1 58.4 90.6 99.1	99.9 87.5 96.3 98.6 97.6 97.4 91.1 50.5 90.5 99.1	99.9 87.4 96.3 98.6 97.6 97.4 91.1 35.2 90.5 99.1
21	[113]	2018	DDoS	ISCX from UNB	SDN	SVM	93.4	-	-	-

			Detection			KNN RF	92.5 92.3	- -	- -	- -
22	[114]	2019	DDoS Detection	Bot-IoT	IoT devices	ANN	99	100	99	99
23	[115]	2019	DDoS Detection	CICIDS2017	loT networks	CNN MLP LSTM SVM Bayes RF	95.14 86.34 96.24 95.5 95.19 94.64	98.14 88.47 98.44 97.72 92.56 90.18	- - - - - -	- - - - - -
24	[116]	2021	DDoS Detection	Boun DDoS	IoT network traffic	RF	97.9	100	2	98.9
25	[117]	2021	DDoS Detection	ISCXIDS 2012 CICIDS 2017 CSE-CIC-IDS 2018 CICDDoS 2019	Cloud	NB /CICDDoS 2019 LR /CICDDoS 2019 DT /CICDDoS 2019 RF/CICDDoS 2019 KNN/CICDDoS 2019 SVM/CICDDoS 2019 MLP/CICDDoS 2019	98.33 99.79 99.97 99.97 99.92 99.79 99.93	99.72 99.92 99.99 99.99 99.96 99.90 99.96	98.44 99.85 99.97 99.98 99.95 99.87 99.95	99.72 99.88 99.98 99.98 99.95 99.88 99.96
26	[118]	2021	DDoS Detection	IoT-23	IoT network traffic	DT RF	99.5 99.0	92.0 95.0	94.0 93.0	92.0 93.0
27	[119]	2018	DDoS Detection	Collected by authors	IoT network traffic	KNN LSVM DT RF NN	99.9 99.1 99.9 99.9 99.9	99.8 99.2 99.6 99.9 98.3	99.3 87.0 99.3 99.8 98.9	99.5 92.7 99.4 99.8 98.6
28	[120]	2020	DDoS Detection	Data Collected by authors combined it with CICDDoS2019	IoT network traffic	ID3 RF NB LR Author Model	- - - - -	78.0 77.0 41.0 25.0 99.47	65.0 56.0 11.0 02.0 99.31	69.0 62.0 05.0 04.0 99.35
29	[121]	2020	DDoS Detection	Collected by authors from 3 IoT devices.	Wireless sensor network (WSN)	SVM NN J-48 RF	- - - -	99.5 99.6 99.6 99.7	99.4 99.6 99.5 99.7	99.4 99.6 99.5 99.7
30	[122]	2021	DDoS Detection	CIC DoS 2017 CIC IDS 2017	Software-Defined Networking (SDN)	J48 RT REP Tree RF SVM MLP FFCNN/ Authors	90.68 91.76 90.37 94.41 93.1 95.01 99.41	65.22 72.83 64.17 78.33 92.0 95.46 97.48	52.8 55.65 50.44 81.86 93.0 94.51 99.54	58.36 63.09 56.48 80.05 93.0 94.98 98.50
31	[123]	2019	Anomaly Detection	MAWILab	Apache Spark	DT LR RF MLP NB	84.70 95.83 95.05 83.07 18.20	- - - - -	- - - - -	- - - - -
32	[124]	2020	Anomaly Detection	Collected by authors	Surveillance systems	NB MP ICO DT RF	95.64 96.56 96.79 95.41 96.56	- - - - -	- - - - -	- - - - -
33	[125]	2019	Anomaly Detection	UNSW-NB15	Smart City	RF	99.34	79.0	97.0	86.0
34	[126]	2021	Anomaly Detection	N-BaIoT	IoT in the internet	Auto-encoder ANN LR	- 96.4 99.98	99.30 93.9 99.9	99.99 95.1 99.96	- 99.13 99.92
35	[127]	2021	Anomaly Detection	CIC-IDS-2017	IoT Network	KNN RF ID3 Adaboost MLP NB Q discr.analysis PHICAD	- - - - - - - -	96 98 98 77 77 88 97 99	96 97 98 84 83 04 88 84	96 97 98 77 76 04 92 91
36	[128]	2021	Anomaly Detection	Motion sensors data collected by authors	Discover Sensor Tampering	AD-ML/ authors	91.62	94	75	83
37	[129]	2020	Anomaly Detection	Collected by authors	Domain Name System (DNS) traffic data	ARBA/ authors ARBA-SVM ARBA-LR ARBA-NC BotDAD DomainObserver	99.7 99.6 99.5 99.5 99.5 99.1	97.1 97.1 93.0 92.9 94.2 87.1	93.2 90.5 90.5 89.1 87.8 82.4	95.1 93.6 91.8 90.9 90.9 84.7
38	[130]	2019	Anomaly	Collected by authors	Smart	Transfer Learning	-	90.68	89.82	90.24

			Detection		Manufacturing	VAR CNN KNN	70-85 90 90	76.52 71.38 73.31	85.64 82.51 81.71	80.82 76.54 77.28
39	[131]	2021	Anomaly Detection	Collected by authors	HTTP anomaly detection	One-class HYBRID Only OC-SVM Only SVDD	98.31 97.2 92.63	95.2 89.44 74.72	97.0 98.67 99.48	96.11 93.82 85.34
40	[132]	2022	Anomaly Detection	UNSW-NB15	IoT Network	K-Means Mean Shift LOF RF LR CatBoost LOF with RF LOF with LR LOF with CatBoost Authors model	71 53 75 69 89 98 99 93 99 99	60 50 46 43 82 93 98 89 99 99	65 50 47 43 76 97 97 84 97 97	60 45 46 43 78 98 97 86 98 98
41	[133]	2019	Anomaly Detection	mobile big data (MBD)/ authors	Mobile Wireless Networks	NB Bayesian SVM RF	89.4 86.5 92.9 91.3	22.6 0.18 97.6 99.9	99.1 31.5 99.6 99.3	36.8 0.4 98.6 99.6
42	[134]	2019	Anomaly Detection	Collected by authors	Cloud computing	SVM RF MLP	- - -	71 95 100	- - -	70 94 100
43	[135]	2021	Anomaly Detection	CSIC + Authors Dataset	Hypertext Transfer Protocol (HTTP)	One-Class HYBRID Only OC-SVM Only SVDD	98.31 97.2 92.63	95.2 89.44 74.72	97.0 98.67 99.48	96.11 93.82 85.34
44	[136]	2022	Anomaly Detection	Modbus based network	IoT Network	Federal Leaning (FL)	98.67	96.25	96.72	96.48
45	[137]	2020	Anomaly Detection	DS2OS traffic traces	IoT Network	LR ANN	99 99	99 99	91 96	94 94
46	[138]	2018	Intrusion Detection	NSL-KDD	wireless packet traffic flow	K-mean SVM SVM+ K-mean	71.45 74.45 98.34	- - -	- - -	- - -
47	[139]	2022	Intrusion Detection	NSL-KDD CIC-IDS2018 TON IoT	IoT Network	CNN	99	-	-	-
48	[140]	2021	Intrusion Detection	TON_IoT	IoT Network	DT RF AdaBoost GB XGBoost KNN SVM	99.19 99.23 95.37 97.74 99.14 83.41 86.02	98.57 98.62 94.85 96.05 98.47 79.64 76.13	99.12 99.19 91.72 97.55 99.09 70.53 87.40	98.85 98.90 93.26 96.79 98.78 74.81 81.37
49	[141]	2020	Anomaly Detection	DARPA KDD99 DEFCON	SDN and NFV	SVM	99.9			
50	[142]	2022	Intrusion Detection	Bot-IoT	IoT Network	DT LR NB Stack	79.06 92.05 59.93 100	35.78 52.62 47.21 100	28.64 50.88 22.04 99.91	28.46 51.17 21.46 99.96
51	[52]	2020	Attack Detection	DS2OS	Industrial IoT system	ANN RaNN	97.97 98.33	97.27 98.05	97.97 98.06	97.62 98.12
52	[143]	2021	Intrusion Detection	UNSW-NB 15	Fog/Edge computing	XGB ADA_Boost RandomForest IG RandomForest Gini GBC DT KNN SVM Gaussian NB	95.54 92.77 92.63 92.59 92.09 91.44 90.12 89.11 50.46	- - - - - - - - -	- - - - - - - - -	- - - - - - - - -
53	[144]	2020	Intrusion Detection	OneM2M	OneM2M	J48 DL	92.32 82	92.95 83.82	93.80 86.62	- -
54	[145]	2020	Intrusion Detection	IoT/IIoT TON-IoT	OS logs and Network traffic of IoT network	LR LDA KNN RF CART NB SVM LSTM	61 92 72 71 77 54 60 98	38 46 71 69 77 59 37 64	62 63 73 72 77 51 61 68	47 51 70 67 75 52 46 63
55	[146]	2021	Intrusion Detection	NSL-KDD	IoT Network	SVM + linear kernel SVM+polymoial kernel	96.5 98.8 99.0	- - -	- - -	- - -

						SVM + RBF kernel SVM + Sigmoid kernel	97.8	-	-	-
56	[147]	2020	Intrusion Detection	NSL-KDD	IoT Network	KNN	99.78	-	-	99.72
57	[148]	2022	Intrusion Detection	IOT_BOTNET	industrial IoT network traffic	PCA-SVM PCA-NN PCA-SVM-NN	21.9 8.1 6.3	94.6 57.1 46.6	17.7 8.0 6.0	29.8 14.0 10.7
58	[149]	2020	Intrusion Detection	IoT Network Intrusion Dataset	IoT Network	LR SVM KNN RF XGBoost	86 79 96 100 96	- - - - -	75 78 96 100 96	80 77 96 100 96
59	[150]	2020	Intrusion Detection	Collected by authors	Blockchain Network	SVM	99	99.98	100	99.98
60	[151]	2022	Intrusion Detection	Collected by authors	IoT Network	LR RF NB SVM	93.52 93.54 93.53 93.53	- - - -	- - - -	- - - -
61	[152]	2021	Intrusion Detection	KDDCUP99 N-BaIot	IoT applications	J48 RF RT REPTree NB	99.92 99.96 99.95 99.93 77.41	- - - - -	- - - - -	- - - - -
62	[153]	2022	Intrusion Detection	CIC-IDS dataset 2017	Network Flow	RF DT ET KNN	99.61 99.53 99.59 99.19	100 80 100 100	100 100 100 100	100 88.88 100 100
63	[154]	2022	Intrusion Detection	UNSW NB15 BoT-IoT	IoT Network	Bi-LSTM LSTM	96.60 96.41	96 97	96 96	96 96
64	[155]	2022	Intrusion Detection	TON-IoT	IoTProtect	RF LR DT Gaussian NB	99.99 65.01 99.99 37.02	- - - -	- - - -	- - - -
65	[156]	2021	DoS detection	NSL-KDD	Wireless Environment	MLP	-	93	71	81
66	[157]	2021	DoS Detection	UNSW-NB15 ISCX	SDN	SVM KNN MLP	- - -	95 95 95	95 97 97	95 95 95
67	[158]	2019	Intrusion Detection	NSL-KDD	IoT Network	DNN + softsign Activation function	98	98	98	98
68	[159]	2021	Intrusion Detection	CICIDS2017	IoT Network	CNN /DDos attack CNN/ infiltration CNN/ Web attack CNN/ Brute force attack	98 97 98 97	96 98 95 98	98 97 97 98	98 97 97 97
69	[160]	2021	Intrusion Detection	BoT-IoT	SDN	RF CNN	90.78 99.95	- -	- -	- -
70	[161]	2019	Intrusion Detection	UNSW-15	IoT authentication	ANN	84	-	-	-
71	[162]	2021	Intrusion Detection	UNSW-NB 15	IoT Network	PLA LR NN DT Voting BoA AdaBoost RF	16.19 17.23 32.81 28.97 19.36 18.092 7.58 68.50	12.31 10.09 12.51 21.30 10.94 09.71 15.29 29.55	10.88 11.11 11.06 12.20 11.63 11.01 15.10 20.94	05.94 06.68 08.56 10.01 06.65 06.66 10.59 19.38
72	[163]	2022	Intrusion Detection	Comp1, Comp2 Comp3, Comp4 Comp5, AIO	IoT devices	CNN/ Comp3 CNN/ AIO	96.73 96.77	90.20 90.31	90.20 90.31	90.20 90.31
73	[164]	2019	Intrusion Detection	NSL-KDD	Edge Computing	J48/ DoS SVM/ DoS NB/ DoS Logistic/ DoS RF/ DoS CNN/ DoS	97.4 97.5 75.7 97.1 96.9 99.2	- - - - - -	- - - - - -	- - - - - -
74	[165]	2021	Intrusion Detection	BOT-IOT	IoT Network	SVM SVM + Chi Square SVM + ExtraTrees SVM + PCA SVM + FA	88.37 98.66 99.29 62.52 99.38	100 99.99 100 100 100	83.37 98.66 99.29 76.92 99.30	98.84 99.32 99.64 62.59 99.78

75	[166]	2021	Attack detection	NSL-KDD	IoT Network	LR	-	-	-	96
						GBC	-	-	-	97.44
						KNN	-	-	-	94.55
						DT	-	-	-	95.39
						XGBoost	-	-	-	96.99
						GSOM	-	-	-	96.47
				KDD99		LR	-	-	-	88.72
						GBC	-	-	-	99.69
						KNN	-	-	-	96.66
						DT	-	-	-	98.02
						XGBoost	-	-	-	99.03
						GSOM	-	-	-	93.60
				CICID		LR	-	-	-	87.36
						GBC	-	-	-	99.19
						KNN	-	-	-	99.48
						DT	-	-	-	98.33
						XGBoost	-	-	-	97.28
						GSOM	-	-	-	99.99
				Bot-IoT		LR	-	-	-	90.15
						GBC	-	-	-	91.64
						KNN	-	-	-	93.56
						DT	-	-	-	95.49
						XGBoost	-	-	-	98.12
						GSOM	-	-	-	9445
76	[167]	2021	Spam Detection	REFIT smart home dataset.	IoT framework	BGLM	79.81	65	100	-
						BLM	83.22	54.1	100	-
						Xgboost	84.35	56.7	100	-
						GLM	88.9	59.8	100	-
77	[168]	2019	Attack detection	IoT Botnet Attacks	IoT critical infrastructures	DT	99.13	-	-	-
						SVM	99.31	-	-	-
						MP	90.81	-	-	-
78	[169]	2019	Attack detection	KDD	Traffic Signal System	K-Mea	75	-	-	-
						K-Med	75	-	-	-
						RF	100	-	-	-
						LAD	100	-	-	-
79	[170]	2021	Attack detection	CVE extracted from the NVD database	Social Media Monitoring	RF	92.71	71.75	91.57	77.65
						DT	89.97	67.20	87.84	72.24
						NB	60.19	55.31	73.72	47.12
						LR	71.42	57.12	78.36	54.24
						SVM	73.90	57.98	80.63	56.25
						KNN	58.09	54.32	69.38	45.25
80	[171]	2020	Attack detection	KDD	IoT Industry	KNC	88.56	-	-	-
						K-mean	82.78	-	-	-
						DT	87.91	-	-	-
						MLP	87.91	-	-	-
						LR	89.52	-	-	-
						SVM	88.32	-	-	-
						CDL	97	-	-	-
						Co-DL2	97.52	-	-	-
						Co-DL3	97.54	-	-	-
				NSLKDD		KNC	94.31	-	-	-
						K-mean	87.05	-	-	-
						DT	93.78	-	-	-
						MLP	90.16	-	-	-
						LR	92.52	-	-	-
						SVM	93.38	-	-	-
						CDL	90.86	-	-	-
						Co-DL2	93.99	-	-	-
						Co-DL3	93.37	-	-	-
				UNSW-NB15		KNC	96.85	-	-	-
						K-mean	86.19	-	-	-
						DT	97.01	-	-	-
						MLP	96.77	-	-	-
						LR	96.2	-	-	-
						SVM	96.74	-	-	-
						CDL	95.67	-	-	-
						Co-DL2	95.6	-	-	-
						Co-DL3	95.67	-	-	-
81	[172]	2020	Attack detection	UNSW NB15	Smart City	ANN	85.16	84	85	84
82	[173]	2018	Attack detection	NSL-KDD	IoT environment	DFEL.GBT	98.54	98.54	98.53	-
						DFEL.KNN	98.82	98.82	98.82	-
						DFEL.DT	98.77	98.77	98.77	-
						DFEL.LG	98.85	98.85	98.85	-

						DFEL.GNB	98.80	98.79	98.79	-
						DFEL.SVM	98.86	98.86	98.87	-
				UNSW-NB15		DFEL.GBT	91.22	90.38	90.69	-
						DFEL.KNN	91.90	91.25	91.21	-
						DFEL.DT	92.29	91.24	92.52	-
						DFEL.LG	92.35	91.50	92.11	-
						DFEL.GNB	92.52	91.45	92.85	-
						DFEL.SVM	92.32	91.41	92.20	-
83	[174]	2022	Attack detection	IoT-MQTT	Message Queuing Telemetry Transport (MQTT)	ANN	98.23	99.58	99.58	98.34
						NB	96.07	99.78	99.78	96.73
						XGB	99.60	99.55	99.55	99.48
						DT	97.74	97.73	97.73	97.16
						KNN	95.86	96.74	96.74	94.73
84	[175]	2022	Attack detection & classification	IOT-23	IoT Traffic	NB	37.60	-	-	25.44
						DT	95.62	-	-	75.92
						BG	95.46	-	-	76.98
						RF	95.49	-	-	77.28
						MIMETC(NET+HDR)	99.93	-	-	91.70
85	[176]	2021	Attack detection	NSLKDD	Medical Smart Environment	PSO-RF	99.76	99.75	96.45	-
						PSO-DT	99.58	99.59	96.27	-
						PSO-KNN	98.90	98.89	92.33	-
						PSO-RC	97.61	97.60	91.06	-
86	[177]	2021	Attack detection	N-BaIoT	IoT system	K-mean/ Dataset1	99.2	-	-	99.0
						K-mean/ Dataset2	97.7	-	-	97.6
						K-mean/ Dataset3	96.7	-	-	96.6
87	[178]	2021	Attack detection	IoT-23	IoT Network	KNN	-	-	-	60.19
						RF	-	-	-	55.18
						AAE+KNN	-	-	-	97.43
						BiGAN+KNN	-	-	-	97.41
88	[179]	2021	Botnet Detection	N-BaIoT	IoT device	KNN	98.64	97.89	-	-
						DT	100	100	-	-
						RF	99.99	99.99	-	-
						AD	93.9	90.13	-	-
						GB	96.52	94.88	-	-
89	[180]	2021	Botnet Detection	Created by authors in [176]	IoT device	K-mean + DT	99.13	-	-	-
90	[181]	2019	Botnet Detection	Mirai-alike	DNS	RF	-	99	98	98
						KNN	-	90	78	80
91	[182]	2022	Botnet Detection	Collected from IoT devices in University of New South Wales's (UNSW)	IOT Network Traffic	LR	94.60	-	-	-
						RF	96.18	-	-	-
						KNN	94.64	-	-	-
						SVM	95.04	-	-	-
						XGB	96.03	-	-	-
						DNN	82.64	-	-	-
92	[183]	2019	Botnet Detection	N-BaIoT	IOT Network Flow	LR	11.1	10.3	11.1	2.3
						NB	7.2	10.7	7.2	8.6
						KNN	93.9	95.2	93.9	93.7
						AB	98.5	98.6	98.5	98.5
						AB	38.3	28.3	38.3	26.2
						RF	93.6	94.9	93.6	93.4
						RSVM	95.8	94.1	94.8	95.1
						LSVM	96.8	95.7	95.1	95.9
						DNN	100	100	100	100
				BoT-IoT		LR	52.6	33.4	26.4	19.8
						NB	52.9	47.1	53.4	24.4
						KNN	99.8	99.6	99.7	99.6
						DT	99.9	98.2	98.2	98.2
						AB	12.2	61.2	35.3	20.5
						RF	99.1	99.5	97.8	98.6
93	[184]	2022	Botnet Detection	N-BaToT	IoT environment	NB	-	14	14	14
						RF	-	91	91	91
						ANN	-	90	90	90
94	[185]	2022	Botnet Detection	IoT-23	IoT Network	NB	-	85	30	21
						SVM	-	60	69	57
						DT	-	77	73	65
95	[186]	2020	Botnet Detection	CICIDS2017	IoT Network	NB	73.71	68.33	99.93	81.16
						KNN	99.59	99.58	99.69	99.64
						RF	100	100	100	100
						LR	100	100	100	100
				CTU-13		NB	71.72	92.89	36.11	52.01
						KNN	97.51	97.67	96.44	97.05
						RF	100	100	100	100
						LR	100	100	100	100
				IoT-23		NB	99.92	99.92	99.92	99.92

						KNN RF LR	99.94 100 99.91	99.94 100 99.91	99.94 100 99.91	99.94 100 99.91
96	[187]	2018	Botnet Detection	Balanced and unbalanced dataset created by authors	IoT Network	SVM /unbalanced DS SVM/ balanced DS	93.15 83.37	96.27 76.86	- -	- -
97	[188]	2022	Botnet Detection	IoT-23	5G Networks	Authors model with WFS feature selection	99.99	100	-	100
98	[189]	2020	DDOS Detection	Collected by authors	IoT networks with WSNs	SVM/Device ID 3 NN/Device ID 3 J-48/Device ID 3 RF/Device ID 3	- - - -	99.5 99.6 99.6 99.7	99.4 99.6 99.5 99.7	99.4 99.6 99.5 99.7
99	[190]	2022	Botnet Detection	MedBIoT	IoT Network	DT LR MLP Ensemble ELM/ authors	81.6 69.9 83.8 90.2 95.4	82.7 68.3 87.4 89.8 94.2	82.6 96.9 87.5 88.9 94.1	82.6 69.0 87.4 89.5 91.1
100	[191]	2018	Botnet Detection	Collected by authors	Linux IoT system	(DGCNN)/size1100	91.6	91.7	96.3	94.0
101	[192]	2020	Botnet Detection	N_BaIoT	IOT Network Traffic	DT /Baby monitor ETC /Baby monitor RF/Baby monitor SVM/Baby monitor	- - - -	83.98 89.46 94.34 91.75	83.44 89.22 94.23 90.71	83.71 89.34 94.29 91.23
102	[193]	2019	Malware Detection	Collected by authors	IoT software vulnerabilities	RF KNN GNB	88.8 94.44 77.78	86 92 75	100 100 100	92 96 86
103	[194]	2020	Botnet Detection	ISOT HTTP Botnet CTU-13 CICDDoS2019	IOT Network Traffic	ANTE/ authors SVM NB MLP	99.23 93 74 94	100 100 40 94	96 62 83 71	98 77 54 81
104	[195]	2018	Botnet Detection	Collected by authors	IOT Network Traffic	DT KNN	98.97 94.97	- -	- -	- -
105	[196]	2022	Botnet Detection	Collected by authors [177]	IOT Network Traffic	BC DT RF	99.9 99.82 99.7	99.8 99.62 99.49	99.9 100 99.87	99.8 99.81 99.68
106	[197]	2022	Botnet Detection	Collected by authors	IoT Edge Devices	DT/ 115 features DT/ 9 features LR NB RF AdaBoost XGBoost SGD ANN	100 100 99.94 96.76 100 100 100 99.90 99.99	100 100 99.98 96.98 100 100 100 99.98 99.99	100 100 99.95 99.95 100 100 100 99.91 99.99	100 100 99.97 9827 100 100 100 99.94 99.99
107	[198]	2020	Botnet Detection	Bot-IoT-2018	IoT Environments	DT SVM Authors model	99.82 88.37 99.99	53 100 99	91 88 100	56 94 100
108	[199]	2021	Botnet Detection	Bot-IoT	IOT Network Traffic	C4.5/ authors- normal NB- normal RF- normal SVM- normal	99.99 99.97 99.98 99.82	93 32.33 95.52 100	- - - -	- - - -
109	[200]	2021	Botnet Detection	Bot-IoT	IOT Network Traffic	CNN LSTM	94.48 86.24	- -	- -	- -
110	[201]	2021	Botnet Detection	Bot-IoT CTU-13	IOT Network Flow	DT/5s time window RF/5s time window XGB/5s time window	92.23 93.11 95.42	91.26 93.02 95.29	93.45 95.48 97.11	92.34 94.23 96.19
111	[202]	2020	Botnet Detection	Bot-IoT UNSW UNSW-NB15	IOT Network	DT NB KNN SVM	99.89 96.90 98.80 83	100 95 99 85	100 99 100 84	100 97 99 85
112	[203]	2021	Botnet Detection	UCI ML Repository	IoT Environments	GNB ANN SVM KNN CNN	75.98 88.70 97.14 97.48 97.98	- - - - -	- - - - -	- - - - -
113	[204]	2021	Anomaly detection	N-BaIoT	IOT Network	Auto-encoder ANN LR CNN RNN LSTM	- 96.4 99.98 - - -	99.30 93.9 99.9 - - -	99.99 95.1 99.96 - - -	- 99.13 99.92 91 41 62
114	[205]	2021	Intrusion Detection	UNSW-NB15	IOT Network traffic	RF IG RF Gini DT	92.63 92.61 91.44	- - -	- - -	- - -

						SVM LR GNB	89.19 89.05 50.46	- - -	- - -	- - -
115	[206]	2021	Botnet Detection	IoT-Botnet	Industrial IOT environment	SVM KNN RF	99.8 98.9 99.3	- - -	- - -	- - -
116	[207]	2021	Botnet Detection	PSI-graph	IOT environment	DT KNN SVM RF	85.22 92.77 95.13 94.45	88.5 92.72 96.96 94.94	87.35 95.78 95.07 96.11	87.92 94.23 96.01 95.52
117	[231]	2022	Botnet Detection	Bot-IoT	IOT Network	RF NB DT GB	98 70 91 99	96 70 91 99	96 89 96 99	98 89 96 99
118	[234]	2022	Attack detection	UNSW-NB15	Real IOT environment	KNN GNB SVM RF	96.67 78.07 97 99.84	97.51 85.53 96.23 99.74	96.41 72.43 98.41 99.97	96.96 78.43 97.30 99.85
119	[235]	2022	Botnet Detection	Ton-IoT	IOT environment	LR GNB RF DT KNN XGB	65 7 98.3 98.2 98.1 98.4	51 68 98.4 98.2 98.2 98.5	65 7 98.3 98.2 98.1 98.4	55 47 98.3 98.2 98.2 98.4

References

[1] K. Shaukat, S. Luo, V. Varadharajan, I. A. Hameed and M. Xu, "A Survey on Machine Learning Techniques for Cyber Security in the Last Decade," in IEEE Access, vol. 8, pp. 222310-222354, 2020, doi: 10.1109/ACCESS.2020.3041951.

[2] E. Bout, V. Loscri and A. Gallais, "How Machine Learning Changes the Nature of Cyberattacks on IoT Networks: A Survey," in IEEE Communications Surveys & Tutorials, vol. 24, no. 1, pp. 248-279, Firstquarter 2022, doi: 10.1109/COMST.2021.3127267.

[3] Liang, X., & Kim, Y. (2021, January). A survey on security attacks and solutions in the IoT network. In 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC) (pp. 0853-0859). IEEE.

[4] Sengupta, N. (2019, May). Designing security system for IoT. In 2019 IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing,(HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS) (pp. 195-199). IEEE.

[5] T. T. Nguyen and V. J. Reddi, "Deep Reinforcement Learning for Cyber Security," in IEEE Transactions on Neural Networks and Learning Systems, doi: 10.1109/TNNLS.2021.3121870.

[6] B. A. Salau, A. Rawal and D. B. Rawat, "Recent Advances in Artificial Intelligence for Wireless Internet of Things and Cyber–Physical Systems: A Comprehensive Survey," in IEEE Internet of Things Journal, vol. 9, no. 15, pp. 12916-12930, 1 Aug.1, 2022, doi: 10.1109/JIOT.2022.3170449.

[7] R. Li, Q. Li, J. Zhou and Y. Jiang, "ADRIoT: An Edge-Assisted Anomaly Detection Framework Against IoT-Based Network Attacks," in IEEE Internet of Things Journal, vol. 9, no. 13, pp. 10576-10587, 1 July1, 2022, doi: 10.1109/JIOT.2021.3122148.

[8] L. Xiao, X. Wan, X. Lu, Y. Zhang and D. Wu, "IoT Security Techniques Based on Machine Learning: How Do IoT Devices Use AI to Enhance Security?," in IEEE Signal Processing Magazine, vol. 35, no. 5, pp. 41-49, Sept. 2018, doi: 10.1109/MSP.2018.2825478.

[9] O. Abdel Wahab, "Intrusion Detection in the IoT Under Data and Concept Drifts: Online Deep Learning Approach," in IEEE Internet of Things Journal, vol. 9, no. 20, pp. 19706-19716, 15 Oct.15, 2022, doi: 10.1109/JIOT.2022.3167005.

[10] S. M. Kasongo, "An Advanced Intrusion Detection System for IIoT Based on GA and Tree Based Algorithms," in IEEE Access, vol. 9, pp. 113199-113212, 2021, doi: 10.1109/ACCESS.2021.3104113.

[11] A. Sarwar, S. Hasan, W. U. Khan, S. Ahmed and S. N. K. Marwat, "Design of an Advance Intrusion Detection System for IoT Networks," 2022 2nd International Conference on Artificial Intelligence (ICAI), 2022, pp. 46-51, doi: 10.1109/ICAI55435.2022.9773747.

[12] D. Gümüşbaş, T. Yıldırım, A. Genovese and F. Scotti, "A Comprehensive Survey of Databases and Deep Learning Methods for Cybersecurity and Intrusion Detection Systems," in IEEE Systems Journal, vol. 15, no. 2, pp. 1717-1731, June 2021, doi: 10.1109/JSYST.2020.2992966.

[13] J. Hunter, B. Huber and F. Kandah, "Towards feasibility of Deep-Learning based Intrusion Detection System for IoT Embedded Devices," 2022 IEEE 19th Annual Consumer Communications & Networking Conference (CCNC), 2022, pp. 947-948, doi: 10.1109/CCNC49033.2022.9700706.

[14] R. Zhao et al., "A Novel Intrusion Detection Method Based on Lightweight Neural Network for Internet of Things," in IEEE Internet of Things Journal, vol. 9, no. 12, pp. 9960-9972, 15 June15, 2022, doi: 10.1109/JIOT.2021.3119055.

[15] F. Hussain, R. Hussain, S. A. Hassan and E. Hossain, "Machine Learning in IoT Security: Current Solutions and Future Challenges," in IEEE Communications Surveys & Tutorials, vol. 22, no. 3, pp. 1686-1721, thirdquarter 2020, doi: 10.1109/COMST.2020.2986444.

[16] E. Rodríguez, B. Otero, N. Gutiérrez and R. Canal, "A Survey of Deep Learning Techniques for Cybersecurity in Mobile Networks," in IEEE Communications Surveys & Tutorials, vol. 23, no. 3, pp. 1920-1955, thirdquarter 2021, doi: 10.1109/COMST.2021.3086296.

[17] M. A. Al-Garadi, A. Mohamed, A. K. Al-Ali, X. Du, I. Ali and M. Guizani, "A Survey of Machine and Deep Learning Methods for Internet of Things (IoT) Security," in IEEE Communications Surveys & Tutorials, vol. 22, no. 3, pp. 1646-1685, thirdquarter 2020, doi: 10.1109/COMST.2020.2988293.

[18] A. Toshniwal, K. Mahesh and R. Jayashree, "Overview of Anomaly Detection techniques in Machine Learning," 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2020, pp. 808-815, doi: 10.1109/I-SMAC49090.2020.9243329.

[19] A. Jamalipour and S. Murali, "A Taxonomy of Machine-Learning-Based Intrusion Detection Systems for the Internet of Things: A Survey," in IEEE Internet of Things Journal, vol. 9, no. 12, pp. 9444-9466, 15 June15, 2022, doi: 10.1109/JIOT.2021.3126811.

[20] L. Basheer and P. Ranjana, "A Comparative Study of Various Intrusion Detections In Smart Cities Using Machine Learning," 2022 International Conference on IoT and Blockchain Technology (ICIBT), 2022, pp. 1-6, doi: 10.1109/ICIBT52874.2022.9807724.

[21] T. Li et al., "Applications of Multi-Agent Reinforcement Learning in Future Internet: A Comprehensive Survey," in IEEE Communications Surveys & Tutorials, vol. 24, no. 2, pp. 1240-1279, Secondquarter 2022, doi: 10.1109/COMST.2022.3160697.

[22] H. Wu, H. Han, X. Wang and S. Sun, "Research on Artificial Intelligence Enhancing Internet of Things Security: A Survey," in IEEE Access, vol. 8, pp. 153826-153848, 2020, doi: 10.1109/ACCESS.2020.3018170.

[23] S. Zaman et al., "Security Threats and Artificial Intelligence Based Countermeasures for Internet of Things Networks: A Comprehensive Survey," in IEEE Access, vol. 9, pp. 94668-94690, 2021, doi: 10.1109/ACCESS.2021.3089681.

[24] B. Bojarajulu, S. Tanwar and A. Rana, "A Synoptic Review on Feature Selection and Machine Learning models used for Detecting Cyber Attacks in IoT," 2021 6th International Conference on Computing, Communication and Security (ICCCS), 2021, pp. 1-7, doi: 10.1109/ICCCS51487.2021.9776344.

[25] N. Koroniotis, N. Moustafa and E. Sitnikova, "Forensics and Deep Learning Mechanisms for Botnets in Internet of Things: A Survey of Challenges and Solutions," in IEEE Access, vol. 7, pp. 61764-61785, 2019, doi: 10.1109/ACCESS.2019.2916717.

[26] M. Mamdouh, M. A. I. Elrukhsi and A. Khattab, "Securing the Internet of Things and Wireless Sensor Networks via Machine Learning: A Survey," 2018 International Conference on Computer and Applications (ICCA), 2018, pp. 215-218, doi: 10.1109/COMAPP.2018.8460440.

[27] S. Abdelhamid, M. Aref, I. Hegazy and M. Roushdy, "A Survey on Learning-Based Intrusion Detection Systems for IoT Networks," 2021 Tenth International Conference on Intelligent Computing and Information Systems (ICICIS), 2021, pp. 278-288, doi: 10.1109/ICICIS52592.2021.9694226.

[28] A. Uprety and D. B. Rawat, "Reinforcement Learning for IoT Security: A Comprehensive Survey," in IEEE Internet of Things Journal, vol. 8, no. 11, pp. 8693-8706, 1 June1, 2021, doi: 10.1109/JIOT.2020.3040957.

[29] G. Lin, S. Wen, Q. -L. Han, J. Zhang and Y. Xiang, "Software Vulnerability Detection Using Deep Neural Networks: A Survey," in Proceedings of the IEEE, vol. 108, no. 10, pp. 1825-1848, Oct. 2020, doi: 10.1109/JPROC.2020.2993293.

[30] Y. Xin et al., "Machine Learning and Deep Learning Methods for Cybersecurity," in IEEE Access, vol. 6, pp. 35365-35381, 2018, doi: 10.1109/ACCESS.2018.2836950.

[31] I. Idrissi, M. Azizi and O. Moussaoui, "IoT security with Deep Learning-based Intrusion Detection Systems: A systematic literature review," 2020 Fourth International Conference On Intelligent Computing in Data Sciences (ICDS), 2020, pp. 1-10, doi: 10.1109/ICDS50568.2020.9268713.

[32] Kaspersky Security Risk Report. Available at: https://www.kaspersky.com/blog/security_risks_report_financial_impact/ (Accessed: September 12, 2022).

[33] I. Ben Abdel Ouahab, L. Elaachak, Y. A. Alluhaidan and M. Bouhorma, "A new approach to detect next generation of malware based on machine learning," 2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), 2021, pp. 230-235, doi: 10.1109/3ICT53449.2021.9581625.

[34] Cisco Annual Report. Available at: https://www.cisco.com/c/dam/en_us/about/annual-report/cisco-annual-report-2022.pdf (Accessed: January 10, 2023).

[35] W. Iqbal, H. Abbas, M. Daneshmand, B. Rauf and Y. A. Bangash, "An In-Depth Analysis of IoT Security Requirements, Challenges, and Their Countermeasures via Software-Defined Security," in IEEE Internet of Things Journal, vol. 7, no. 10,

pp. 10250-10276, Oct. 2020, doi: 10.1109/JIOT.2020.2997651.

[36] M. T. Mahmood, S. R. A. Ahmed and M. R. A. Ahmed, "Using Machine Learning To Secure IOT Systems," 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), 2020, pp. 1-7, doi: 10.1109/ISMSIT50672.2020.9254304.

[37] V. Hassija, V. Chamola, V. Saxena, D. Jain, P. Goyal and B. Sikdar, "A Survey on IoT Security: Application Areas, Security Threats, and Solution Architectures," in IEEE Access, vol. 7, pp. 82721-82743, 2019, doi: 10.1109/ACCESS.2019.2924045.

[38] M. Mainuddin, Z. Duan and Y. Dong, "Network Traffic Characteristics of IoT Devices in Smart Homes," 2021 International Conference on Computer Communications and Networks (ICCCN), 2021, pp. 1-11, doi: 10.1109/ICCCN52240.2021.9522168.

[39] Muhammad Azhar Iqbal; Sajjad Hussain; Huanlai Xing; Muhammad Ali Imran, "Internet of Things (IoT) Fundamentals," in Enabling the Internet of Things: Fundamentals, Design and Applications , IEEE, 2021, pp.1-28, doi: 10.1002/9781119701460.ch1.

[40] Y. Kumar and B. Subba, "A lightweight machine learning based security framework for detecting phishing attacks," 2021 International Conference on COMmunication Systems & NETworkS (COMSNETS), 2021, pp. 184-188, doi: 10.1109/COMSNETS51098.2021.9352828.

[41] F. Li et al., "Online Distributed IoT Security Monitoring With Multidimensional Streaming Big Data," in IEEE Internet of Things Journal, vol. 7, no. 5, pp. 4387-4394, May 2020, doi: 10.1109/JIOT.2019.2962788.

[42] M. Raza et al., "Challenges and Limitations of Internet of Things Enabled Healthcare in COVID-19," in IEEE Internet of Things Magazine, vol. 4, no. 3, pp. 60-65, September 2021, doi: 10.1109/IOTM.0001.2000176.

[43] W. Yang, M. N. Johnstone, L. F. Sikos and S. Wang, "Security and Forensics in the Internet of Things: Research Advances and Challenges," 2020 Workshop on Emerging Technologies for Security in IoT (ETSecIoT), 2020, pp. 12-17, doi: 10.1109/ETSecIoT50046.2020.00007.

[44] N. M. Min, V. Visoottiviseth, S. Teerakanok and N. Yamai, "OWASP IoT Top 10 based Attack Dataset for Machine Learning," 2022 24th International Conference on Advanced Communication Technology (ICACT), 2022, pp. 317-322, doi: 10.23919/ICACT53585.2022.9728969.

[45] Y. Wang et al., "IoT Device Identification Using Supervised Machine Learning," 2022 IEEE International Conference on Consumer Electronics (ICCE), 2022, pp. 1-6, doi: 10.1109/ICCE53296.2022.9730354.

[46] C. D. McDermott, F. Majdani and A. V. Petrovski, "Botnet Detection in the Internet of Things using Deep Learning Approaches," 2018 International Joint Conference on Neural Networks (IJCNN), 2018, pp. 1-8, doi: 10.1109/IJCNN.2018.8489489.

[47] X. Liang and Y. Kim, "A Survey on Security Attacks and Solutions in the IoT Network," 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC), 2021, pp. 0853-0859, doi: 10.1109/CCWC51732.2021.9376174.

[48] P. K. Sharma, M.-Y. Chen and J. H. Park, ”A software defined fog node based distributed blockchain cloud architecture for IoT”, IEEE Access, vol. 6, pp. 115-124, 2018.

[49] S. A. Sokolov, T. B. Iliev and I. S. Stoyanov, "Analysis of Cybersecurity Threats in Cloud Applications Using Deep Learning Techniques," 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2019, pp. 441-446, doi: 10.23919/MIPRO.2019.8756755.

[50] T. U. Chai, H. G. Goh and V. Ponnusamy, "A Study of Security Threat for Internet of Things in Smart Factory," 2021 IEEE International Conference on Computing (ICOCO), 2021, pp. 97-102, doi: 10.1109/ICOCO53166.2021.9673550.

[51] H. Djuitcheu, M. Debes, M. Aumüller and J. Seitz, "Recent review of Distributed Denial of Service Attacks in the Internet of Things," 2022 5th Conference on Cloud and Internet of Things (CIoT), 2022, pp. 32-39, doi: 10.1109/CIoT53061.2022.9766655.

[52] S. Latif, Z. Zou, Z. Idrees and J. Ahmad, "A Novel Attack Detection Scheme for the Industrial Internet of Things Using a Lightweight Random Neural Network," in IEEE Access, vol. 8, pp. 89337-89350, 2020, doi: 10.1109/ACCESS.2020.2994079.

[53] Z. Chkirbene, S. Eltanbouly, M. Bashendy, N. AlNaimi and A. Erbad, "Hybrid Machine Learning for Network Anomaly Intrusion Detection," 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), 2020, pp. 163-170, doi: 10.1109/ICIoT48696.2020.9089575.

[54] A. Al-Bakaa and B. Al-Musawi, "Improving the Performance of Intrusion Detection System through Finding the Most Effective Features," 2021 International Congress of Advanced Technology and Engineering (ICOTEN), 2021, pp. 1-9, doi: 10.1109/ICOTEN52080.2021.9493564.

[55] F. Pistorius, D. Grimm, F. Erdösi and E. Sax, "Evaluation Matrix for Smart Machine-Learning Algorithm Choice," 2020 1st International Conference on Big Data Analytics and Practices (IBDAP), 2020, pp. 1-6, doi: 10.1109/IBDAP50342.2020.9245610.

[56] K. Sharma and R. Nandal, "A Literature Study On Machine Learning Fusion With IOT," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), 2019, pp. 1440-1445, doi: 10.1109/ICOEI.2019.8862656.

[57] Y. Xin et al., "Machine Learning and Deep Learning Methods for Cybersecurity," in IEEE

Access, vol. 6, pp. 35365-35381, 2018, doi: 10.1109/ACCESS.2018.2836950.

[58] E. Elbasi, S. Mathew, A. E. Topcu and W. Abdelbaki, "A Survey on Machine Learning and Internet of Things for COVID-19," 2021 IEEE World AI IoT Congress (AIIoT), 2021, pp. 0115-0120, doi: 10.1109/AIIoT52608.2021.9454241.

[59] F. Liang, W. G. Hatcher, W. Liao, W. Gao and W. Yu, "Machine Learning for Security and the Internet of Things: The Good, the Bad, and the Ugly," in IEEE Access, vol. 7, pp. 158126-158147, 2019, doi: 10.1109/ACCESS.2019.2948912.

[60] S. Malik and R. Chauhan, "Securing the Internet of Things using Machine Learning: A Review," 2020 International Conference on Convergence to Digital World - Quo Vadis (ICCDW), 2020, pp. 1-4, doi: 10.1109/ICCDW45521.2020.9318666.

[61] M. A. Mahmood and A. M. Zeki, "Securing IOT against DDOS attacks using machine learning," 3rd Smart Cities Symposium (SCS 2020), 2020, pp. 471-476, doi: 10.1049/icp.2021.0905.

[62] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, ``Deep reinforcement learning: A brief survey," IEEE Signal Process. Mag., vol. 34, no. 6, pp. 2638, Nov. 2017.

[63] J. Moradi, H. Shahinzadeh, H. Nafisi, M. Marzband and G. B. Gharehpetian, "Attributes of Big Data Analytics for Data-Driven Decision Making in Cyber-Physical Power Systems," 2020 14th International Conference on Protection and Automation of Power Systems (IPAPS), 2019, pp. 83-92, doi: 10.1109/IPAPS49326.2019.9069391.

[64] R. Manzano S., N. Goel, M. Zaman, R. Joshi and K. Naik, "Design of a Machine Learning Based Intrusion Detection Framework and Methodology for IoT Networks," 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC), 2022, pp. 0191-0198, doi: 10.1109/CCWC54503.2022.9720857.

[65] Top Search Terms from IEEE Xplore. Available at: https://ieeexplore-ieee-org.proxy.lib.odu.edu/popular/all (Accessed: January 12, 2023).

[66] Y. Wang et al., "A smart campus internet of things framework," 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), 2017, pp. 498-503, doi: 10.1109/UEMCON.2017.8249106.

[67] R. Gandhi and Y. Li, "Comparing Machine Learning and Deep Learning for IoT Botnet Detection," 2021 IEEE International Conference on Smart Computing (SMARTCOMP), 2021, pp. 234-239, doi: 10.1109/SMARTCOMP52413.2021.00053.

[68] L. Holbrook and M. Alamaniotis, "Internet of Things Security Analytics and Solutions with Deep Learning," 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI), 2019, pp. 178-185, doi: 10.1109/ICTAI.2019.00033.

[69] M. Munir, M. A. Chattha, A. Dengel and S. Ahmed, "A Comparative Analysis of Traditional and Deep Learning-Based Anomaly Detection Methods for Streaming Data," 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA), 2019, pp. 561-566, doi: 10.1109/ICMLA.2019.00105.

[70] B. Majhi and Prastavana, "An Improved Intrusion Dectection System using BoT-IoT Dataset," 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT), 2022, pp. 488-492, doi: 10.1109/CSNT54456.2022.9787639.

[71] T.Fawcett, „ROC Graphs: Notes and Practical Considerations for Researchers.” Kluwer Academic Publishers, 2004.

[72] Vujović, Z. (2021). Classification model evaluation metrics. International Journal of Advanced Computer Science and Applications, 12(6), 599-606.

[73] B. Urooj, M. A. Shah, C. Maple, M. K. Abbasi and S. Riasat, "Malware Detection: A Framework for Reverse Engineered Android Applications Through Machine Learning Algorithms," in IEEE Access, vol. 10, pp. 89031-89050, 2022, doi: 10.1109/ACCESS.2022.3149053.

[74] Meidan, Y., Bohadana, M., Mathov, Y., Mirsky, Y., Shabtai, A., Breitenbacher, D., & Elovici, Y. (2018). N-baiot—network-based detection of iot botnet attacks using deep autoencoders. IEEE Pervasive Computing, 17(3), 12-22.

[75] Dataset Source: https://github.com/dung4883/PSI_graphIoTBotnet , Accessed 15 Nov 2021

[76] Nour Moustafa, October 16, 2019, "The Bot-IoT dataset", IEEE Dataport. [Dataset]. Available: https://ieee-dataport.org/documents/bot-iot-dataset . [Accessed: Nov. 17, 2022].

[77] M. Tavallaee, E. Bagheri, W. Lu, and A. Ghorbani,, 2009, “A Detailed Analysis of the KDD CUP 99 Data Set,” Submitted to Second IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA). [Dataset]. Available: https://www.unb.ca/cic/datasets/nsl.html . [Accessed: Nov. 23, 2022].

[78] Nour Moustafa, October 16, 2019, "UNSW_NB15 dataset", IEEE Dataport, [Dataset]. Available: https://research.unsw.edu.au/projects/unsw-nb15-dataset . [Accessed: Nov. 13, 2022].

[79] Iman Sharafaldin, Arash Habibi Lashkari, and Ali A. Ghorbani, “Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization”, 4th International Conference on Information Systems Security and Privacy (ICISSP), Portugal, January 2018.). [Dataset]. Available: https://www.unb.ca/cic/datasets/ids-2017.html . [Accessed: Dec. 2, 2022].

[80] Y. Meidan, M. Bohadana, Y. Mathov, Y. Mirsky, D. Breitenbacher, A. Shabtai, and Y. Elovici 'N-BaIoT: Network-based Detection of IoT Botnet

Attacks Using Deep Autoencoders', IEEE Pervasive Computing, Special Issue - Securing the IoT (July/Sep 2018).). [Dataset]. Available: https://www.unb.ca/cic/datasets/nsl.html . [Accessed: Dec. 2, 2022].

[81] "Sebastian Garcia, Agustin Parmisano, & Maria Jose Erquiaga. (2020). IoT-23: A labeled dataset with malicious and benign IoT network traffic (Version 1.0.0) [Data set]. Zenodo. [Dataset]. Available: https://www.stratosphereips.org/datasets-iot23. [Accessed: Nov. 26, 2022].

[82] Stephen D. Bay and Dennis F. Kibler and Michael J. Pazzani and Padhraic Smyth. The UCI KDD Archive of Large Data Sets for Data Mining Research and Experimentation. SIGKDD Explorations, 2. 2000.). [Dataset]. Available: https://archive.ics.uci.edu/ml/datasets/KDD+Cup+1999+Data . [Accessed: Dec. 9, 2022].

[83] Ali Shiravi, Hadi Shiravi, Mahbod Tavallaee, Ali A. Ghorbani, Toward developing a systematic approach to generate benchmark datasets for intrusion detection, Computers & Security, Volume 31, Issue 3, May 2012, Pages 357-374, ISSN 0167-4048, 10.1016/j.cose.2011.12.012.). [Dataset]. Available: https://www.unb.ca/cic/datasets/ids.html . [Accessed: Dec. 14, 2022].

[84] Iman Sharafaldin, Arash Habibi Lashkari, Saqib Hakak, and Ali A. Ghorbani, "Developing Realistic Distributed Denial of Service (DDoS) Attack Dataset and Taxonomy", IEEE 53rd International Carnahan Conference on Security Technology, Chennai, India, 2019.). [Dataset]. Available: https://www.unb.ca/cic/datasets/ddos-2019.html. [Accessed: Dec. 16, 2022].

[85] Sebastian Garcia, Martin Grill, Jan Stiborek and Alejandro Zunino. "An empirical comparison of botnet detection methods" Computers and Security Journal, Elsevier. 2014. Vol 45, pp 100-123. [Dataset]. Available: https://www.stratosphereips.org/datasets-ctu13 . [Accessed: Dec. 14, 2022].

[86] Nour Moustafa, October 16, 2019, "ToN_IoT datasets", IEEE Dataport. [Dataset]. Available: https://ieee-dataport.org/documents/toniot-datasets . [Accessed: De. 16, 2022].

[87] Iman Sharafaldin, Arash Habibi Lashkari, and Ali A. Ghorbani, "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization", 4th International Conference on Information Systems Security and Privacy (ICISSP), Portugal, January 2018). [Dataset]. Available: https://www.unb.ca/cic/datasets/ids-2018.html. [Accessed: Nov. 17, 2022].

[88] Seiya Kato, Rui Tanabe, Katsunari Yoshioka, Tsutomu Matsumoto, "Adaptive Observation of Emerging Cyber Attacks targeting Various IoT Devices," IFIP/IEEE International Symposium on Integrated Network Management (IM), 2021). [Dataset]. Available: https://sec.ynu.codes/iot/. [Accessed: Dec. 14, 2022].

[89] External Data Source, "VirusShare Dataset." IMPACT, 2018). [Dataset]. Available: http://www.secrepo.com/Datasets%20Description/PE_malware/VirusShare.html . [Accessed: Nov. 22, 2022].

[90] Ds2os Traffic Traces. Available online: https://www.kaggle.com/francoisxa/ds2ostraffictraces (accessed on 22 May 2022).). [Dataset]. Available: https://www.kaggle.com/datasets/francoisxa/ds2ostraffictraces . [Accessed: Nov. 22, 2022].

[91] Anubis: Analyzing Unknown Binaries. [Dataset]. Available: https://www.unb.ca/cic/datasets/nsl.html . [Accessed: Dec. 16. 2022].

[92] NIST Software Assurance Reference Dataset (SARD). [Dataset]. Available: https://samate.nist.gov/SARD/ . [Accessed: Nov. 17, 2022].

[93] Das Malwerk. Malware Downloading Website. Accessed: Jan. 15, 2021. [Dataset]. Available: https://dasmalwerk.eu/ . [Accessed: Nov. 22, 2022].

[94] M. Shobana and S. Poonkuzhali, "A novel approach to detect IoT malware by system calls using Deep learning techniques," 2020 International Conference on Innovative Trends in Information Technology (ICITIIT), 2020, pp. 1-5, doi: 10.1109/ICITIIT49094.2020.9071531.

[95] A. P. Kuruvila, S. Kundu and K. Basu, "Analyzing the Efficiency of Machine Learning Classifiers in Hardware-Based Malware Detectors," 2020 IEEE Computer Society Annual Symposium on VLSI (ISVLSI), 2020, pp. 452-457, doi: 10.1109/ISVLSI49217.2020.00-15.

[96] P. R. K. Varma, K. P. Raj and K. V. S. Raju, "Android mobile security by detecting and classification of malware based on permissions using machine learning algorithms," 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2017, pp. 294-299, doi: 10.1109/I-SMAC.2017.8058358.

[97] Q. -G. Lin, N. Li, Q. Qi and J. -B. Hu, "Classification of IoT Malware based on Convolutional Neural Network," 2020 International Conference on Service Science (ICSS), 2020, pp. 51-57, doi: 10.1109/ICSS50103.2020.00016.

[98] Y. -T. Lee et al., "Cross Platform IoT-Malware Family Classification Based on Printable Strings," 2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), 2020, pp. 775-784, doi: 10.1109/TrustCom50675.2020.00106.

[99] K. Singh, S. S. Grover and R. K. Kumar, "Cyber Security Vulnerability Detection Using Natural Language Processing," 2022 IEEE World AI IoT Congress (AIIoT), 2022, pp. 174-178, doi: 10.1109/AIIoT54504.2022.9817336.

[100] Vijayaraj, S. M, R. M, U. Vijayaraj, D. Kamaleshwar and D. Rajalakshmi, "Decision Trees to Detect Malware in a Cloud Computing Environment," 2022 International Conference on Electronic Systems and Intelligent Computing (ICESIC), 2022, pp. 299-303, doi: 10.1109/ICESIC53714.2022.9783547.

[101] G. AbdulsalamYa'u, G. K. Job, S. M. Waziri, B. Jaafar, N. A. SabonGari and I. Z. Yakubu, "Deep Learning for Detecting Ransomware in Edge Computing Devices Based On Autoencoder Classifier," 2019 4th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT), 2019, pp. 240-243, doi: 10.1109/ICEECCOT46775.2019.9114576.

[102] M. Ficco, "Detecting IoT Malware by Markov Chain Behavioral Models," 2019 IEEE International Conference on Cloud Engineering (IC2E), 2019, pp. 229-234, doi: 10.1109/IC2E.2019.00037.

[103] W. Niu, X. Zhang, X. Du, T. Hu, X. Xie and N. Guizani, "Detecting Malware on X86-Based IoT Devices in Autonomous Driving," in IEEE Wireless Communications, vol. 26, no. 4, pp. 80-87, August 2019, doi: 10.1109/MWC.2019.1800505.

[104] A. Kumar and T. J. Lim, "EDIMA: Early Detection of IoT Malware Network Activity Using Machine Learning Techniques," 2019 IEEE 5th World Forum on Internet of Things (WF-IoT), 2019, pp. 289-294, doi: 10.1109/WF-IoT.2019.8767194.

[105] S. Ali, O. Abusabha, F. Ali, M. Imran and T. ABUHMED, "Effective Multitask Deep Learning for IoT Malware Detection and Identification Using Behavioral Traffic Analysis," in IEEE Transactions on Network and Service Management, 2022, doi: 10.1109/TNSM.2022.3200741.

[106] I. Vaccari, S. Narteni, M. Aiello, M. Mongelli and E. Cambiaso, "Exploiting Internet of Things Protocols for Malicious Data Exfiltration Activities," in IEEE Access, vol. 9, pp. 104261-104280, 2021, doi: 10.1109/ACCESS.2021.3099642.

[107] G. Bendiab, S. Shiaeles, A. Alruban and N. Kolokotronis, "IoT Malware Network Traffic Classification using Visual Representation and Deep Learning," 2020 6th IEEE Conference on Network Softwarization (NetSoft), 2020, pp. 444-449, doi: 10.1109/NetSoft48620.2020.9165381.

[108] R. Yumlembam, B. Issac, S. M. Jacob and L. Yang, "IoT-based Android Malware Detection Using Graph Neural Network With Adversarial Defense," in IEEE Internet of Things Journal, 2022, doi: 10.1109/JIOT.2022.3188583.

[109] K. Wehbi, L. Hong, T. Al-salah and A. A. Bhutta, "A Survey on Machine Learning Based Detection on DDoS Attacks for IoT Systems," 2019 SoutheastCon, 2019, pp. 1-6, doi: 10.1109/SoutheastCon42311.2019.9020468.

[110] Roopak, M., Tian, G. Y., & Chambers, J. (2020, January). An intrusion detection system against ddos attacks in iot networks. In 2020 10th annual computing and communication workshop and conference (CCWC) (pp. 0562-0567). IEEE.

[111] K. Gurulakshmi and A. Nesarani, "Analysis of IoT Bots Against DDOS Attack Using Machine Learning Algorithm," 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), 2018, pp. 1052-1057, doi: 10.1109/ICOEI.2018.8553896.

[112] I. Cvitić, D. Perakovic, B. B. Gupta and K. -K. R. Choo, "Boosting-Based DDoS Detection in Internet of Things Systems," in IEEE Internet of Things Journal, vol. 9, no. 3, pp. 2109-2123, 1 Feb.1, 2022, doi: 10.1109/JIOT.2021.3090909.

[113] J. N. Bakker, B. Ng and W. K. G. Seah, "Can Machine Learning Techniques Be Effectively Used in Real Networks against DDoS Attacks?," 2018 27th International Conference on Computer Communication and Networks (ICCCN), 2018, pp. 1-6, doi: 10.1109/ICCCN.2018.8487445.

[114] Y. N. Soe, P. I. Santosa and R. Hartanto, "DDoS Attack Detection Based on Simple ANN with SMOTE for IoT Environment," 2019 Fourth International Conference on Informatics and Computing (ICIC), 2019, pp. 1-5, doi: 10.1109/ICIC47613.2019.8985853.

[115] M. Roopak, G. Yun Tian and J. Chambers, "Deep Learning Models for Cyber Security in IoT Networks," 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), 2019, pp. 0452-0457, doi: 10.1109/CCWC.2019.8666588.

[116] B. Özçam, H. H. Kilinc and A. H. Zaim, "Detecting TCP Flood DDoS Attack by Anomaly Detection based on Machine Learning Algorithms," 2021 6th International Conference on Computer Science and Engineering (UBMK), 2021, pp. 512-516, doi: 10.1109/UBMK52708.2021.9558989.

[117] O. R. Sanchez, M. Repetto, A. Carrega and R. Bolla, "Evaluating ML-based DDoS Detection with Grid Search Hyperparameter Optimization," 2021 IEEE 7th International Conference on Network Softwarization (NetSoft), 2021, pp. 402-408, doi: 10.1109/NetSoft51509.2021.9492633.

[118] D. Nanthiya, P. Keerthika, S. B. Gopal, S. B. Kayalvizhi, T. Raja and R. S. Priya, "SVM Based DDoS Attack Detection in IoT Using Iot-23 Botnet Dataset," 2021 Innovations in Power and Advanced Computing Technologies (i-PACT), 2021, pp. 1-7, doi: 10.1109/i-PACT52855.2021.9696569.

[119] R. Doshi, N. Apthorpe and N. Feamster, "Machine Learning DDoS Detection for Consumer Internet of Things Devices," 2018 IEEE Security and Privacy Workshops (SPW), 2018, pp. 29-35, doi: 10.1109/SPW.2018.00013.

[120] Y. Jia, F. Zhong, A. Alrawais, B. Gong and X. Cheng, "FlowGuard: An Intelligent Edge Defense Mechanism Against IoT DDoS Attacks," in IEEE Internet of Things Journal, vol. 7, no. 10, pp. 9552-9562, Oct. 2020, doi: 10.1109/JIOT.2020.2993782.

[121] Aysa, M. H., Ibrahim, A. A., & Mohammed, A. H. (2020, October). Iot ddos attack detection using machine learning. In 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) (pp. 1-7). IEEE.

[122] Ilango, H. S., Ma, M., & Su, R. (2021, December). Low Rate DoS Attack Detection in IoT-SDN using Deep Learning. In 2021 IEEE International Conferences on Internet of Things (iThings) and IEEE Green Computing & Communications (GreenCom) and IEEE Cyber, Physical & Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics) (pp. 115-120). IEEE.

[123] Pwint, P. H., & Shwe, T. (2019, November). Network traffic anomaly detection based on Apache Spark. In 2019 international conference on advanced information technologies (ICAIT) (pp. 222-226). IEEE.

[124] Elbasi, E. (2020, December). Reliable abnormal event detection from IoT surveillance systems. In 2020 7th International Conference on Internet of Things: Systems, Management and Security (IOTSMS) (pp. 1-5). IEEE.

[125] Alrashdi, I., Alqazzaz, A., Aloufi, E., Alharthi, R., Zohdy, M., & Ming, H. (2019, January). 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) (pp. 0305-0310). IEEE.

[126] Abbasi, F., Naderan, M., & Alavi, S. E. (2021, May). Anomaly detection in Internet of Things using feature selection and classification based on Logistic Regression and Artificial Neural Network on N-BaIoT dataset. In 2021 5th International Conference on Internet of Things and Applications (IoT) (pp. 1-7). IEEE.

[127] Huč, A., & Trček, D. (2021). Anomaly detection in IoT networks: From architectures to machine learning transparency. IEEE Access, 9, 60607-60616.

[128] Pathak, A. K., Saguna, S., Mitra, K., & Åhlund, C. (2021, June). Anomaly detection using machine learning to discover sensor tampering in IoT systems. In ICC 2021-IEEE International Conference on Communications (pp. 1-6). IEEE.

[129] Rosenthal, G., Kdosha, O. E., Cohen, K., Freund, A., Bartik, A., & Ron, A. (2020). ARBA: Anomaly and reputation based approach for detecting infected IoT devices. IEEE Access, 8, 145751-145767.

[130] Hsieh, R. J., Chou, J., & Ho, C. H. (2019, November). Unsupervised online anomaly detection on multivariate sensing time series data for smart manufacturing. In 2019 IEEE 12th Conference on Service-Oriented Computing and Applications (SOCA) (pp. 90-97). IEEE.

[131] An, Y., Li, J., Yu, F. R., Chen, J., & Leung, V. C. (2020). A Novel HTTP Anomaly Detection Framework Based on Edge Intelligence for the Internet of Things (IoT). IEEE Wireless Communications, 28(2), 159-165.

[132] Abdelmoumin, G., Rawat, D. B., & Rahman, A. (2021). On the performance of machine learning models for anomaly-based intelligent intrusion detection systems for the internet of things. IEEE Internet of Things Journal, 9(6), 4280-4290.

[133] Ma, J., & Lin, S. (2019, December). Big data enabled anomaly user detection in mobile wireless networks. In 2019 IEEE 5th International Conference on Computer and Communications (ICCC) (pp. 479-484). IEEE.

[134] Ayad, A., Zamani, A., Schmeink, A., & Dartmann, G. (2019, October). Design and implementation of a hybrid anomaly detection system for IoT. In 2019 Sixth International Conference on Internet of Things: Systems, Management and Security (IOTSMS) (pp. 1-6). IEEE.

[135] An, Y., Yu, F. R., Li, J., Chen, J., & Leung, V. C. (2020). Edge intelligence (EI)-enabled HTTP anomaly detection framework for the Internet of Things (IoT). IEEE Internet of Things Journal, 8(5), 3554-3566.

[136] Mothukuri, V., Khare, P., Parizi, R. M., Pouriyeh, S., Dehghantanha, A., & Srivastava, G. (2021). Federated-Learning-Based Anomaly Detection for IoT Security Attacks. IEEE Internet of Things Journal, 9(4), 2545-2554.

[137] Sahu, N. K., & Mukherjee, I. (2020, June). Machine learning based anomaly detection for IoT network:(Anomaly detection in IoT network). In 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184) (pp. 787-794). IEEE.

[138] Gulhare, A. K., Badholia, A., & Sharma, A. (2022, July). Mean-Shift and Local Outlier Factor-Based Ensemble Machine Learning Approach for Anomaly Detection in IoT Devices. In 2022 International Conference on Inventive Computation Technologies (ICICT) (pp. 649-656). IEEE.

[139] R. Kale, Z. Lu, K. W. Fok and V. L. L. Thing, "A Hybrid Deep Learning Anomaly Detection Framework for Intrusion Detection," 2022 IEEE 8th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS), 2022, pp. 137-142, doi: 10.1109/BigDataSecurityHPSCIDS54978.2022.00034.

[140] G. Guo, "A Machine Learning Framework for Intrusion Detection System in IoT Networks Using an Ensemble Feature Selection Method," 2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2021, pp. 0593-0599, doi: 10.1109/IEMCON53756.2021.9623082.

[141] M. Bagaa, T. Taleb, J. B. Bernabe and A. Skarmeta, "A Machine Learning Security Framework for Iot Systems," in IEEE Access, vol. 8, pp. 114066-114077, 2020, doi: 10.1109/ACCESS.2020.2996214.

[142] S. A. Abdulkareem, C. H. Foh, F. Carrez and K. Moessner, "FI-PCA for IoT Network Intrusion Detection," 2022 International Symposium on Networks, Computers and Communications (ISNCC), 2022, pp. 1-6, doi: 10.1109/ISNCC55209.2022.9851723.

[143] Divakar, S., Priyadarshini, R., Barik, R. K., & Roy, D. S. (2021, January). An Intelligent Intrusion Detection Scheme Powered by Boosting Algorithm. In 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 205-209). IEEE.

[144] N. Chaabouni, M. Mosbah, A. Zemmari and C. Sauvignac, "A OneM2M Intrusion Detection and Prevention System based on Edge Machine Learning," NOMS 2020 - 2020 IEEE/IFIP Network Operations and Management Symposium, 2020, pp. 1-7, doi: 10.1109/NOMS47738.2020.9110473.

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

[146] J. Liu, D. Yang, M. Lian and M. Li, "Research on Classification of Intrusion Detection in Internet of Things Network Layer Based on Machine Learning," 2021 IEEE International Conference on Intelligence and Safety for Robotics (ISR), 2021, pp. 106-110, doi: 10.1109/ISR50024.2021.9419529.

[147] Ravi, N., & Shalinie, S. M. (2020). Semisupervised-learning-based security to detect and mitigate intrusions in IoT network. IEEE Internet of Things Journal, 7(11), 11041-11052.

[148] G. Abdelmoumin, D. B. Rawat and A. Rahman, "On the Performance of Machine Learning Models for Anomaly-Based Intelligent Intrusion Detection Systems for the Internet of Things," in IEEE Internet of Things Journal, vol. 9, no. 6, pp. 4280-4290, 15 March15, 2022, doi: 10.1109/JIOT.2021.3103829.

[149] Liu, Z., Thapa, N., Shaver, A., Roy, K., Yuan, X., & Khorsandroo, S. (2020, August). Anomaly detection on iot network intrusion using machine learning. In 2020 International conference on artificial intelligence, big data, computing and data communication systems (icABCD) (pp. 1-5). IEEE.

[150] M. A. Cheema, H. Khaliq Qureshi, C. Chrysostomou and M. Lestas, "Utilizing Blockchain for Distributed Machine Learning based Intrusion Detection in Internet of Things," 2020 16th International Conference on Distributed Computing in Sensor Systems (DCOSS), 2020, pp. 429-435, doi: 10.1109/DCOSS49796.2020.00074.

[151] R. M and V. Ananthanarayanan, "Machine Learning based Prediction Analysis in Intrusion Detection," 2022 International Conference on Electronics and Renewable Systems (ICEARS), 2022, pp. 1153-1159, doi: 10.1109/ICEARS53579.2022.9752061.

[152] A. Yahyaoui, H. Lakhdhar, T. Abdellatif and R. Attia, "Machine learning based network intrusion detection for data streaming IoT applications," 2021 21st ACIS International Winter Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD-Winter), 2021, pp. 51-56, doi: 10.1109/SNPDWinter52325.2021.00019.

[153] A. P. Singh, S. Kumar, A. Kumar and M. Usama, "Machine Learning based Intrusion Detection System for Minority Attacks Classification," 2022 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), 2022, pp. 256-261, doi: 10.1109/CISES54857.2022.9844381.

[154] K. Saurabh et al., "LBDMIDS: LSTM Based Deep Learning Model for Intrusion Detection Systems for IoT Networks," 2022 IEEE World AI IoT Congress (AIIoT), 2022, pp. 753-759, doi: 10.1109/AIIoT54504.2022.9817245.

[155] M. M. Alani, "IoTProtect: A Machine-Learning Based IoT Intrusion Detection System," 2022 6th International Conference on Cryptography, Security and Privacy (CSP), 2022, pp. 61-65, doi: 10.1109/CSP55486.2022.00020.

[156] V. Ponnusamy and B. Sharma, "Investigation on IoT Intrusion Detection in Wireless Environment," 2021 International Conference on Computer & Information Sciences (ICCOINS), 2021, pp. 7-13, doi: 10.1109/ICCOINS49721.2021.9497203.

[157] J. Ashraf, N. Moustafa, A. D. Bukhshi and A. Javed, "Intrusion Detection System for SDN-enabled IoT Networks using Machine Learning Techniques," 2021 IEEE 25th International Enterprise Distributed Object Computing Workshop (EDOCW), 2021, pp. 46-52, doi: 10.1109/EDOCW52865.2021.00031.

[158] C. Liang, B. Shanmugam, S. Azam, M. Jonkman, F. D. Boer and G. Narayansamy, "Intrusion Detection System for Internet of Things based on a Machine Learning approach," 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN),

2019, pp. 1-6, doi: 10.1109/ViTECoN.2019.8899448.

[159] R. V. Mendonça et al., "Intrusion Detection System Based on Fast Hierarchical Deep Convolutional Neural Network," in IEEE Access, vol. 9, pp. 61024-61034, 2021, doi: 10.1109/ACCESS.2021.3074664.

[160] B. Susilo and R. F. Sari, "Intrusion Detection in Software Defined Network Using Deep Learning Approach," 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC), 2021, pp. 0807-0812, doi: 10.1109/CCWC51732.2021.9375951.

[161] S. Hanif, T. Ilyas and M. Zeeshan, "Intrusion Detection In IoT Using Artificial Neural Networks On UNSW-15 Dataset," 2019 IEEE 16th International Conference on Smart Cities: Improving Quality of Life Using ICT & IoT and AI (HONET-ICT), 2019, pp. 152-156, doi: 10.1109/HONET.2019.8908122.

[162] J. Jose, D. V. Jose, K. S. Rao and J. Janz, "Impact of Machine Learning Algorithms in Intrusion Detection Systems for Internet of Things," 2021 International Conference on Advances in Computing and Communications (ICACC), 2021, pp. 1-6, doi: 10.1109/ICACC-202152719.2021.9708404.

[163] D. Lightbody, D. -M. Ngo, A. Temko, C. Murphy and E. Popovici, "Host-Based Intrusion Detection System for IoT using Convolutional Neural Networks," 2022 33rd Irish Signals and Systems Conference (ISSC), 2022, pp. 1-7, doi: 10.1109/ISSC55427.2022.9826188.

[164] C. Jiang, J. Kuang and S. Wang, "Home IoT Intrusion Prevention Strategy Based on Edge Computing," 2019 IEEE 2nd International Conference on Electronics and Communication Engineering (ICECE), 2019, pp. 94-98, doi: 10.1109/ICECE48499.2019.9058536.

[165] R. Samdekar, S. M. Ghosh and K. Srinivas, "Efficiency Enhancement of Intrusion Detection in Iot Based on Machine Learning Through Bioinspire," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 2021, pp. 383-387, doi: 10.1109/ICICV50876.2021.9388392.

[166] V. Christopher et al., "Minority Resampling Boosted Unsupervised Learning With Hyperdimensional Computing for Threat Detection at the Edge of Internet of Things," in IEEE Access, vol. 9, pp. 126646-126657, 2021, doi: 10.1109/ACCESS.2021.3111053.

[167] A. Makkar, S. Garg, N. Kumar, M. S. Hossain, A. Ghoneim and M. Alrashoud, "An Efficient Spam Detection Technique for IoT Devices Using Machine Learning," in IEEE Transactions on Industrial Informatics, vol. 17, no. 2, pp. 903-912, Feb. 2021, doi: 10.1109/TII.2020.2968927.

[168] I. Kotenko, I. Saenko, A. Kushnerevich and A. Branitskiy, "Attack Detection in IoT Critical Infrastructures: A Machine Learning and Big Data Processing Approach," 2019 27th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP), 2019, pp. 340-347, doi: 10.1109/EMPDP.2019.8671571.

[169] Y. Zhang, C. Dukkipati and L. -C. Cheng, "Clustering Methods for Identification of Attacks in IoT Based Traffic Signal System," 2019 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC), 2019, pp. 24-28, doi: 10.1109/SDPC.2019.00013.

[170] S. Alevizopoulou, P. Koloveas, C. Tryfonopoulos and P. Raftopoulou, "Social Media Monitoring for IoT Cyber-Threats," 2021 IEEE International Conference on Cyber Security and Resilience (CSR), 2021, pp. 436-441, doi: 10.1109/CSR51186.2021.9527964.

[171] T. V. Khoa et al., "Collaborative Learning Model for Cyberattack Detection Systems in IoT Industry 4.0," 2020 IEEE Wireless Communications and Networking Conference (WCNC), 2020, pp. 1-6, doi: 10.1109/WCNC45663.2020.9120761.

[172] M. M. Rashid, J. Kamruzzaman, T. Imam, S. Kaisar and M. J. Alam, "Cyber Attacks Detection from Smart City Applications Using Artificial Neural Network," 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), 2020, pp. 1-6, doi: 10.1109/CSDE50874.2020.9411606.

[173] Y. Zhou, M. Han, L. Liu, J. S. He and Y. Wang, "Deep learning approach for cyberattack detection," IEEE INFOCOM 2018 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), 2018, pp. 262-267, doi: 10.1109/INFCOMW.2018.8407032.

[174] A. B. M. Sultan, S. Mehmood and H. Zahid, "Man in the Middle Attack Detection for MQTT based IoT devices using different Machine Learning Algorithms," 2022 2nd International Conference on Artificial Intelligence (ICAI), 2022, pp. 118-121, doi: 10.1109/ICAI55435.2022.9773590.

[175] A. Nascita, F. Cerasuolo, D. D. Monda, J. T. A. Garcia, A. Montieri and A. Pescapè, "Machine and Deep Learning Approaches for IoT Attack Classification," IEEE INFOCOM 2022 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), 2022, pp. 1-6, doi: 10.1109/INFOCOMWKSHPS54753.2022.9797971.

[176] Y. K. Saheed and M. O. Arowolo, "Efficient Cyber Attack Detection on the Internet of Medical Things-Smart Environment Based on Deep Recurrent Neural Network and Machine Learning Algorithms," in IEEE Access, vol. 9, pp. 161546-161554, 2021, doi: 10.1109/ACCESS.2021.3128837.

[177] L. T. Hong Van, P. Van Huong, T. Q. Hua, L. Duc Thuan and N. H. Minh, "Feature Generation by K-means for Convolutional Neural Network in Detecting IoT System Attacks," 2021 IEEE

International Conference on Machine Learning and Applied Network Technologies (ICMLANT), 2021, pp. 1-5, doi: 10.1109/ICMLANT53170.2021.9690532.

[178] N. Abdalgawad, A. Sajun, Y. Kaddoura, I. A. Zualkernan and F. Aloul, "Generative Deep Learning to Detect Cyberattacks for the IoT-23 Dataset," in IEEE Access, vol. 10, pp. 6430-6441, 2022, doi: 10.1109/ACCESS.2021.3140015.

[179] Susanto, D. Stiawan, M. A. S. Arifin, J. Rejito, M. Y. Idris and R. Budiarto, "A Dimensionality Reduction Approach for Machine Learning Based IoT Botnet Detection," 2021 8th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), 2021, pp. 26-30, doi: 10.23919/EECSI53397.2021.9624299.

[180] M. G. Desai, Y. Shi and K. Suo, "A Hybrid Approach for IoT Botnet Attack Detection," 2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2021, pp. 0590-0592, doi: 10.1109/IEMCON53756.2021.9623102.

[181] O. P. Dwyer, A. K. Marnerides, V. Giotsas and T. Mursch, "Profiling IoT-Based Botnet Traffic Using DNS," 2019 IEEE Global Communications Conference (GLOBECOM), 2019, pp. 1-6, doi: 10.1109/GLOBECOM38437.2019.9014300.

[182] H. Gandhi and V. Ribeiro, "Packet Batching for Reducing System Resource Consumption for Botnet Detection using Network Traffic Analysis," 2022 14th International Conference on COMmunication Systems & NETworkS (COMSNETS), 2022, pp. 1-6, doi: 10.1109/COMSNETS53615.2022.9668511.

[183] S. Sriram, R. Vinayakumar, M. Alazab and S. KP, "Network Flow based IoT Botnet Attack Detection using Deep Learning," IEEE INFOCOM 2020 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), 2020, pp. 189-194, doi: 10.1109/INFOCOMWKSHPS50562.2020.9162668.

[184] T. C. Tran and T. Khanh Dang, "Machine Learning for Multi-Classification of Botnets Attacks," 2022 16th International Conference on Ubiquitous Information Management and Communication (IMCOM), 2022, pp. 1-8, doi: 10.1109/IMCOM53663.2022.9721811.

[185] F. Jeelani, D. S. Rai, A. Maithani and S. Gupta, "The Detection of IoT Botnet using Machine Learning on IoT-23 Dataset," 2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM), 2022, pp. 634-639, doi: 10.1109/ICIPTM54933.2022.9754187.

[186] F. Hussain, S. G. Abbas, U. U. Fayyaz, G. A. Shah, A. Toqeer and A. Ali, "Towards a Universal Features Set for IoT Botnet Attacks Detection," 2020 IEEE 23rd International Multitopic Conference (INMIC), 2020, pp. 1-6, doi: 10.1109/INMIC50486.2020.9318106.

[187] S. Nõmm and H. Bahşi, "Unsupervised Anomaly Based Botnet Detection in IoT Networks," 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), 2018, pp. 1048-1053, doi: 10.1109/ICMLA.2018.00171.

[188] M. Lefoane, I. Ghafir, S. Kabir and I. -U. Awan, "Unsupervised Learning for Feature Selection: A Proposed Solution for Botnet Detection in 5G Networks," in IEEE Transactions on Industrial Informatics, vol. 19, no. 1, pp. 921-929, Jan. 2023, doi: 10.1109/TII.2022.3192044.

[189] Aysa, M. H., Ibrahim, A. A., & Mohammed, A. H. (2020, October). Iot ddos attack detection using machine learning. In 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) (pp. 1-7). IEEE.

[190] N. Hasan, Z. Chen, C. Zhao, Y. Zhu and C. Liu, "IoT Botnet Detection framework from Network Behavior based on Extreme Learning Machine," IEEE INFOCOM 2022 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), 2022, pp. 1-6, doi: 10.1109/INFOCOMWKSHPS54753.2022.9798307.

[191] H. -T. Nguyen, Q. -D. Ngo and V. -H. Le, "IoT Botnet Detection Approach Based on PSI graph and DGCNN classifier," 2018 IEEE International Conference on Information Communication and Signal Processing (ICICSP), 2018, pp. 118-122, doi: 10.1109/ICICSP.2018.8549713.

[192] S. Joshi and E. Abdelfattah, "Efficiency of Different Machine Learning Algorithms on the Multivariate Classification of IoT Botnet Attacks," 2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), 2020, pp. 0517-0521, doi: 10.1109/UEMCON51285.2020.9298095.

[193] A. Kumar and T. J. Lim, "EDIMA: Early Detection of IoT Malware Network Activity Using Machine Learning Techniques," 2019 IEEE 5th World Forum on Internet of Things (WF-IoT), 2019, pp. 289-294, doi: 10.1109/WF-IoT.2019.8767194.

[194] A. B. de Neira, A. M. Araujo and M. Nogueira, "Early Botnet Detection for the Internet and the Internet of Things by Autonomous Machine Learning," 2020 16th International Conference on Mobility, Sensing and Networking (MSN), 2020, pp. 516-523, doi: 10.1109/MSN50589.2020.00087.

[195] H. Bahşi, S. Nõmm and F. B. La Torre, "Dimensionality Reduction for Machine Learning Based IoT Botnet Detection," 2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV), 2018, pp. 1857-1862, doi: 10.1109/ICARCV.2018.8581205.

[196] P. R. K. Pranav, S. Verma, S. Shenoy and S. Saravanan, "Detection of Botnets in IoT Networks using Graph Theory and Machine Learning," 2022

6th International Conference on Trends in Electronics and Informatics (ICOEI), 2022, pp. 590-597, doi: 10.1109/ICOEI53556.2022.9777117.

[197] M. Raghavendra and Z. Chen, "Detecting IoT Botnets on IoT Edge Devices," 2022 IEEE International Conference on Communications Workshops (ICC Workshops), 2022, pp. 373-378, doi: 10.1109/ICCWorkshops53468.2022.9814555.

[198] M. Injadat, A. Moubayed and A. Shami, "Detecting Botnet Attacks in IoT Environments: An Optimized Machine Learning Approach," 2020 32nd International Conference on Microelectronics (ICM), 2020, pp. 1-4, doi: 10.1109/ICM50269.2020.9331794.

[199] M. Shafiq, Z. Tian, A. K. Bashir, X. Du and M. Guizani, "CorrAUC: A Malicious Bot-IoT Traffic Detection Method in IoT Network Using Machine-Learning Techniques," in IEEE Internet of Things Journal, vol. 8, no. 5, pp. 3242-3254, 1 March1, 2021, doi: 10.1109/JIOT.2020.3002255.

[200] H. Qiao, B. Novikov and J. O. Blech, "Concept Drift Analysis by Dynamic Residual Projection for Effectively Detecting Botnet Cyber-Attacks in IoT Scenarios," in IEEE Transactions on Industrial Informatics, vol. 18, no. 6, pp. 3692-3701, June 2022, doi: 10.1109/TII.2021.3108464.

[201] C. Long, X. Xiao, W. Wan, J. Zhao, J. Wei and G. Du, "Botnet Detection Based on Flow Summary and Graph Sampling with Machine Learning," 2021 International Conference on Computer Engineering and Application (ICCEA), 2021, pp. 309-317, doi: 10.1109/ICCEA53728.2021.00068.

[202] Alshamkhany, M., Alshamkhany, W., Mansour, M., Khan, M., Dhou, S., & Aloul, F. (2020, November). Botnet attack detection using machine learning. In 2020 14th International Conference on Innovations in Information Technology (IIT) (pp. 203-208). IEEE.

[203] K. N. Karaca and A. Çetin, "Botnet Attack Detection Using Convolutional Neural Networks in the IoT Environment," 2021 International Conference on INnovations in Intelligent SysTems and Applications (INISTA), 2021, pp. 1-6, doi: 10.1109/INISTA52262.2021.9548445.

[204] F. Abbasi, M. Naderan and S. E. Alavi, "Anomaly detection in Internet of Things using feature selection and classification based on Logistic Regression and Artificial Neural Network on N-BaIoT dataset," 2021 5th International Conference on Internet of Things and Applications (IoT), 2021, pp. 1-7, doi: 10.1109/IoT52625.2021.9469605.

[205] U. Garg, V. Kaushik, A. Panwar and N. Gupta, "Analysis of Machine Learning Algorithms for IoT Botnet," 2021 2nd International Conference for Emerging Technology (INCET), 2021, pp. 1-5, doi: 10.1109/INCET51464.2021.9456246.

[206] T. N. Nguyen, Q. -D. Ngo, H. -T. Nguyen and G. L. Nguyen, "An Advanced Computing Approach for IoT-Botnet Detection in Industrial Internet of Things," in IEEE Transactions on Industrial Informatics, vol. 18, no. 11, pp. 8298-8306, Nov. 2022, doi: 10.1109/TII.2022.3152814.

[207] Q. -D. Ngo, H. -T. Nguyen, V. -D. Nguyen, C. -M. Dinh, A. -T. Phung and Q. -T. Bui, "Adversarial Attack and Defense on Graph-based IoT Botnet Detection Approach," 2021 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), 2021, pp. 1-6, doi: 10.1109/ICECCE52056.2021.9514255.

[208] "Cyber crime," FBI. [Online]. Available: https://www.fbi.gov/investigate/cyber#Respond-and%20Report. [Accessed: 04-May-2022].

[209] T. Liggett, ''Evolution of endpoint detection and response platforms,'' Ph.D. dissertation, School Arts, Utica College, Utica, NY, USA, 2018.

[210] F. Liang, W. G. Hatcher, G. Xu, J. Nguyen, W. Liao, and W. Yu, ''Towards online deep learning-based energy forecasting,'' in Proc. 28th Int. Conf. Comput. Commun. Netw. (ICCCN), 2019, pp. 1–9.

[211] Y. Cui, W. He, C. Ni, C. Guo, and Z. Liu, ''Energy-efficient resource allocation for cache-assisted mobile edge computing,'' in Proc. IEEE 42nd Conf. Local Comput. Netw. (LCN), Oct. 2017, pp. 640–648

[212] W. Liao, C. Luo, S. Salinas, and P. Li, ''Efficient secure outsourcing of large-scale convex separable programming for big data,'' IEEE Trans. Big Data, vol. 5, no. 3, pp. 368–378, Sep. 2019

[213] G. Lin,W. Xiao, J. Zhang, and Y. Xiang, "Deep learning-based vulnerable function detection: A benchmark," in Proc. Int. Conf. Inf. Commun. Secur. Cham, Switzerland: Springer, 2019, pp. 219–232.

[214] E. Fernandes, A. Rahmati, K. Eykholt, and A. Prakash, "Internet of Things security research: A rehash of old ideas or new intellectual challenges?" IEEE Security Privacy, vol. 15, no. 4, pp. 79–84, Aug. 2017.

[215] Z. Zhou, X. Chen, E. Li, L. Zeng, K. Luo, and J. Zhang, ``Edge intelligence: Paving the last mile of articial intelligence with edge computing," Proc. IEEE, vol. 107, no. 8, pp. 17381762, Aug. 2019.

[216] J. Sengupta, R.Kubendran, E. Neftci, and A. Andreou, ``High-speed, realtime, spike-based object tracking and path prediction on Google edge TPU," in Proc. 2nd IEEE Int. Conf. Artif. Intell. Circuits Syst. (AICAS), Aug. 2020, pp. 134135.

[217] M. A. Al-Garadi, A. Mohamed, A. Al-Ali, X. Du, I. Ali, and M. Guizani, ``A survey of machine and deep learning methods for Internet of Things (IoT) security," IEEE Commun. Surveys Tuts., early access, Apr. 20, 2020, doi: 10.1109/COMST.2020.2988293.

[218] P. Shukla, "ML-IDS: A machine learning approach to detect wormhole attacks in Internet of

Things," in Proc. Intell. Syst. Conf. (IntelliSys), Sep. 2017, pp. 234–240

[219] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," in Proc. Int. Conf. Inf. Syst. Secur. Privacy, 2018, pp. 108–116.

[220] J. A. Harer et al., "Automated software vulnerability detection with machine learning," 2018, arXiv:1803.04497. [Online]. Available: http://arxiv.org/abs/1803.04497

[221] R. Russell et al., "Automated vulnerability detection in source code using deep representation learning," in Proc. 17th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA), Dec. 2018, pp. 757–762.

[222] Z. Li et al., "Vuldeepecker: A deep learning-based system for vulnerability detection," in Proc. NDSS, 2018, pp. 1–15.

[223] G. Lin et al., "Cross-project transfer representation learning for vulnerable function discovery," IEEE Trans. Ind. Informat., vol. 14, no. 7, pp. 3289–3297, Jul. 2018.

[224] M. Allamanis, E. T. Barr, P. Devanbu, and C. Sutton, "A survey of machine learning for big code and naturalness," ACM Comput. Surv., vol. 51, no. 4, p. 81, 2018.

[225] Z. A. El Houda, B. Brik and S. -M. Senouci, "A Novel IoT-Based Explainable Deep Learning Framework for Intrusion Detection Systems," in IEEE Internet of Things Magazine, vol. 5, no. 2, pp. 20-23, June 2022, doi: 10.1109/IOTM.005.2200028.

[226] M. Abrishami et al., "Classification and Analysis of Adversarial Machine Learning Attacks in IoT: a Label Flipping Attack Case Study," 2022 32nd Conference of Open Innovations Association (FRUCT), Tampere, Finland, 2022, pp. 3-14, doi: 10.23919/FRUCT56874.2022.9953823.

[227] S. Laazizi, J. Ben Azzouz and A. Jemai, "cybclass: classification approach for cybersecurity in industry 4.0," 2022 IEEE 9th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT), Hammamet, Tunisia, 2022, pp. 378-384, doi: 10.1109/SETIT54465.2022.9875643.

[228] K. Singh, S. S. Grover and R. K. Kumar, "Cyber Security Vulnerability Detection Using Natural Language Processing," 2022 IEEE World AI IoT Congress (AIIoT), Seattle, WA, USA, 2022, pp. 174-178, doi: 10.1109/AIIoT54504.2022.9817336.

[229] S. T. Mehedi, A. Anwar, Z. Rahman, K. Ahmed and R. Islam, "Dependable Intrusion Detection System for IoT: A Deep Transfer Learning Based Approach," in IEEE Transactions on Industrial Informatics, vol. 19, no. 1, pp. 1006-1017, Jan. 2023, doi: 10.1109/TII.2022.3164770.

[230] Z. Anastasakis et al., "Enhancing Cyber Security in IoT Systems using FL-based IDS with Differential Privacy," 2022 Global Information Infrastructure and Networking Symposium (GIIS), Argostoli, Greece, 2022, pp. 30-34, doi: 10.1109/GIIS56506.2022.9936912.

[231] K. Ibrahimi and H. Benaddi, "Improving the IDS for BoT-IoT Dataset-Based Machine Learning Classifiers," 2022 5th International Conference on Advanced Communication Technologies and Networking (CommNet), Marrakech, Morocco, 2022, pp. 1-6, doi: 10.1109/CommNet56067.2022.9993869.

[232] M. Venkatasubramanian, A. H. Lashkari and S. Hakak, "IoT Malware Analysis using Federated Learning: A Comprehensive Survey," in IEEE Access, doi: 10.1109/ACCESS.2023.3235389.

[233] O. Abdel Wahab, "Intrusion Detection in the IoT Under Data and Concept Drifts: Online Deep Learning Approach," in IEEE Internet of Things Journal, vol. 9, no. 20, pp. 19706-19716, 15 Oct.15, 2022, doi: 10.1109/JIOT.2022.3167005.

[234] N. Karmous, M. O. -E. Aoueileyine, M. Abdelkader and N. Youssef, "IoT Real-Time Attacks Classification Framework Using Machine Learning," 2022 IEEE Ninth International Conference on Communications and Networking (ComNet), Hammamet, Tunisia, 2022, pp. 1-5, doi: 10.1109/ComNet55492.2022.9998441.

[235] A. Ahmed and C. Tjortjis, "Machine Learning based IoT-BotNet Attack Detection Using Real-time Heterogeneous Data," 2022 International Conference on Electrical, Computer and Energy Technologies (ICECET), Prague, Czech Republic, 2022, pp. 1-6, doi: 10.1109/ICECET55527.2022.9872817.