IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 7 August 2024, accepted 21 August 2024, date of publication 26 August 2024, date of current version 6 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3450180

TOPICAL REVIEW

A Detailed Review on Enhancing the Security in Internet of Things-Based Smart City Environment Using Machine Learning Algorithms

ARUNKUMAR MUNISWAMY AND R. RATHI
School of Computer Science Engineering and Information Systems, Vellore Institute of Technology, Vellore, Tamil Nadu 632014, India

Corresponding author: R. Rathi (rathi.r@vit.ac.in)

This work was supported by the Vellore Institute of Technology, Vellore, Tamil Nadu, India.

ABSTRACT Over the past few years, smart cities have seamlessly integrated into our daily lives, offering convenience and simplicity. However, as these cities become increasingly interconnected and reliant on the Internet of Things (IoT), ensuring heightened security measures becomes paramount. The potential compromise of IoT devices due to vulnerabilities poses significant risks, including the theft of personal data, leading to severe hazards for individuals. Thus, Security plays a pivotal role in safeguarding IoT devices. In this modern era, integrating security measures with machine learning has emerged as a solution to automate and streamline security protocols. This requires a comprehensive analysis of enhancing security levels in IoT devices within innovative city environments. Our study extensively surveys security issues across various facets of IoT infrastructure, including hardware, cloud environments, applications, data, software, and networks. Through thorough examination, we identify the effects of these issues and propose countermeasures to bolster Security, mainly focusing on IoT devices. Furthermore, our study delves into various machine learning algorithms, providing examples, detailing attack types, and assessing accuracy rates for each algorithm. We offer a quick reference guide that outlines the benefits and drawbacks of different machine-learning algorithms and their applications. Additionally, we aim to identify and mitigate various security threats by exploring diverse datasets, evaluation metrics, IoT threats, and machine-learning techniques. By thoroughly exploring these aspects, our study equips future researchers with the knowledge to effectively identify potential security threats and implement robust safeguards against them.

INDEX TERMS IoT, security, machine learning, smart city, attacks.

I. INTRODUCTION

IoT devices have sensors and actuators where the sensors are used to sense the particular environment and use the internet to transmit the sensed input to the decision-making processor, and in turn, it sends the specific signals to activate the actuators to take necessary actions in the environment, for example, the temperature sensor senses the temperature inside a particular room and sends the input to the processor, and it decides whether the room temperature to be increased or decreased, such that the processor will send back the signal to the actuators to cool or heat that room. We utilize billions of IoT devices in our everyday lives, Global security and risk management spending is expected to increase by 14.3% in 2024, outpacing IT spending at 8%, according to Gartner. Fortune Business Insights predicts the cybersecurity market will reach $424.97 billion by 2030, nearly 2.5 times its 2023 value [1].

The concept of an intelligent city environment includes elements like smart homes and smart Agriculture, as depicted in Figure 1. These factors significantly influence the use of Internet service providers, data storage, and access, consequently introducing the possibility of vulnerabilities and threats into the scenario. Nowadays, there are more security issues in any IoT device's hardware and software side due to the device manufacturers' lack of encryption

The associate editor coordinating the review of this manuscript and approving it for publication was Kashif Saleem.

and standardization. The software can easily be hacked by hackers using bots.

A smart city powered by IoT devices has been used by **Smart Buildings** – how to Secure the home as well as office buildings and also to automate all the work in and around it [2], **Smart Agriculture** – deals with how to maximize the profit in the cultivation of agricultural products with less investment and to improving the utilization natural farming instead by avoid using chemical fertilizers or pesticides [3]

Smart Manufacturing – to automate, and speed up the process and product workflow in manufacturing industries [4], **Smart Education** – to enhance the process of delivering the classes in digital forms such as audio and video and to increase the teaching and learning process between a teacher and a student [5], **Smart Environment** – how to safeguard our environment by making a green environment and encouraging about waste management, recycling of wastes, air quality management, pollution management, etc. [6]

Smart Grid – dealing the way of the utilization of energy resources for power such that proper generation, transmission, distribution from various kinds of power resources(wind, water, air, atomic) [7], **Smart Healthcare** – to make the physicians work smarter and lesser the intelligent devices and wearables are used to monitor the patient condition all the time, thereby Robots are also used for diagnosis and to do critical tasks to make more affordable in the Healthcare sectors, [8], **Smart Transportation's** – to have a more brilliant way of mobility, transportation, parking garages, electric vehicle charging stations, delivery of goods, self-driving cars and more [9].

The table labeled Table 1 compiles abbreviations alongside their respective definitions.

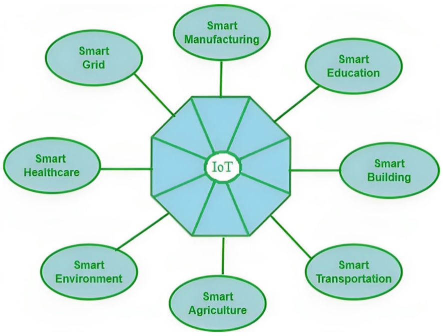

FIGURE 1. Illustration of smart city environment with internet of things.

A. PROCESS OF SLR

This study examines the present research on enhancing Security within the Internet of Things (IoT) domain. Employing a systematic mapping study, we scrutinize the existing literature. The investigation aims to address the following five Research Questions (RQs):

- RQ1: What are the areas to concentrate on in IoT Security? The answer to the above question is given in Section II-A through II-F
- RQ2: What are the different types of attacks and their corresponding countermeasures? This above question is answered in Section III-A through III-F
- RQ3: In what ways is ML influencing Security in IoT? We have the solution to the above question in Section IV and V
- RQ4: How can you design an IoT system securely? The answer to the above question is given in Section VI-A through VI-D

Figure 2 illustrates the findings and selection of articles. Initially, we searched based on keywords, screening approximately 986 articles. After applying our selection criteria, we included 198 articles. Among these, we excluded survey articles, leaving 137 for further consideration. We then thoroughly examined the title, abstract, and full text of 96 articles. Subsequently, we chose 59 articles for further Analysis based on the publisher's quality, as shown in Figure 3.

Figure 3 represents the selection of articles based on the publishers.

1) ORGANIZATION OF THIS SURVEY

The following is the outline for this paper. Section I introduces a 'smart city' concept and the Research Questions to be Solved. Section II provides an overview of Essential Focus Areas for IoT Security and different possible attacks. Section III explains various attacks in IoT and discusses their countermeasures. Section IV discusses the need for a Machine Learning Algorithm in IoT security. Section V represents a literature survey on various types of Machine Learning algorithms. Section VI addresses Security Issues with the Internet of Things. Section VII explores Secure Designs with problems and Solutions, and Section VIII delves into Future Scope and Research Directions, concluding with final remarks and References.

II. SIX ESSENTIAL FOCUS AREAS FOR IoT SECURITY

The significant studies' tiered architectures had anywhere from three to seven tiers depending on how comprehensive a given study was. These tiers housed the various components of the different IoT platform types. Researchers introduced the three-layer architecture as the first IoT paradigm [10]. The sensing layer is the foundation and comprises sensors and controllers as objects, the cloud layer handles data storage and processing, and the application layer facilitates user participation. This three-layer structure is expanded to a four-layer form by adding a company component [11].

Another description of IoT, which has a middleware-based, five-layer structure [12]. Composing services, managing those services, and abstracting them from the rest of the system are all middleware-based activities. In addition to the five-based model, which already includes the edge and mixed edge cre, the six-layer adds a fog layer or a gateway layer [13]. Cisco's latest suggestion for the Internet of Things'

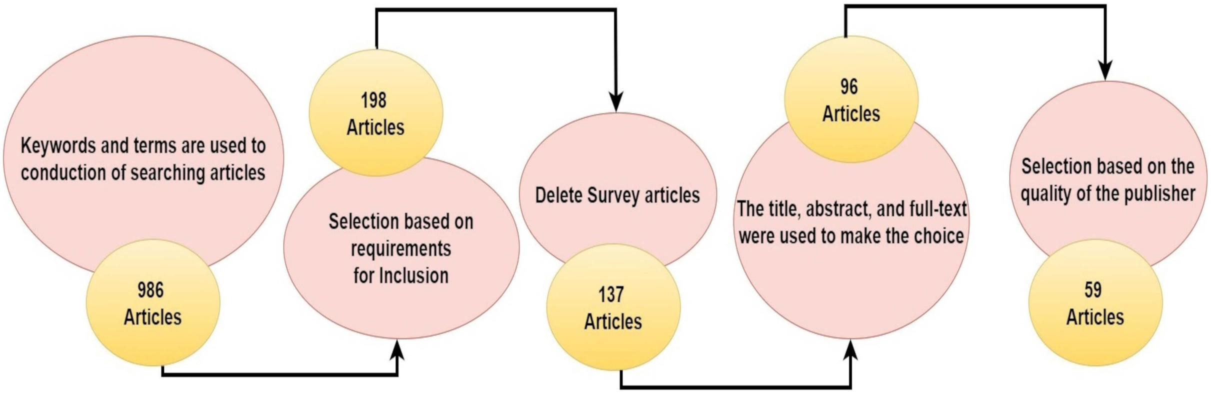
Keywords and terms are used to conduction of searching articles
986 Articles
198 Articles
Selection based on requirements for Inclusion
Delete Survey articles
137 Articles
96 Articles
The title, abstract, and full-text were used to make the choice
Selection based on the quality of the publisher
59 Articles

FIGURE 2. The steps involved in discovering and selecting articles.

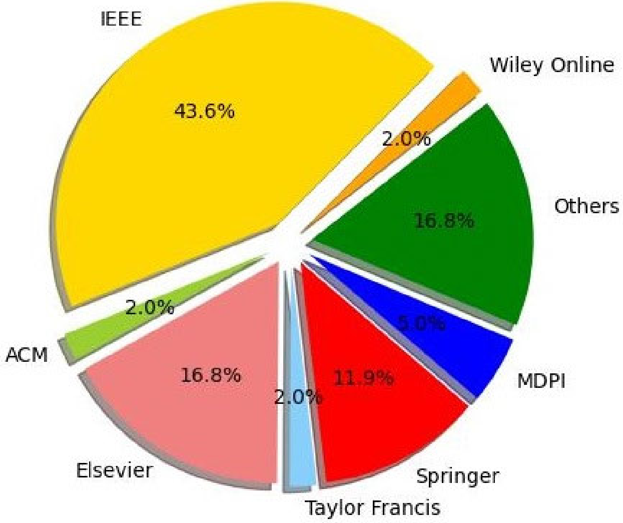

FIGURE 3. The arrangement of the chosen articles based on the publishers.

underlying infrastructure is a seven-tiered plan. Coatings for the end user and processes and a layer for peripheral processing were added to the prior design [14]. With the increasing intricacy of IoT systems and the general direction of research in this field, we have chosen to concentrate on the seven-layer design paradigm. We have decided to focus on this investigation's 3, 5, and 7-layer design types. Figure 4 illustrates how these seven levels are translated into six critical regions for attention.

As mentioned above, the IoT might have avoided many of these assaults if it had protected all 6 of the following emphasis areas.

The following are the primary areas where Security should be a top priority while developing your IoT solutions which is shown in Figure 5.

A. DEVICE/HARDWARE SECURITY

Due to computing and power limitations, security alternatives. Because many IoT installations are widely distributed, these constraints allow attackers to easily locate a network access point through methods such as brute force, Rowhammer, fuzzing, or side-channel attacks. For sensors and gateways, organizations must provide an outer layer of protection. They might begin by actively looking for or building devices with built-in security features. Tamper-resistant casings and gadgets disabling when tampered with are examples of security methods. IoT devices need hardware-level Security since software-based security solutions can easily miss assaults on this level. Software security cannot stop assaults on its own [15].

B. CLOUD SECURITY

IT administrators should look at gadgets when securing the Internet of Things. IoT gadgets, including temperature sensors, security cameras, and wearable medical equipment, come in various shapes and sizes. IoT device security involves additional specialized best practices, such as device detection and segmentation, while hardware security also includes measures to safeguard specific devices [15]. Every organization must start with device detection. IT managers will not take other security measures, like changing a default password, rolling out an update, or turning off idle devices, if they are unaware that a device even exists [15].

C. APPLICATION SECURITY

Some vulnerabilities in IoT apps that hackers exploit include insecure network links, open data storing, outdated IoT application segments, weak passwords, and inadequate upgrade procedures. IT managers must prepare for basic attacks like spoofing and privilege escalation [15].

Applications may be well-protected using industry-standard best security practices, such as frequent software upgrades, firewalls, and access permission. IT administrators must secure all the connections between devices and apps, API integrations, and surrounding technologies. However, even if everything is safe, they must watch for dangers and unexpected behavior in IoT programs [15].

TABLE 1. Table of abbreviations used.

Abbreviation	Definition	Abbreviation	Definition
ABE	Attribute-Based Encryption	MDP	Markov Decision Process
AI	Artificial Intelligence	ML	Machine learning
ANN	Artificial Neural Networks	MQTT	Message Queuing Telemetry Transport
APA	An Authentication Protocol to Prevent Phishing	NB	Naive Bayes
API	Application Programming Interface	NN	Neural Network
ASLR	Address Space Layout Randomization	NOS	Networked Smart Object
AUC	Area Under the Curve	PCA	Principle Component Analysis
BAN	Body Area Network	PDOS	Permanent Denial of Service
BIOS	Basic Input Output System	PUF	Physically Unclonable Function
BN	Batch Normalization	R2L	Root to Local
CNN	Convolution Neural Networks	RAM	Random Access Memory
CRC	Cyclic Redundancy Check	RBAC	Role Based Access Control
CUTE	Customizable and Trustable End device	ReLU	Rectified Linear unit
DBSCAN	Density-based spatial clustering of applications with noise	RF	Random Forest
DDoD	Dual Denial of Decision	RFID	Radio Frequency Identifier
DDoS	Distributed Denial of Service	RL	Reinforcement Learning
DOM	Document Object Model	RNN	Recurrent Neural Networks
DoS	Denial of Service	RQ	Research Question
DoS	Denial of Service	SAML	Security Assertion Markup Language
DPP	Dynamic Privacy Protection	SOAP	Simple Object Access Protocol
DQN	Deep Q Network	SQL	Structured Query Language
DS	Data Science	SSL	Secure Socket Layer
DT	Decision Tree	SVM	Support Vector Machine
EDoS	Economic Denial of Sustainability	SYN	Synchronize Sequence Number
EL	Ensemble Learning	TCP	Transmission Control Protocol
GMM	Gaussian mixture models	TD	Temporal Difference
HER	Electronic Health Record	TLS	Transport Layer Security
HLS	High-Level Synthesis	U2R	User to Root
HTML	Hyper Text Markup Language	USB	Universal Serial Bus
IoT	Internet of Things	WAFs	Web Application Firewalls
ISDD	Improved Secure Directed Diffusion	WS	Web Services
IT	Information Technology	WSN	Wireless Sensor Networks
KNN	K-Nearest Neighbours	XGBOOST	Extreme Gradient Boosting
LDA	Linear Discriminant Analysis	XML	Extensible Markup Language
LSTM	Long Short Time Memory	XSS	Cross-site scripting

D. DATA SECURITY

The business insights provided by the data gathered make IoT valuable. The information can support procedures or guarantee a patient's well-being and safety. However, data security is among the most challenging topics for many firms. IoT implementations involve the continuous exchange of enormous amounts of data. Multiple protection layers must cooperate to safeguard users' data privacy and ensure the uninterrupted operation of IoT devices. Additionally, businesses must choose where to save and arrange their IoT data. Legislative agencies have started to broaden the data protection and privacy rules that firms must abide by in addition to practical factors [15].

IT managers may take the first measures to restrict access by frequently upgrading all devices and resetting all passwords. SSL and other IoT data encryption technologies guarantee that data will not be intercepted. To safeguard the machines, they must also set up firewalls and monitor how users and software programs handle critical information [15].

E. SOFTWARE SECURITY

The IoT industry needs to catch up regarding security norms and legislation compared to other technology areas. Because it would take more time and money to incorporate Security, IoT technology does not come with it by default. Organizations must include Security in all of their hardware and IoT software. IoT software developers need to carefully choose their platforms, languages, and tools due to security vulnerabilities present in various libraries and APIs. To expedite their work, IoT developers could use open-source software, but they must consider the offered assistance and if the community takes proactive action to fix problems. Access control measures and software vulnerability testing are essential components of software security [15].

F. NETWORK SECURITY

Last, the network is becoming the primary focus of IoT protection. After devices join the network, the network has access to all data and tasks. With this method, hackers can gain access to anything on the web. By utilizing the network,

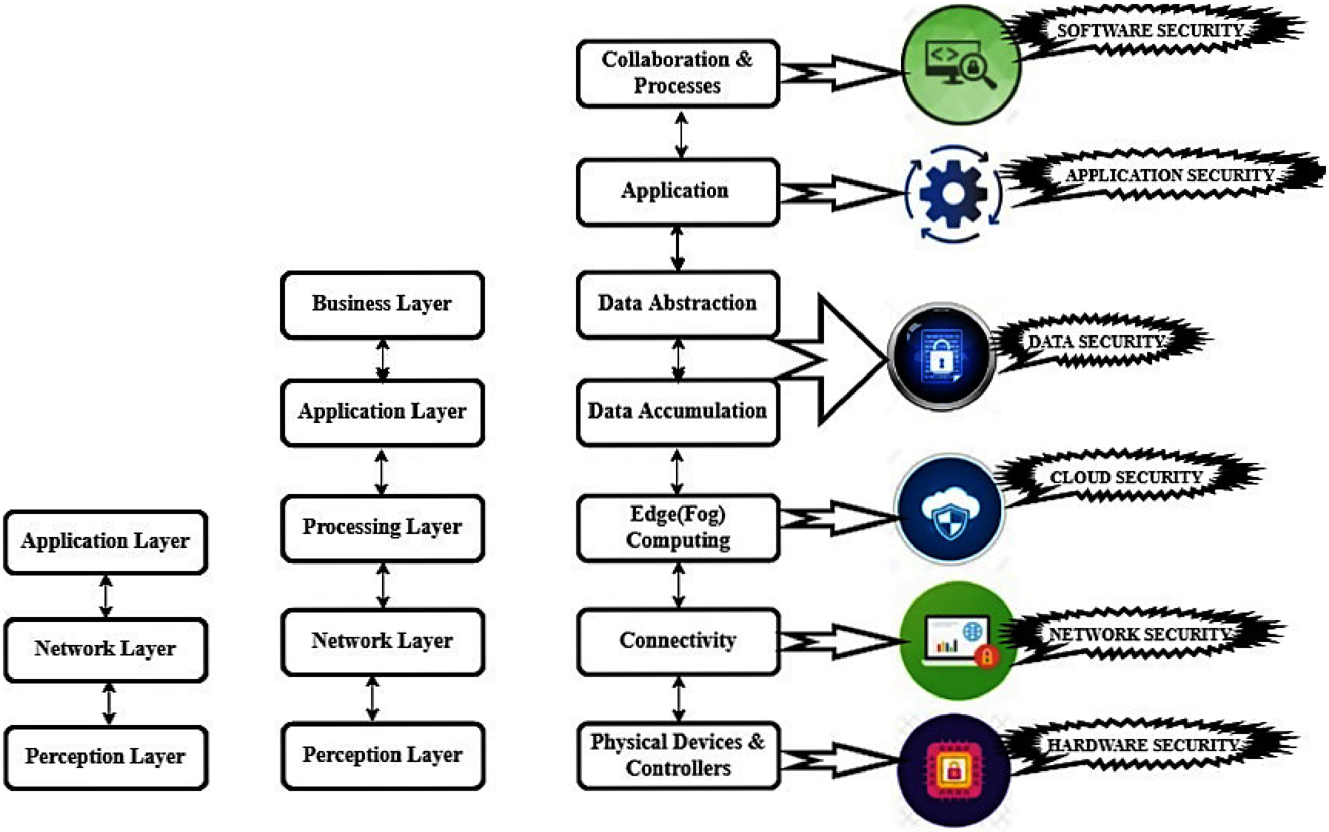

FIGURE 4. Illustration of how the different layers are translated into six key regions for security.

the Verification of users and devices can occur, enabling the establishment of guidelines and regulations to control access, behaviour, and the detection of anomalies [15]. The network is essential for spotting an attack as it happens and responding to the assault by shutting down, restricting, or slowing down the device. You can accomplish a lot with the network infrastructure. Most of the time, IT professionals will want to rely on network firewalls to solve security issues, but they only utilize one for some of the switch ports. The network itself has to have security features [15].

III. ATTACKS IN IoT

Various types of attacks are taking place in different areas mentioned above; we will discuss the attacks in particular domains of IoT as discussed below. Figure 6 summarizes the various attacks in IoT.

A. ATTACKS ON IoT DEVICE/HARDWARE

Here, we provide a comprehensive summary of these assaults. Physical attacks can be launched against network or system hardware if the intruder maintains proximity to it [16]. The following is a list of typical physically made Device or Hardware assaults or Attacks:

1) TAMPERING

The act of physically altering a gadget (such as an RFID) or communication connection [17].

2) MALICIOUS CODE INJECTION

By infecting a tangible object with harmful code, an assailant gains a foothold from which to launch further attacks [16] and [17].

3) RADIO FREQUENCY INTERFERENCE/JAMMING

An attacker produces and transmits noise signals across Wireless Sensor Network (WSN) and radio frequency (RF) signals to conduct denial-of-service (DoS) assaults on RFID tags in addition to sensor nodes and obstruct connectivity [18].

4) FAKE NODE INJECTION

To manipulate data flow between two genuine network nodes, an attacker inserts a fake node between them [19].

5) SLEEP DENIAL ATTACK

A Sleep Denial Attack targets battery-powered devices, like those in IoT, by continuously sending data or requests to keep them awake. This prevents low-power sleep mode, quickly draining the battery and reducing the device's lifespan [20].

TABLE 2. Device /Hardware threats, outcomes, and defenses.

Name of the Attack	Threats	Proposed Measures for Prevention	References
Malicious Code Injection and Tampering	The ability to obtain private data and launch denial-of-service attacks	An Authentication Method Relying on a Physically Unclonable Function (PUF), SVM.	[17]
Jamming/ RF Interference	DoS: Cause Conversation Problems	CUTE Mote, RF, SVM, RL, and K means are Used.	[18]
Fake Node Injection	Manage information traffic, act as an intermediary	Key Establishment and Pervasive Authentication in Wireless Sensor Networks for Distributed Internet of Things Applications (PAuthKey), SVM, RF, K-Means, and DB-SCAN are used.	[19]
Sleep Attack of Denial	The shutdown of a Node	Support Vector Machine (SVM), CUTE (customizable and trustable end device) Mote, SVM, RF, K-Means, and DB-SCAN are used.	[20]
Side Channel Attack	Gather the Decryption Codes	Authentication using PUF and Masking technique, SVM, PCA, LDA, RF, and CNN are used.	[21]
Permanent Denial of Service	Destruction of Resources	NetwOrked Smart object (NOS) Middleware, SVM, RF, K-Means, and DBSCAN are used.	[22]

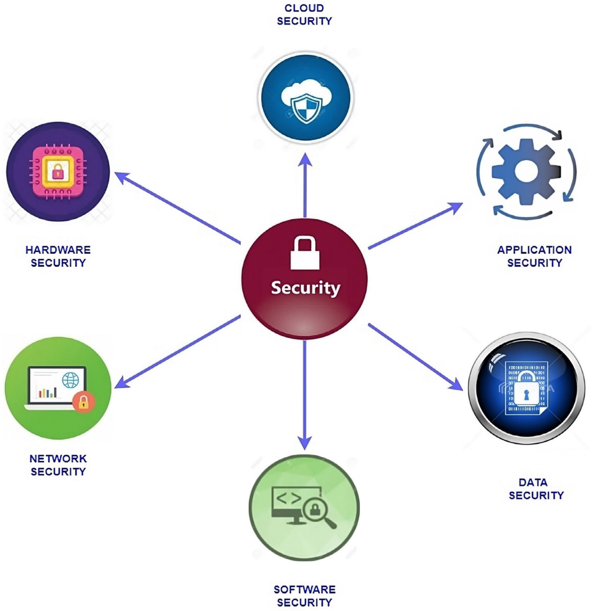

FIGURE 5. Six essential areas for IoT security.

6) SIDE-CHANNEL ATTACK

During this assault, the assailant gathers time, power, fault attacks, and other techniques on the system's hardware; one can decrypt keys [21]. These keys enable it to encrypt and decode sensitive data.

7) PERMANENT DENIAL OF SERVICE(PDoS)

Denial-of-service attacks like phishing entail using malicious software to wipe down an IoT device. The malware used to conduct the attack uploads corrupted BIOS or destroys firmware [23]. Table 2 summarizes the various types of attacks and their Effects and Countermeasures in IoT Devices/Hardware.

B. ATTACKS ON IoT CLOUD

1) FLOODING ATTACKS IN THE CLOUD

An SYN message overflow attack aims to exhaust a server's capacity by sending high volumes of SYN requests. Therefore, the server is unable to serve the actual users. SYN flooding attacks impact PaaS and IaaS layers [24]. Transmission Control Protocol (TCP) SYN messages link a legitimate client to a server. (TCP). The server then sends the authorized user a request to confirm receipt of the SYN communication, abbreviated as (SYN-ACK). The final step in establishing a link is for the authenticated user to submit an ACK message to the server. SYN overflow is inevitable if many messages are sent to the server from the assailants, but the three-way negotiation technique will fail. As a result, the cloud system's resource efficiency suffers as the computer waits for all those messages to complete. Two approaches are used in the PaaS layer to connect to a legitimate user request. The SYN cache mechanism is one, while the SYN cookies defense technique is another. Due to its 15% longer request-response time and subpar performance, neither is at a suitable level [24].

2) CLOUD MALWARE INJECTION ATTACK

The malware infiltration process damages cloud computing's credibility because it compromises the reliability of the extraordinary services for which the technology is known. Attackers try to introduce their malicious virtual machines, services, and applications into the cloud system; as a result, services are marked as genuine instances of harmful software. Cloud-based anti-malware services are available to all authorized users, including Trojan packages or virus programs that create hazardous zones in dispersed environments. Malware programs put system hardware and cloud instances in danger and make data unavailable throughout the cloud system [25]. There are two suggested actions to stop an assault using cloud malware injection. The first and most crucial stage is to forward all inbound queries to the service instance security check. The second stage is to implement a system that requires a correct hash value for each user.

3) SIGNATURE WRAPPING ATTACK

The Signature Wrapping Attack is implemented using SOAP-Simple Object Access Protocol, which underlies the XML signature attack. A signature element in the security

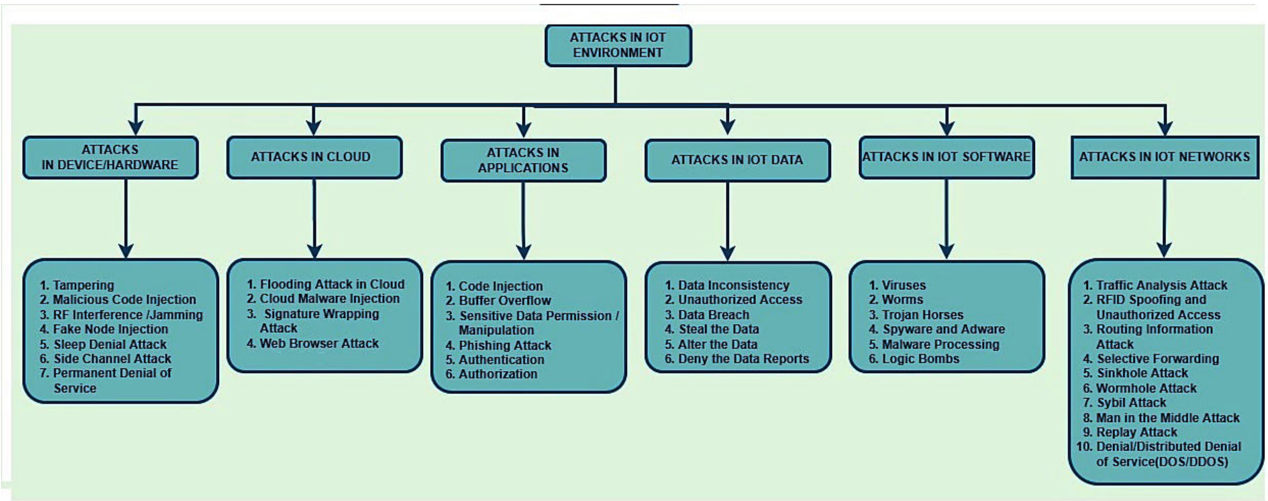

FIGURE 6. Illustration of various attacks in IoT environment.

header of a SOAP message addresses extra message sections. Since XML messages' IDs are used for validation, a user ID for the online application can be used to launch an assault on the data. The receiver of a SOAP message will typically check the signature's authenticity by matching the received ID to a predetermined set of numbers. An XML signature attack is launched against the user who fraudulently uses a copied SOAP message as a heading element in the request. The malicious attacker carried out every SOAP message control there. To prevent an XML Signature attack, one must adhere to the WS-Security policy [26].

4) WEB BROWSER ATTACK

Cross-site scripting adware inserts harmful code onto a legitimate website to initiate an attack on a user's browser. The most prevalent XSS attack is DOM-based, followed by non-persistent (or dynamic) XSS attacks and irreversible (or static) XSS attacks. Reflected XSS [27] includes DOM-based attacks as a subtype, the injected script is temporarily diverted away from the web server, and persistent type attack involves the lasting storage of an inserted script. Due to their dynamism, dynamic websites are more susceptible to being hurt by XSS attacks than static websites. When trusted users click pop-up windows on the screen, private information is accidentally disclosed. Several strategies are used to stop this threat, including technology for detecting vulnerabilities in web applications, content-based methods for preventing data leaks, and Active Content Filtering [27].

Table 3 summarizes the various types of attacks and their Effects and Countermeasures in the IoT Cloud Environment.

C. ATTACKS ON IoT APPLICATIONS

1) INJECTION

In this attack, harmful code is inserted into the system by exploiting programming vulnerabilities [25], [28]. Injecting malicious code into a software application has many uses, including data stealing, system takeover, and malware distribution [25], [28]. Standard methods of assault include script injection and HTML code injection. The system may become unstable, users' privacy may exposed, or the entire system may shut down if attacked in this way.

2) BUFFER OVERFLOW

In this type of assault, program flaws are exploited to overflow the memory holding the code or data. Many programs use a standard memory organization to keep blocks of code and data. The attacker floods a preset residence zone by writing a lengthy data stream to a designated place. The application's flow management can be disrupted, malevolent code can be executed (when the pattern crosses into a logical block of code.), and other data can be changed (where another data buffer's data section meets the series). Some typical methods include memory overflows on the heap or stack, arithmetic errors, double frees, and attacks based on format strings [29].

3) AUTHORIZATION TO MANIPULATE/ACCESS SENSITIVE INFORMATION

This assault compromises users' Security by allowing third parties to see and potentially change their data [30]. This attack often takes advantage of permission model design shortcomings [30]. Attackers can manipulate programs in intelligent homes by taking advantage of flaws in the authorization paradigm, leading to theft and break-ins [30]. In addition, earlier research [30] examined the events involved in communication between Smart Apps and Smart Devices. Keep in mind that smart devices and intelligent apps provide a particularly challenging issue for data security.

4) PHISHING ATTACKS

An intruder may masquerade as a user or a genuine organization to access sensitive material [31]. Email is a typical target for this attack, and when recipients open the email, the attacker gets access to essential data.

TABLE 3. Cloud threats, outcomes, and defences.

Name of the Attack	Threats	Proposed Measures for Prevention	References
Flooding Attack	Performance and Quality of Service on the Cloud	By setting up a firewall or intrusion detection system (IDS), you may prevent attacks on your server from unauthorized users, SVM, K-Means, CNN, and DBSCAN are used.	[24]
Cloud Malware Injection Attacks	Stealing Data or Eavesdropping.	To verify the authenticity of incoming requests to a service instance using their hash values, SVM, RF, XGBOOST, and CNN are used.	[25]
Signature Wrapping Attacks	Bypassing signature-based security measures, Tampering with message integrity, and Gaining unauthorized access.	Anomaly Detection can implemented by SVM, Behavioral Analysis by RNN, and Deep Packet Inspection by Using CNN.	[26]
Browser Security Attack	Identity theft	(Transport Layer Security) TLS with x509 certificate. Public key with SAML (Security Assertion Markup Language). A server's public key is used. Legally Binding Session. SVM, NB, SVM, RF, NN, and DBSCAN are used.	[27]

5) AUTHENTICATION AND AUTHORIZATION

The preservation of IoT security and privacy depends heavily on the authentication procedure. Fine-grained Verification is impossible with the current authentication techniques [32]. Malicious files, for instance, can be obtained alongside program updates, allowing attackers to exert direct control over a device [32]. However, there are problems with the authorization paradigm as well. Over-privilege is a common problem [32] because it enables a device to obtain data without utilizing all permissions to be granted. Additionally, the permission issue results from utilizing the default setup. When a file or location is given the wrong rights, an intruder can exploit it in several ways [32]. The smart card is vulnerable to distant verification attacks in a specific application, which might result in user information leakage and manipulation [32]. Additionally, an attacker can carry out unlawful actions, like opening the door, because the smart home lacks a complete authentication method [33]. Table 4 summarizes the various types of attacks and their Effects and Countermeasures in IoT Applications.

D. ATTACKS ON IoT DATA

The computational resources required to sustain the degree of connection and data collecting that IoT's growth and development in the manufacturing sector are putting pressure on the market for connected products [34]. Cloud computing entered the scene, serving as the framework for all the IoT offers. Cloud computing simplifies the deployment of virtual servers, the launch of a database instance, and the construction of data channels to support the functioning of IoT systems [34]. Cloud services may also be helpful in this context due to the importance placed on data protection by providing suitable login methods, hardware and software upgrade processes, etc. In this article, we discuss the most severe recent data leaks in the IoT industry:

1) DATA INCONSISTENCY

Attacks on data honesty, also called Data Inconsistency, can lead to problems with widely kept or in-transit data in the IoT; erroneous Analysis and decisions made due to inconsistent data might lead to system failure or unexpected results [35].

2) UNAUTHORIZED ACCESS

Ensuring that only genuine users are granted access and that malicious users are prevented from gaining data proprietorship or access to sensitive information is crucial, as regulated entry aims to prevent unauthorized users [36].

3) DATA BREACH

The revealing of data, also known as memory leaking, is the illegal use of private, delicate, or secret information [37].

4) STEAL THE DATA

Data theft is the unauthorized or intentional acquisition of any data not meant for sharing. Data access has been more widely available in recent years, which has led to a rise in data theft. Data may stolen from business databases, desktop computers, mobile phones, portable devices, flash drives, and cameras [38].

5) ALTER THE DATA

Any circumstance has the potential for data interception and tampering. The basis of digital messaging is the transmission of data. Unencrypted data transmissions are particularly vulnerable because hackers can easily tamper with them and redirect them to different destinations. The system program could have a security vulnerability while the data is at rest, allowing an intruder to damage the data or the base computer code with harmful code [38].

6) DENY THE DATA REPORTS

The report's Security is robust; it allows or disallows precise access to data pieces in a package, data source, or even row-level access in a database. The ability to run a report is often granted to specific Groups or Roles using Read and Execute permissions. However, you may give or limit access to a message so we can use it at particular times of the day [38].

TABLE 4. IoT Application threats, outcomes, and defenses.

Name of the Attack	Threats	Proposed Measures for Prevention	References
Code Injection	Inserting Code Insert harmful code into a software program or HTML document.	Black box testing uses web crawlers to establish the upper bounds of SQL speed and Web Application Firewalls (WAFs) is used to track how the application reacts to these limits, DT and SVM are used.	[25] and [28]
Buffer Overflow	Buffer Overflow Overwrite an application's memory	ASLR randomizes data region address spaces. Randomizing address spaces makes buffer overflow attacks unfeasible; SVM, DT, RF, and NN are used.	[29]
Authorization to Manipulate / Access Sensitive Information	Unauthorized manipulation or sensitive data access.	Utilize DT or RF in Machine Learning Access Control for adaptive authorization based on contextual data.	[30]
Phishing Attacks	Concealment of the attacker's identity to take sensitive information via "phishing."	An Authentication Protocol to Prevent Phishing (APA): It employs a zero-knowledge password proof and a two-factor authentication system. NB, SVM, RF, and XG-BOOST are used.	[31]
Authentication and Authorization	Unauthorized access to IoT devices, Identity spoofing, Unauthorized data access.	Implement Face Recognition for Bio-metric Authentication, utilize Reinforcement Learning for Behavioral Analysis, and employ DT or RF in Machine Learning Access Control for adaptive authorization based on contextual data.	[32]

Table 5 summarizes the various types of attacks and their Effects and Countermeasures in IoT Data.

E. ATTACKS ON IoT SOFTWARE

Below are some examples of software assaults that an attacker may conduct by taking advantage of a system's related software or security flaws:

1) VIRUSES

An attacker could compromise the system with this malicious software to perform unauthorized changes, theft, and a denial-of-service assault [45].

2) WORMS

According to [46], White worms are self-propagating programs to safeguard and guard IoT devices. They do this by utilizing the dissemination and infection mechanism of malevolent botnets. White worms, however, inherit specific problems from malevolent botnets since they resemble them in some ways.

3) TROJAN HORSES

The Internet of Things (IoT) provides a new arena in which malware can be used to construct formidable botnets. Mirai, a novel Linux Trojan attack, is elusive and widespread. The danger is a new variant of Gafgyt, also called BASHLITE or Torlus, the standard software used by DDoS service providers [47] and [48].

4) SPYWARE AND ADWARE

A type of virus called spyware includes secretly watching the person actively using a device. The network makes it possible for harmful operations, including surveillance, keystroke collection, data harvesting of account credentials and banking and credit card information, and monitoring of keystrokes. It could also corrupt a device's four software security settings. It can exploit vulnerabilities in open software and spread itself by attaching itself to various software [49] and [50].

Adware is a malicious program designed to display advertisements on your computer, often through an Internet browser. Many security experts consider it the precursor to today's potentially harmful software. It usually employs a dishonest strategy to either pass for legitimacy or disguise another piece of software to trick users into installing it on their computer, laptop, or smartphone. Malware is a nightmare for any contemporary enterprise. New harmful software is constantly accessible by attackers and cyber-criminals to assault their targets. Despite their most significant efforts, security firms must find millions of new malware each month to fully defend themselves against malware attacks [49] and [50].

5) MALWARE

Malware could infect the cloud or data centers if it gains access to information on the Internet of Things devices, as stated by [45].

6) LOGIC BOMBS

Several security professionals agree that a logic bomb is a deliberately implanted bit of code in a program that goes off only when a predetermined set of conditions is met. If a coder worries about being dismissed from their job, they might hide some code that deletes all of the company's data. (a database payment prompt is one such example.) [51].

Table 6 summarizes the various types of attacks and their Effects and Countermeasures in IoT Software.

F. ATTACKS ON IoT NETWORK

Creating harm to the network attacks manipulates IoT network infrastructure. Without being near the network,

TABLE 5. IoT Data threats, outcomes, and defences.

Name of the Attack	Threats	Proposed Measures for Prevention	References
Data Inconsistency	Discordance in the Data	Blockchain architecture, Chaos-based scheme, k-means or hierarchical Clustering, KNN is used	[39], and [40].
Unauthorized Access	There has been a breach of data privacy.	Blockchain-based Attribute-Based Encryption(ABE), ABE that Protects User Privacy, SVM, RF, and DT are used.	[41] and, [36].
Data Breach	Information Loss	Dynamic Privacy Protection (DPP), Improved Secure Directed Diffusion (ISDD), Two Factor Authentication, CNN, K-Means, DBSCAN, and SVM are used.	[42], [43], and [44].
Stealing the Data	Unauthorized data theft.	Implement anomaly detection using Isolation Forest or One-Class SVM.	[38]
Altering the Data	Unauthorized manipulation of data.	Employ anomaly detection models like Autoencoders to identify data anomalies.	[38]
Deny for Data Reports	Blocking or denying access to data reports.	Use machine learning-based access control, such as RBAC, to enforce policies.	[38]

TABLE 6. Software threats, outcomes, and defences.

Name of the Attack	Threats	Proposed Measures for Prevention	References
Trojan Horses, Worms, Viruses, Adware and Spyware.	Destruction of resources	High-Level Synthesis (HLS), and Lightweight framework, SVM, CNN, RF, NB, and DT are used.	[47] and [48].
Malware	Infection of Data	Neural Networks with Lightweight Framework and Classification of Malware Images, SVM, CNN, RF, NB, DT, and XG-BOOST are used.	[49] and [50].
Logic Bombs	Malicious code are triggered under specific conditions	Use unsupervised anomaly detection to identify potential logic bombs.	[51]

it may deployed with ease. The following list of network assaults' most prevalent types:

1) TRAFFIC ANALYSIS ATTACK

Without physically accessing the network, an adversary can still collect sensitive information by ''sniffing'' data in transit between and among the devices [52].

2) RFID SPOOFING

An RFID signal must be faked first by the invader [53] to gain access to the RFID tag's data. Using the initial tag ID, the intruder can transmit data while making it look like it came from a legitimate source.

3) RFID UNAUTHORIZED ACCESS

Unauthorized access to RFID transpires when an individual gains entry to the system without proper authorization, enabling them to steal sensitive information or manipulate the data. The scenario above may compromise sensitive data, fraudulent use of personal identity, or depletion of financial resources [53].

4) ROUTING INFORMATION ASSAULTS

Active assaults include things such as establishing up routing wraps and sending out messages of error [54]. The perpetrator is attempting to cause disturbance by forging or modifying routing data.

5) SELECTIVE FORWARDING

In the event of an attack, it is possible for a non-compliant node to selectively alter, exclude, or solely transmit certain messages to additional nodes within the network [55]. Consequently, the transmission of information to its intended recipient needs to be improved.

6) SINKHOLE ASSAULT

The attack in uncertainty involves compromising a node situated closer to the sink, commonly known as a ''sinkhole node.'' This node makes it seem more appealing than others in the network, so traffic would flow to it instead of elsewhere [56].

7) WORMHOLE ATTACK

The perpetrator establishes the conduit by stealing data from one part of the network and sending it via another means to another part of the network. Using a hacked device or setting up an unauthorized access point are two standard methods of achieving this [57].

8) SYBIL ATTACK

One malevolent node spreads itself across the network by adopting multiple names (termed Sybil nodes) [58]. As a result, many valuable resources are divided unfairly.

9) MAN IN THE MIDDLE ATTACK (MITM)

One who opposes listens in on or monitors a discussion between the two IoT gadgets to eavesdrop on private data [59].

10) REPLAY ATTACK

An adversary can repeatedly send a verified message to the target after capturing it, as stated by [60]. A DoS attack could happen if network traffic remains high.

11) DENIAL OF SERVICE/DISTRIBUTED DENIAL OF SERVICE (DOS/DDOS) ATTACKS

DDoS attacks, which differ from DoS attacks, involve many corrupted nodes sending information or requests for connection to a single target to overwhelm it and make it unavailable or collapse [61].

Table 7 summarizes the various types of attacks and their Effects and Countermeasures in IoT Networks.

Table 8 summarizes the various types of attacks and specifies the type of machine learning applied to each.

IV. NEED FOR A MACHINE-LEARNING ALGORITHM IN IoT SECURITY

Figure 7 represents the current state and projected growth of the international Internet of Things industry graphically, which illustrates the need for IoT devices will increase along with threats and vulnerabilities will also increase in the future.

Machine learning (ML) can find and comprehend patterns in massive datasets, which is essential to the Security of the Internet of Things. Because the Internet of Things produces enormous volumes of data daily, machine learning algorithms may be trained to recognize emerging patterns and apply this understanding to identify and mitigate network hazards efficiently. These algorithms can identify typical activity patterns by examining data from registered IoT devices. Any departures from these patterns may interpreted as possible security risks. ML techniques, such as reinforcement learning, supervised, unsupervised, and semi-supervised techniques, are useful in identifying various threats, such as DDoS attacks, intrusions, and authentication attacks [63].

By employing machine learning, we can recognize and stop these threats before they do much damage. Because ML models can train continuously, their accuracy improves over time, which makes them crucial for tackling ongoing security issues in IoT networks. We'll examine several machine learning techniques, assess their efficacy, and describe how they improve IoT network security in the next section [63].

V. MACHINE LEARNING (ML)

Table 9 tabulates overall Machine Learning Design Approaches with different techniques.

Artificial intelligence plays a vital role in our day-to-day activities; machine-learning concepts are used to implement Security in an innovative city environment. The goal of machine learning is to enable the development of self-improving machines. Data science, at the confluence of computer science and statistics and serves as the foundation for (Artificial Intelligence) AI and (Data Science) DS, is one of the most quickly increasing fields of technology. Recent progress in machine learning can be attributed to increased accessible internet data, advancements in learning theory and algorithms, and the decreasing cost of computing power [65].

The Figure 8 quickly summarizes various types of Machine Learning Algorithms.

A. SUPERVISED

In supervised learning, machines are tasked with determining how inputs lead to desired outcomes by studying examples of such couples [66]. It uses diverse training instances and labelled data from training to deduce a function. Data is collected when determined to achieve specific goals or use a task-based approach [67]. The two most common guided tasks are classifying (which seeks to group the data) and prediction (which seeks to match the data). Text classification, in which a text snippet, like a short message or a product review, is analyzed to determine its class name or sentiment, is an application of supervised learning. All machine learning algorithms work on Classification, Clustering, Regression, and Dimensionality Reduction principles. Since classifying data includes the prediction of a name for a particular sample, it falls under the umbrella of guided learning in machine learning [66]. A goal, designation, or group (Y) is calculated analytically from an incoming variable (X). It can determine the type of the given data elements and whether they are organized or unstructured. Email service companies need help with a categorization problem when determining what should and should not marked as junk.

Mathematically, supervised learning is given in Equation(1): X is input, and Y is output. Training data is represented as

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n), \tag{1}$$

where x_i represents the i-th input and y_i represents the i-th output. We want to learn a function f(X) that translates inputs to outputs $f(x_i) = y_i$for all i. The genuine output is y, and the expected output is f(X). Minimize the expected and real output discrepancy to learn f(X). The loss function L(y, f(X)) is used to assess the difference between the real output and the anticipated output [68], [69], [70].

1) CLASSIFICATION TYPES

Binary Classification: Binary Classification: This term describes jobs that need the categorization of two classes, such as "true" and "false" or "yes" and "no" [66].

a: NAIVE BAYES(NB)

The theorem of Bayes is the basis for the naïve Bayes algorithm, which further asserts the independence of each set of traits [71]. In many actual circumstances, such as spam filtering, document or text categorization, etc., It is helpful for binary and multiclass classification tasks and performs well. Anomaly detection and Intrusion detection on the network layer [72], [73], [74] are two of the most prevalent examples of NB in the IoT. The most straightforward equation notation

TABLE 7. Network threats, outcomes, and defenses.

Name of the Attack	Threats	Proposed Measures for Prevention	References
Traffic Analysis Attack	Leakage of Data (Information in Network)	Oblivious communication structure that protects user privacy, RF, SVM, KNN, and CNN are used.	[52]
Unauthorized Access and RFID Spoofing.	Modification and Data Manipulation (Write, Delete, Read)	Personal User Function (PUF) with Static RAM, DT, KNN, RF, and SVM are used.	[53]
Routing Information Attacks	Repetitive routes	Verification using a Hash Chain's Integrity, DT, KNN, RF, and SVM are used.	[54]
Selective Forwarding	Destructive Messages	Verification using a Hash Chain's Integrity, a Method dependent on using monitors, RF, SVM, DT, and XGBOOST are used. RF, SVM, DT, and XGBOOST are used.	[55]
Sinkhole Attack	Data tampering or disclosure	Verification using a Hash Chain's Integrity, Detected Intrusions, RF, SVM, DT, and XGBOOST are used.	[56]
Wormhole Attack	Packet-based tunnelling.	Cluster Analysis for Intrusion Detection, RF, SVM, DT, and XGBOOST are used.	[57]
Sybil Attack	Poor resource distribution, duplication of effort	Informed-Trust Protocol, SVM, and RF are used.	[58]
Man in the Middle Attack	A Violation of Personal Space or Data.	Securing Message queuing telemetry support (MQTT), Authentication Between Devices, NB, SVM, and RF are used.	[59]
Replay Attack	DoS attacks, network bottlenecks	Signcryption (encrypting and digital signing capabilities), SVM, DT, and RF are used.	[60]
DDoS/DoS Attacks	Overloading and crashing of networks	Economic Denial of Sustainability (EDoS) Servers, IoT architecture built on software-defined networking, SVM, RF, KNN, and DT are used.	[62]

TABLE 8. Summary of area of attacks and machine learning algorithm applied.

Area of the Attack	Machine Learning Algorithm Applied
Hardware Attacks	Anomaly Detection Algorithms (e.g., Isolation Forest, One-Class SVM)
Cloud Attacks	Intrusion Detection Systems (e.g., Random Forest, SVM)
Application Attacks	Web Application Firewalls (e.g., Decision Trees, Deep Learning models)
Data Attacks	Data Loss Prevention Systems (Machine Learning-based anomaly detection)
Software Attacks	Behavioral Analysis Algorithms (e.g., Hidden Markov Models, LSTM)
Network Attacks	Network Intrusion Detection Systems (e.g., k-Nearest Neighbors, Deep Learning)

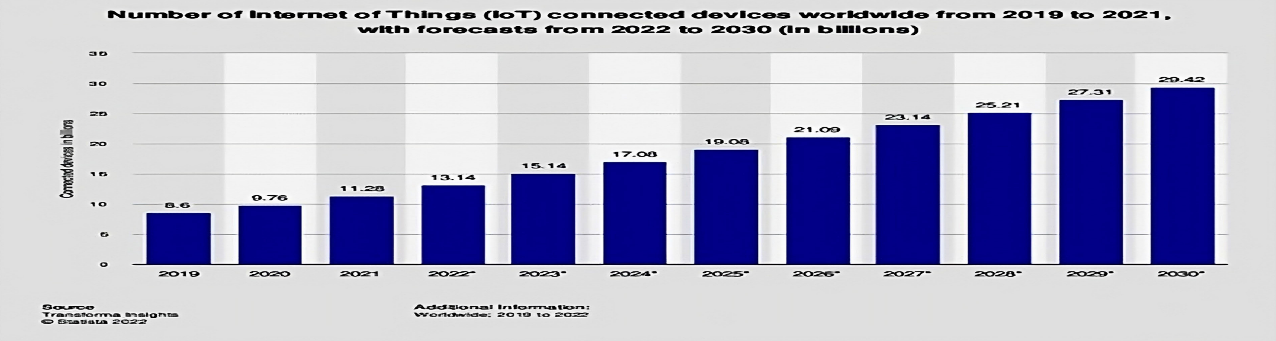

FIGURE 7. The current state and projected growth of the international Internet of Things industry are graphically represented and discussed [64].

TABLE 9. Machine learning design approach with techniques used.

Learning Method	How a Model is Designed?	Techniques Used
Supervised Learning	Algorithms, and models may be trained with the use of statistics with labels (task-oriented approach)	Classification, Regression
Unsupervised Learning	Learning models and algorithms may process data without labels (Data-oriented Approach)	Dimensionality reduction, Clustering, and associations
Reinforcement Learning	Motives such as reward and punishment form the basis of models (environment-driven approach)	Classification, control

for Naive Bayes is given in Equation(2):

$$[y = \arg\max_{c} P(c) \cdot P(x|c) \tag{2}$$

When a fresh data point is added, its expected category is denoted by y. c is the set of categories to which the fresh data point may be assigned. P(c) represents the expected frequency of class c. The probability that a new data element x conforms to category c is denoted by $P(x|c)$.

b: K-NEAREST NEIGHBOURS (KNN)

In certain instances, KNN is called a "lazy learning" method, [75] is an "instance-based learning" or "non-generalizing" learning technique. In place of a generic internal model, an n-dimensional space contains all instances similar to the training set. Using closeness measures, KNN classifies new data with the help of previously collected information. (Euclidean distance function is used) [71]. In the IoT, the KNN approach is used to detect intrusions, viruses, and anomalies. The KNN algorithm is simple, low-cost, and straightforward to implement [76]. Simple k-nearest neighbor's algorithm is given in Equation(3):

1. If we have a set of training data with labels, then

$$X = (x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n) \tag{3}$$

where x_i is an instance of data and y_i is a class name or goal variable.

2. For a new input data point x*, locate the k training instances from X that are closest in some distance measure.

3. Classify x* with the most common label from the KNN. Regression: The average value of the KNN target variable is denoted by x* [68], [69].

c: SUPPORT VECTOR MACHINE (SVM)

One of the popular techniques in machine learning is Support vector machines (SVMs) [77]. In addition to its usage in Regression and Classification, SVMs also have additional applications. An SVM may produce a single hyper-plane or a collection of hyper-planes in high or infinite multidimensional space. Due to the high accuracy levels, SVM has many useful security features; it can be put to good use in a wide variety of IoT security apps, including breach detection [78] [79], intelligent grid attacks [80], malware detection [81], and so on [82]. Simplified SVM algorithm equation: Input sample x_i is sample number i and y_i is the output label, with Let y_i be an element in the set $\{+1, -1\}$.

SVM locates a hyperplane that differentiates negative and positive data to the greatest extent possible. The hyperplane equation is given in Equation(4)

$$w.x + b = 0 \tag{4}$$

where b is the bias term, and w is the weight vector. SVM maximizes the margin between the hyperplane and the nearest positive and negative samples. The optimization problem for SVM is given in Equation(5):

$$y_i(w.x_i + b) \geq 1\text{for every } i \tag{5}$$

The constraint ensures that all samples are correctly classified with a margin of at least 1. The optimization issue may addressed using several approaches, such as quadratic programming or gradient descent, to achieve the optimum values of w and b that define the hyperplane [68], [69].

d: DECISION TREE (DT)

Among non-parametric-assisted techniques, the decision tree is widely used [83]. We turn to DT learning techniques to solve both problems—categorization and Regression [9]. In this context, the words "ID3" [84], "C4.5" [83], and "CART" [85] refer to well-known DT algorithms. DDoS and attack detection are two examples of security apps that use DTs as filters [86]. Decision trees are mathematical functions that translate an input feature vector X to an output label Y. The decision tree function is given in Equation(6)

$$Y = f(X) \tag{6}$$

Each decision rule in the function f corresponds to a tree node for a vector of input features and an output label. The algorithm picks a branch after testing one of X's characteristics. Each unit leads to a separate decision node until a leaf node with the output label is reached [87].

e: RANDOM FOREST (RF)

One popular ensemble categorization technique in many data science and machine learning fields is the random forest classifier [88]. Multiple decision tree classifiers, called "Parallel Ensembling," can be assembled concurrently. RF is frequently employed in the context of network surface attacks for the detection of Dual Denial of Decision (DDoD) attacks [89], abnormality detection [90], and the identity of unlawful Internet of Things (IoT) devices [91]. According to research conducted in the past [89], RF provides superior results than SVM, ANN, and KNN when identifying DDoS attacks. Here is a simplified Equation(7) and Equation(8) for the Random Forest algorithm:

$$y = \begin{cases} T_1(x) & \text{if } T_1(x) \text{ is the mode,} \\ T_2(x) & \text{if } T_2(x) \text{ is the mode,} \\ & \vdots \\ T_n(x) & \text{if } T_n(x) \text{ is the mode.} \end{cases} \tag{7}$$

where: y is the predicted output variable (categorical or continuous) for a given input x. The expected values

$$T_1(x), T_2(x), \ldots, T_n(x) \tag{8}$$

are the outputs of n distinct decision trees, each of which is trained using the experimental data and the characteristics that are split randomly but in different ways [87].

2) REGRESSION ANALYSIS

Several machine-learning techniques used in regression analysis enable the prediction of a continuous (y) variable of interest as a function of the values of more or one (x) predictors [66]. Classification differs from Regression in that it forecasts discrete classes, while Regression simplifies the forecasting of a continuous variable.

a: LINEAR REGRESSION

This strategy is among the most often used machine learning models and is widely used as a regression approach. This technique employs a linear regression line, a continuous dependent variable, and a set of independent factors. In linear Regression, the relationship between the independent (X) and dependent (Y) factors is determined by finding the best-fit straight line (sometimes called the regression line) [66]. Here is the notation for linear Regression in Equation (9): The number of indicators, p, and the number of data, n, define the dimensions of the

$$\mathrm{X}_{n\times p} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np}, \end{bmatrix} \tag{9}$$

(also known as features). Consider the response variable y to be an n-dimensional vector. The linear regression model hypothesizes that the relationship between the predictors and the response variable is linear as given in Equation (10).

$$y = X\beta + \epsilon \tag{10}$$

where: β is the p-dimensional vector representing the coefficients for each predictor, and epsilon is "a," an n-dimensional vector that is the error term. It is supposed that the error component follows a normal distribution with a constant variation and zero means. Least squares estimation is used to calculate the beta values, which involves finding the values of β that minimize the sum of squared errors is in Equation (11):

$$SSE(\beta) = \sum_{i} (y_i - X_i\beta)^2 \tag{11}$$

where: "i" indexes the observations in the dataset. The least squares estimate the coefficients can be found using matrix algebra, specifically the formula is in Equation (12):

$$\beta = (X^T * X)^{(-1)} * X^T * y \tag{12}$$

where: $(X^T * X)^{(-1)}$ is the inverse of the matrix product $X^T * X$ and X^T is the transpose of X. This formula is known as the normal equations [59], [92].

b: LOGISTIC REGRESSION

Logistic Regression is a powerful statistical method designed for binary classification problems, determining the probability that an instance belongs to a specific class. At the core of the logistic regression model lies the logistic function, commonly known as the sigmoid function, which plays a pivotal role in its formulation [93].

The logistic function (sigmoid) is defined in Equation (13):

$$S(z) = \frac{1}{1 + e^{-z}} \tag{13}$$

In logistic Regression, the linear combination of input features is transformed using the logistic (sigmoid) function, given in Equation (14):

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n)}} \tag{14}$$

Here:

P(Y=1|X) is the probability of the dependent variable (Y) being one given the input features (X). β_0 is the intercept term. $\beta_1, \ldots, \beta_n$ are the coefficients associated with each input feature $x_1, \ldots, x_n$. e is the base of the natural logarithm. The logistic regression model outputs probabilities and a decision threshold is applied to convert these probabilities into class labels. If the predicted probability is greater than or equal to 0.5, the instance is classified as belonging to class 1; otherwise, it is classified as belonging to class 0.

Logistic Regression is vital in IoT security for binary classification tasks such as intrusion detection, device authentication, anomaly detection, and risk assessment. It predicts network threats, authenticates device logins, detects anomalies in device behavior, and assesses the associated risk, providing concise insights into potential security concerns in IoT environments [93].

c: NEURAL NETWORK REGRESSION

Modeling complicated non-linear connections between input factors and outcome variables is the goal of the potent machine learning method known as neural network regression. Here is the mathematical equation notation for a neural network regression with a single hidden layer is given in Equation (15):

$$y = f(w_2 \cdot f(w_1 \cdot X + b_1) + b_2) \tag{15}$$

where y is the output variable (the factor whose future you wish to forecast). X is the input variable (a vector of features), the w_1 matrix represents the weights used to pass data from the input layer to the concealed layer. This hidden-to-output layer weight matrix (w_2) is what connects the two layers, b_1 is the bias vector added to the hidden layer, b_2 is the bias scalar added to the output layer, the total of the inputs with weights, and f is a triggering function performing a non-linear transformation. The hidden layer performs non-linear transformations; its neuron count influences model complexity. Sigmoid, hyperbolic tangent, and Rectified Linear unit (ReLU) functions are common activation functions. The neural network outputs a scalar value that predicts the output variable [69], [70], [94].

B. UNSUPERVISED

Unsupervised learning is a technique that examines unlabeled datasets in a data-driven manner, without human intervention, as described in [66]. Data collection is undertaken for various reasons, including identifying significant patterns and structures, classifying trial results, and eliminating defining characteristics. Unsupervised learning is commonly used for Clustering, learning features, density estimation, association

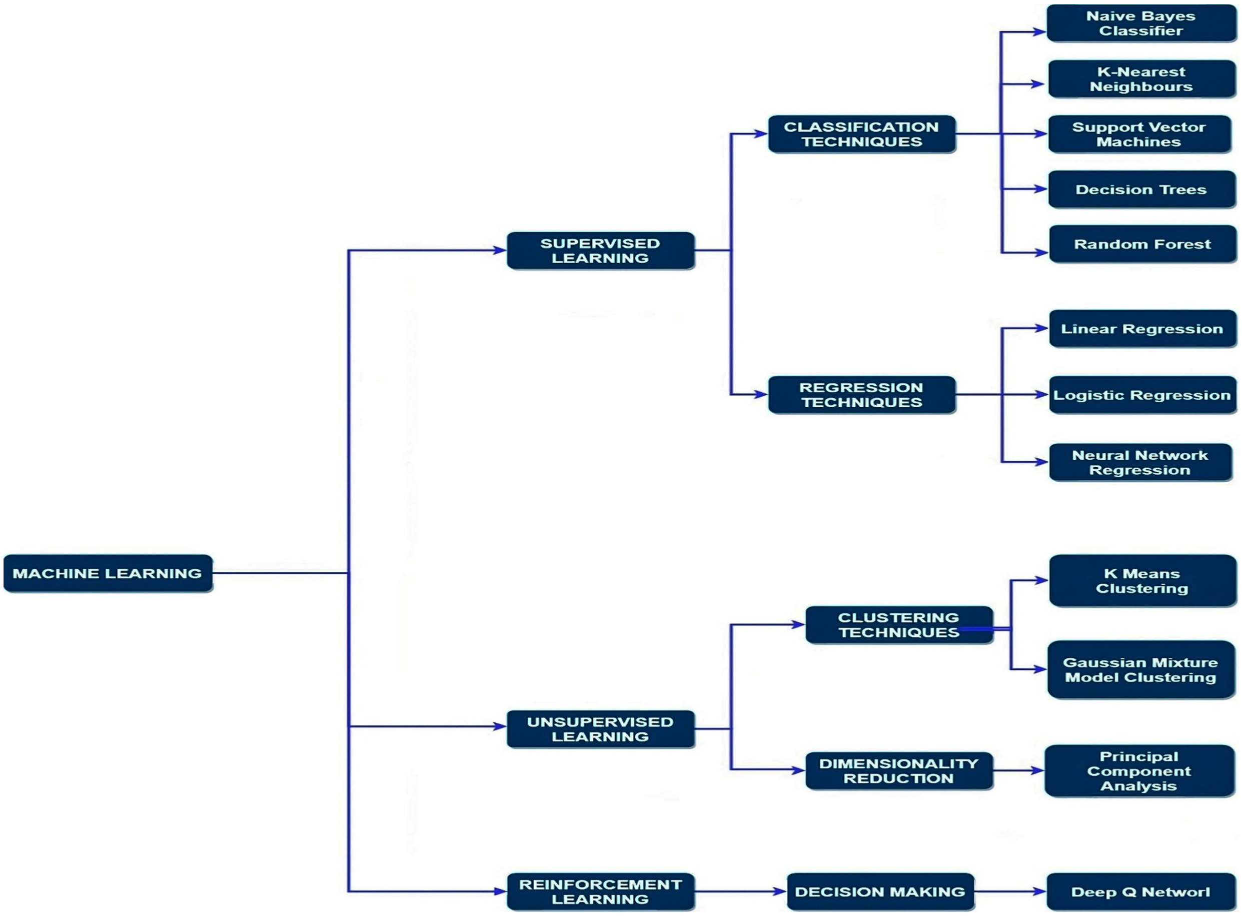

FIGURE 8. Illustration of various types of machine learning algorithms.

rule finding, anomalous detection, dimensionality reduction, and many other tasks. The mathematical equation notation for unsupervised machine learning is given in Equation (16):

$$X = f(Y) \tag{16}$$

where f is a function that maps Y to X, Y is a matrix of latent variables, and X is a matrix of input variables. Unsupervised machine learning finds the best function f that maps hidden variables Y to input variables X without knowing their connection [68], [70], [95].

1) CLUSTER ANALYSIS

Without worrying about getting the results precisely right, "Cluster analysis" (or "Clustering") is a machine learning technique for finding and organizing clusters of related data points in large datasets. Clustering is performed based on shared traits that increase the similarity among members within the same cluster compared to those in different clusters [66]. Frequent data analysis techniques divide customers into distinct groups according to their preferences and other characteristics. Cluster analysis has many potential uses, including internet security, health statistics, e-commerce, and mobile data processing.

a: K-MEANS CLUSTERING

Separating data sets using K-means clustering [96] is a quick and simple process that yields accurate results. In this method, data points are clustered so that the reciprocal distance separating them from the centroid is minimized as much as possible. K-means grouping is susceptible to anomalies because high values can easily influence a mean so the outcomes can differ depending on the data. Clustering using K-medoids [97] is an alternative to K-means that can be more robust against noise and anomalies. The K-mean clustering approach is often used in situations requiring the identification of anomalies [98], [99] as well as Sybil attacks [100]. K-means clustering has the following mathematical notation is given in Equation (17): In the presence of some facts,

$$X = x_1, x_2, \ldots, x_n, \tag{17}$$

where: x_i is a data point in a d-dimensional space, and several clusters k known in advanced; the K-means algorithm defines the following terms: Cluster centroids: A set of k points $m_1, m_2, \ldots, m_k$ that represent the centers of the k clusters. Cluster assignments: A list of n identifiers $c_1, c_2, \ldots, c_n$ that specifies the position within the collection where each data element resides. Using these terms, K-means sets out the following goal function to achieve is given in Equation (18):

$$\min_{m_j} \sum_{i=1}^{n} \|x_i - m_j\|^2 \tag{18}$$

where $j = c_i$ is the number of the group that the i-th data item belongs to x_i is assigned, and m_j represents the cluster's center j. The K-means algorithm iteratively updates the cluster centroids and cluster assignments to minimize this objective function:

1. Initialize k cluster centroids randomly.
2. Place each observation in the region around its center.
3. Make sure the average of all the data points in each cluster is used as the new centroid.
4. The above process 2-3 is repeated until convergence (when cluster allocations stop changing or when a specified amount of cycles is reached).

The resulting set of k clusters is the output of the K-means algorithm [101].

b: GMM CLUSTERING

Gaussian mixture models (GMMs), a distribution-based clustering approach, are often employed for data clustering. In a probabilistic model known as a "Gaussian mixture," a small number of Gaussian distributions with random values are combined to produce all the data points [102]. The expectation-maximization (EM) [102] optimization approach may be used to determine the cluster-specific Gaussian parameters. EM, an iterative technique, employs a statistical model to estimate the parameters. Here's a simplified mathematical notation for the GMM clustering algorithm: Let $X = x_1, x_2, \ldots, x_n$ takes n data values as inputs, and K is the number of clusters to be formed.

1. Choose the number of clusters K.
2. Initialize the parameters of K Gaussian distributions: Mean: μ_k Covariance matrix: Σ_k.
3. Place each data point in the cluster whose center is geographically closest to it: Calculate the probability of each data point x_i belonging to cluster k using Bayes' theorem is given in Equation (19):

$$Pr(\text{cluster } k|\text{data point } x_i) = \frac{Pr(\text{data point } x_i|\text{cluster } k) \cdot Pr(\text{cluster } k)}{Pr(\text{data point } x_i)} \tag{19}$$

Assign x_i to the cluster with the highest probability.

4. Update the mean and covariance matrix of each cluster based on the data points assigned to it: Calculate the new mean μ_k and covariance matrix Σ_k using the data points assigned to cluster k is given in Equation (20) and Equation (21):

$$\mu_k = \frac{1}{N_k} \sum_{x_i \text{ in cluster } k} x_i \tag{20}$$

$$\Sigma_k = \frac{1}{N_k} \sum_{x_i \text{ in cluster } k} (x_i - \mu_k) \cdot (x_i - \mu_k)^T \tag{21}$$

where N_k is the total amount of observations in cluster k.

5. Update the mixing coefficient for each cluster: Calculate the new mixing coefficient π_k, which represents the probability of selecting cluster k in Equation (22):

$$\pi_k = \frac{N_k}{n} \tag{22}$$

where the number of observations is denoted by n.

6. Repeat steps 3-5 until convergence, i.e., the process stops after some fixed number of repetitions or when the groups stop evolving. The final output of the GMM algorithm is the K clusters, each represented by a mean μ_k, covariance matrix Σ_k, and mixing coefficient π_k [103].

2) DIMENSIONALITY REDUCTION AND FEATURE LEARNING

Dimensionality reduction in unsupervised learning is vital for improved human interpretation, reduced computing costs, and preventing model complexity. It includes feature selection or extraction, with the former keeping some original traits and the latter creating new ones [104], [105]. Feature selection narrows relevant variables, enhancing data science and machine learning by removing unnecessary information, simplifying model complexity, speeding up algorithm training, and addressing overfitting [106]. Simultaneously, feature extraction methods, like "feature extraction" [106], enhance data understanding, boost prediction accuracy, and cut processing expenses in machine learning systems. Overall, feature extraction generates new features, summarizing the original set for a streamlined feature collection.

a: PRINCIPLE COMPONENT ANALYSIS

It is customary to apply the dimensionality reduction method of principal components Analysis (PCA) to generate brand-new elements from preexisting traits in a collection [105]. PCA is a popular unsupervised learning method in deep learning and data analytics. PCA [107], [108] is a statistical technique for reducing many independent variables to a smaller number of independent variables. Constructing a robust security protocol using PCA in conjunction with several other machine-learning techniques is possible. PCA and other classifier methods, such as KNN and softmax regression, are used in a recently developed model that results in an effective system [109]. PCA's basic algebraic solution notation is given from Equations (23) to Equation (26): X is a n x d matrix with n data values and d variables or traits.

1. Mean center:

$$Z = X - \mu. \tag{23}$$

where μ is the mean vector of X

2. Covariance matrix:

$$C = (1/n) * Z^T * Z \quad (24)$$

3. Eigenvectors and eigenvalues:

$$C * v = \lambda * v \quad (25)$$

4. Choose the top k eigenvectors with the highest eigenvalues as principal components

5. Transform X to lower-dimensional space using a projection matrix [68], [69], [110].

$$W = [v_1, v_2, \ldots, v_k], X_{new} = Z * W \quad (26)$$

C. REINFORCEMENT LEARNING

To improve efficacy, software bots, and machines can use reinforcement learning to autonomously determine the best behavior in a given setting or situation [111]. This incentive- or penalty-based learning strategy uses environmental activism expertise to increase benefit or reduce risk [112]. It is not advised for fundamental issues. It can teach AI models to automate or improve complex systems like robotics, automated driving, production, and supply chain networks.

Reinforcement learning (RL) allows agents to learn by doing in dynamic environments. Unlike directed learning, which relies on sample data, RL emphasizes environmental interaction. Reinforcement learning (RL) addresses the Markov Decision Process (MDP) [113] of sequential decisions. The ideal action may be inferred from a model of the environment using paradigm-based RL by performing activities and watching the results, such as the following phase and a quick reward [114]. Modeling is at the heart of AlphaZero and AlphaGo [115]. The Bellman equation, which describes the core concept of reinforcement learning, states that a state's value equals the total of its instant compensation and the deferred value of the next state is given in Equation (27).

$$V(s) = \max_a \sum_{s'} P(s'|s, a)[R(s, a, s') + \gamma V(s')] \quad (27)$$

The anticipated total benefit from beginning at state s and following the optimum strategy is denoted by V(s), where s is an arbitrary integer, and max_a chooses a to-do that will maximize the total anticipated prize or reward. $\sum_{s'}$ is the cumulative total of all future situations s'. Given an action, the chance of shifting from the state of s to the state of s' is denoted by $P(s'|s, a)$. When shifting from the state of s to the state of s'. via action a, the instant recompense is denoted by R(s, a, s'), and the reduction component is denoted by γ. Many reinforcement learning methods use the above-mentioned formula to adjust the value function or strategy continuously [116].

a: DEEP Q NETWORK (DQN)

A reinforcement learning technique called Deep Q Network (DQN) expands on the conventional Q-learning methodology by utilizing deep neural networks to manage intricate and high-dimensional state spaces. The central concept approximates the Q-function—a measure of the expected future rewards for actions taken in a given state—using a deep neural network [117].

In DQN, the Q-function is written as Q(s,α,θ)

where s stands for state, α for action, and θ for deep neural network parameters. Through repeated updates of the Q-value, the temporal difference (TD) error is utilized in Equation (28):

$$Q(s, a; \theta_{i+1}) = (1 - \alpha) \cdot Q(s, a; \theta_i) + \alpha \cdot (r + \gamma \cdot \max_{a'} Q(s', a'; \theta_i)) \quad (28)$$

Here: The rate of learning is α. The instant prize is r. The discount factor is γ. The following state is s'.

The difference between the goal Q-value and the anticipated Q-value is known as the temporal difference error, and the deep neural network is trained to reduce it.

DQN strengthens IoT security by learning optimal strategies for intrusion detection, adapting to dynamic threats, optimizing resource allocation, and aiding in anomaly detection. Its adaptability and efficiency make it a powerful tool for addressing complex security challenges in IoT environments [117].

So far, we have discussed the Machine Learning Algorithms in detail, which we consolidated in Table 10, providing examples of the many machine learning methods or attack detection, followed by their respective Accuracies [117].

Table 11 Summarizes the Name of the Dataset used for Machine Learning along with Evaluation metrics and Different types of attacks.

VI. CREATING SECURE DESIGNS: PROBLEMS AND SOLUTIONS

In this section, we examine IoT security from a systemic perspective. We first examine a few conditions that must be satisfied and certain preventative actions. Table 12 displays a few important specifications

A. CONFIDENTIALITY

Specialized encryption process techniques must used to ensure data confidentiality, [128], [129], which will prevent unauthorized usage of IoT infrastructure and the leakage of sensitive data. This service is intended to keep unauthorized users from networks and secure essential information.

B. AUTHENTICATION AND AUTHENTICITY

By allowing only authorized users to access and take control of the protected resources, the system can maintain the Security of the IoT Network [128], [130]. Networks, databases, computer systems, and other network-based

TABLE 10. Illustration of various machine learning along with application and Accuracy.

ML Type	ML Technique	Representation	Potential Uses/Detection of Attacks	Acc(%)	Ref.
SUPERVISED	NB	Probabilistic classification based on Bayes' theorem	Detection of Intruders	50-78	[74]
SUPERVISED	KNN	Assigning based on proximity in feature space	Intruder Detection/Malware Detection	99.6	[76]
SUPERVISED	SVM	Finding hyperplanes to separate classes	Intrusion Detection	99.86	[82].
SUPERVISED	DT	Tree-like model based on input features	Invasion and Unusual Traffic Source Detection	99.9	[87]
SUPERVISED	RF	Ensemble of decision trees	Intruder Detection/Malware Detection.	99.69	[118]
SUPERVISED	Linear Regression	Modeling the relationship between variables linearly	Anomaly detection for abnormal behaviour in IoT sensor readings	91.04-99.95	[119], [59].
SUPERVISED	Logistic Regression	Probability modelling for binary outcomes	Intrusion detection in IoT networks	98.3	[93].
SUPERVISED	NN	Hierarchical interconnected nodes	Network Security in IoT	99	[94]
UNSUPERVISED	K-Means	Partitioning data into 'k' clusters based on centroids	Abnormal Behavior, Data Tampering, and Sybil Detection in medical IoT.	94.5	[96]
UNSUPERVISED	GMM Clustering	Modeling data as a mixture of Gaussian distributions	Anomaly detection, Botnet detection, Intrusion detection	93	[103]
UNSUPERVISED	PCA	Reducing dimensionality while preserving variance	Real-Time Detection System, Intrusion Detection.	98.2	[107].
REINFORCEMENT	RL	Training algorithms through rewards to optimize decision-making in dynamic environments	DoS	96.5	[116]
REINFORCEMENT	DQN	Extends Q-learning with deep neural networks as function approximators	Proxy detection and Botnet detection, Intrusion detection, Cyber-physical system security	93-96.1	[117]

TABLE 11. Name of the dataset used for machine learning along with evaluation metrics and different types of attacks.

DATASET NAME	EVALUATION METRICS	IOT ATTACKS	ML TECHNIQUES	References
CONFICKER Worm, UNINA traffic traces, and CAIDA.	Area under the curve (AUC), False-positive rate, f-measure (sensitivity), specialization (specificity), Accuracy	DDoS	SVM and BN(Batch Normalization).	[120]
RPL-NIDDS17	Accuracy and AUC	We've got the Sinkhole, local fix assaults, Sybil, black hole, distributed denial of service, selective forwarding, and hello flooding.	Ensemble Learning(EL)	[121]
BoT-IoT	Precision, Reliability, F-measure, and Recall	Scams, denial-of-service assaults, infiltration, and denial-of-service	NN	[122]
intel IoT	Detection rate, False Positive rate, and Accuracy	Probing, DOS, U2R "User to Root" and R2L "Root to Local."	Clustering.	[123]
DS2OS	Precision, accuracy, the F1 Score, and Recall	Probing and DOS.	NN	[124]
MedBIoT	F1 score	DOS	NN	[125]
UNSW-NB15 and KDD-CUP99	Precision, F1 Score, erroneous positive rate, and overall efficacy.	Backdoor, Reconnaissance R2L, U2R, DOS, Analysis, generic, fuzzes, and shell-code.	Bayesian, Decision Trees (DT), and Clustering.	[126]
UNSW-NB15, CICIDS2017	Accuracy, false positive rate, specificity, and sensitivity.	DOS	DT.	[127]

services may be among the resources. It is primarily used to confirm the user's identification and to establish the client privilege levels for the various kinds of resources in the IoT Network.

C. INTEGRITY

Maintaining data security means keeping private and priceless information safe from hackers. There are many potential threats to data security, including computer outages. By adding a fixed-length value to maintain data integrity and identify message encryption issues in the Internet of Things (IoT), the Cyclic Redundancy Check (CRC) may be used [128], [131]. The system should primarily increase the reliability of data transmitted through networks while maintaining correctness and consistency.

D. AVAILABILITY

Data accessibility is essential for the Internet of Things because it assures consumers of the Security and dependability of the data they may access. An IoT system must offer a backup of crucial data to avoid data loss. Attacks like DoS and DDoS attacks might damage data availability [128], [131]. The IoT Network should always be accessible, regardless of system failures, hardware, or software issues. Predicting the bottlenecks should be done to supply the bandwidth.

VII. MAJOR CHALLENGES

A. DATA PRIVACY AND SECURITY

1) DATA BREACHES

IoT devices collect vast amounts of data, often sensitive, which cyber-criminals can target. Data encryption, secure storage, and transmission are crucial [36].

2) PRIVACY CONCERNS

With so many interconnected devices, maintaining the privacy of individuals' data is challenging. Methods to anonymize data without losing its utility are needed [36].

B. SCALABILITY

1) DEVICE MANAGEMENT

Managing and securing many heterogeneous devices in a smart city environment is difficult [132].

2) RESOURCE CONSTRAINTS

Many IoT devices have limited processing power, memory, and battery life, making it challenging to implement robust security measures [132].

C. INTEROPERABILITY

1) STANDARDIZATION

The lack of standardized protocols and security measures across different IoT devices and manufacturers complicates security implementations [133].

2) INTEGRATION

Ensuring seamless and secure integration of various systems and devices is critical [133].

D. REAL-TIME THREAT DETECTION

1) TIMELY RESPONSES

Detecting and responding to security threats in real-time is essential but challenging due to the volume of data and potential latency issues [134].

2) FALSE POSITIVES/NEGATIVES

ML algorithms need to be highly accurate to avoid false alarms and missed threats [134].

E. COMPLEX ATTACK VECTORS

1) ADVANCED PERSISTENT THREATS (APTs)

Sophisticated, multi-phase attacks that are difficult to detect and mitigate [135].

2) PHYSICAL SECURITY

IoT devices are often deployed in public spaces, making them vulnerable to physical tampering or attacks [135].

VIII. UNSOLVED PROBLEMS

A. TRUST MANAGEMENT

1) DEVICE AUTHENTICATION

Establishing the identity of devices in a trustworthy manner is challenging due to the diversity and scale of IoT networks. Robust authentication mechanisms that can handle millions of devices without compromising Security or performance are required [136].

2) DATA INTEGRITY

Ensuring the integrity of data collected and transmitted by IoT devices is crucial. Techniques for verifying that data has not been tampered with, both in transit and at rest, are essential to maintaining trust [136].

B. ADAPTIVE SECURITY MECHANISMS

1) DYNAMIC THREAT LANDSCAPES

The threat landscape for IoT systems is constantly evolving. Security mechanisms must adapt quickly to new threats without requiring manual updates. Which requires ongoing learning and adjustment by security systems [137].

2) CONTEXT-AWARE SECURITY

Security measures that can adapt based on the context in which a device is operating (e.g., location, current network conditions) can provide more effective protection [137].

C. ENERGY-EFFICIENT SECURITY PROTOCOLS

1) LOW-POWER DEVICES

Many IoT devices operate on limited power sources, such as batteries, making energy efficiency a critical concern. Developing cryptographic algorithms and security protocols that minimize power consumption while maintaining robust Security is a significant challenge [138].

TABLE 12. Quality features and IoT security overview.

Quality features	IoT security overview
Data integrity	Data integrity ensures trustworthiness and precision by showing that information has not been tampered with or removed.
Data Confidentiality	Data confidentiality strives to keep information hidden from unauthorized parties, safeguarding users' privacy and sensitive information from being obtained by attackers. The information is only accessible to authorized users.
Data availability	To ensure that resources (such as data and services) are available, data availability is utilized.
Authentication	Authentication is the process by which the names of users requesting access to a resource are checked and filtered. Authentication techniques are essential for inter-thing contact in the Internet of Things.
Authorization	The procedure for providing, rejecting, and limiting access to entities is known as authorization. The permission scheme executes several actions by various entities.

2) BALANCING SECURITY AND PERFORMANCE

Ensuring that security measures do not excessively degrade the performance or lifespan of IoT devices is crucial for practical deployments [138].

D. DECENTRALIZED SECURITY SOLUTIONS

1) SCALABILITY OF BLOCKCHAIN

While blockchain offers promising security benefits, its scalability remains an issue. Developing blockchain solutions that can handle many transactions and devices in an innovative city environment without compromising speed or efficiency is a critical research area [139].

2) CONSENSUS MECHANISMS

Traditional consensus mechanisms used in blockchain, such as Proof of Work (PoW), are unsuitable for IoT due to their high computational and energy requirements. Researching alternative consensus mechanisms that are lightweight and efficient is essential [139].

E. HUMAN FACTORS

1) USER EDUCATION AND AWARENESS

Many security breaches occur due to human error. Educating users about security best practices and designing user-friendly security interfaces can help mitigate this risk [140].

2) INSIDER THREATS

Addressing the risk of insiders with legitimate access to IoT systems but misusing their privileges is a complex problem requiring sophisticated monitoring and anomaly detection techniques [140].

IX. POTENTIAL DIRECTIONS AND METHODS FOR FUTURE RESEARCH

A. ADVANCED MACHINE LEARNING TECHNIQUES

1) DEEP LEARNING AND NEURAL NETWORKS

Utilizing deep learning models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), can significantly enhance the detection of complex patterns and anomalies in IoT data. Research can focus on developing lightweight versions of these models suitable for deployment on resource-constrained devices [141].

2) REINFORCEMENT LEARNING

Reinforcement learning can be used to develop adaptive security systems that improve their effectiveness over time. For example, reinforcement learning agents can learn optimal intrusion detection and response strategies based on feedback from the environment [142].

B. FEDERATED LEARNING

1) PRIVACY-PRESERVING TRAINING

Federated learning allows models to be trained on decentralized data sources without centralizing the data, thereby preserving privacy. Research can explore how to enhance federated learning algorithms to ensure robustness against adversarial attacks and improve their efficiency [143].

2) COLLABORATIVE MODEL IMPROVEMENT

Developing techniques for collaborative model improvement where multiple IoT devices contribute to a global model can enhance the overall security posture without compromising individual device security [143].

C. BLOCKCHAIN AND DISTRIBUTED LEDGER TECHNOLOGIES

1) LIGHTWEIGHT BLOCKCHAIN PROTOCOLS

Researching lightweight blockchain protocols that reduce the computational and storage requirements can make blockchain more suitable for IoT environments. Which includes exploring sidechains and off-chain transactions to reduce the load on the main blockchain [144].

2) SMART CONTRACTS FOR IOT SECURITY

Utilizing smart contracts to automate and enforce security policies in IoT networks can provide a robust and transparent mechanism for managing security rules and actions [145].

D. QUANTUM-RESISTANT SECURITY PROTOCOLS

1) POST-QUANTUM CRYPTOGRAPHY

Developing cryptographic algorithms resistant to quantum attacks is crucial for future-proofing IoT security. Research can focus on evaluating and implementing post-quantum cryptographic standards in IoT devices [146].

2) HYBRID CRYPTOGRAPHIC APPROACHES

Combining traditional cryptographic techniques with quantum-resistant algorithms to provide a layered security approach that can transition smoothly as quantum computing becomes more prevalent [146].

E. IOT-SPECIFIC INTRUSION DETECTION SYSTEMS (IDS)

1) BEHAVIORAL ANALYSIS

Developing IDS that leverage machine learning to analyze the behavior of IoT devices and detect deviations from normal patterns can enhance threat detection. Which includes creating models that understand the typical usage patterns and operational behaviors of devices [147].

2) DISTRIBUTED IDS

Implementing distributed IDS operating at the network's edge to monitor and analyze traffic locally can reduce latency and provide faster threat detection and response [148].

F. EDGE COMPUTING

1) REAL-TIME DATA PROCESSING

Utilizing edge computing to process data and execute security algorithms closer to the source can enhance real-time threat detection and mitigation. Research can focus on optimizing edge computing frameworks for security tasks [149].

2) COLLABORATIVE EDGE SECURITY

Developing mechanisms for collaborative Security among edge devices, sharing threat intelligence, and coordinating responses can provide a more resilient security infrastructure [150].

G. COLLABORATIVE SECURITY MODELS

1) CROSS-INDUSTRY COLLABORATION

Encouraging collaboration among different industries, government bodies, and academia can lead to the development of standardized security practices and shared threat intelligence networks [151].

2) PUBLIC-PRIVATE PARTNERSHIPS

Leveraging public-private partnerships to fund and support security research and implementation can accelerate the development and deployment of robust security solutions in smart cities [151].

H. ETHICAL AI IN IoT SECURITY

1) TRANSPARENCY AND EXPLAINABILITY

Ensuring that AI and ML algorithms used for Security are transparent and their decision-making processes are explainable can increase trust and compliance with ethical standards [152].

2) BIAS MITIGATION

Researching techniques to identify and mitigate biases in AI models can ensure that security measures are fair and do not disproportionately affect certain groups or individuals [152].

I. SUSTAINABLE IoT SOLUTIONS

1) GREEN IoT DESIGN

Developing IoT devices with sustainability in mind, such as using eco-friendly materials, designing for durability, and minimizing energy consumption [153].

2) RENEWABLE ENERGY INTEGRATION

Researching methods to integrate renewable energy sources, such as solar or wind power, to supply IoT devices, reducing the overall carbon footprint of intelligent city deployments [153].

3) ENERGY HARVESTING

Exploring technologies that allow IoT devices to harvest energy from their environment, such as solar, thermal, or kinetic energy, to extend their operational life and reduce reliance on batteries [153].

X. CONCLUSION

This paper introduces the intelligent city and describes the need for a Machine Learning Algorithm in IoT security. Then, a literature review of Machine Learning methods, their uses, accuracy, benefits, and drawbacks are presented, as well as the usage of various attacks in Datasets and Machine learning. An overview of Essential Focus Areas for IoT Security areas and different types of possible attacks are given, and a detailed countermeasure for various attacks is given. At last, IoT's various potential applications and security concerns are in the planning phase. The information presented in this article will be helpful to scholars in determining the various Machine Learning algorithm classes, understanding their operation, and mapping out potential attack vectors. Here, the different types of attacks in other areas are discussed, which helps to identify the possible risks. Then, the security issues are defined to avoid or handle those types of challenges in the future. This paper gives an idea about the research challenges and future Security with a detailed literature review.

REFERENCES

[1] S. S. E. A. Irei. (2024). *Cybersecurity Market Researchers Forecast Significant Growth*. Accessed: 2024-05-17. [Online]. Available: https://www.techtarget.com/searchsecurity/feature/Cybersecurity-market-researchers-forecast-significant-growth

[2] M. Jia, A. Komeily, Y. Wang, and R. S. Srinivasan, "Adopting Internet of Things for the development of smart buildings: A review of enabling technologies and applications," *Autom. Construct.*, vol. 101, pp. 111–126, May 2019.

[3] B. Deepa, C. Anusha, and P. C. Devi, "Smart agriculture using IoT," in *Intelligent System Design*. Berlin, Germany: Springer, 2021, pp. 11–19.

[4] H. Yang, S. Kumara, S. T. Bukkapatnam, and F. Tsung, "The Internet of Things for smart manufacturing: A review," *IISE Trans.*, vol. 51, no. 11, pp. 1190–1216, 2019.

[5] M. Abdel-Basset, G. Manogaran, M. Mohamed, and E. Rushdy, "Internet of Things in smart education environment: Supportive framework in the decision-making process," *Concurrency Comput., Pract. Exper.*, vol. 31, no. 10, p. 4515, May 2019.

[6] B. B. Sinha and R. Dhanalakshmi, "Recent advancements and challenges of Internet of Things in smart agriculture: A survey," *Future Gener. Comput. Syst.*, vol. 126, pp. 169–184, Jan. 2022.

[7] J. Sakhnini, H. Karimipour, A. Dehghantanha, R. M. Parizi, and G. Srivastava, "Security aspects of Internet of Things aided smart grids: A bibliometric survey," *Internet Things*, vol. 14, Jun. 2021, Art. no. 100111.

[8] F. Alshehri and G. Muhammad, "A comprehensive survey of the Internet of Things (IoT) and AI-based smart healthcare," *IEEE Access*, vol. 9, pp. 3660–3678, 2021.

[9] B. Jan, H. Farman, M. Khan, M. Talha, and I. U. Din, "Designing a smart transportation system: An Internet of Things and big data approach," *IEEE Wireless Commun.*, vol. 26, no. 4, pp. 73–79, Aug. 2019.

[10] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions," *Future Gener. Comput. Syst.*, vol. 29, no. 7, pp. 1645–1660, Sep. 2013.

[11] H. Muccini and M. T. Moghaddam, "IoT architectural styles: A systematic mapping study," in *Proc. 12th Eur. Conf. Softw. Archit.* Madrid, Spain: Springer, Sep. 2018, pp. 68–85.

[12] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A survey on enabling technologies, protocols, and applications," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2347–2376, 4th Quart., 2015.

[13] J. Pan and J. McElhannon, "Future edge cloud and edge computing for Internet of Things applications," *IEEE Internet Things J.*, vol. 5, no. 1, pp. 439–449, Feb. 2018.

[14] J. Green, "The Internet of Things reference model," in *Internet of Things World Forum*. San Jose, CA, USA: CISCO, 2014, pp. 1–12.

[15] K. Gloss. (2021). *6 IoT Security Layers to Shape the Ultimate Defense Strategy*. Accessed: 2023-05-09. [Online]. Available: https://www.techtarget.com/iotagenda/tip/6-IoT-security-layers-to-shape-the-ultimate-defense-strategy

[16] M. M. Ahemd, M. A. Shah, and A. Wahid, "IoT security: A layered approach for attacks & defenses," in *Proc. Int. Conf. Commun. Technol. (ComTech)*, Apr. 2017, pp. 104–110.

[17] S. S. Sivaraju, V. Mani, A. Umaamaheshvari, P. D. Banu, T. Anuradha, and S. Srithar, "An attack resistant physical unclonable function smart optical sensors for Internet of Things for secure remote sensing," *Meas., Sensors*, vol. 29, Oct. 2023, Art. no. 100882.

[18] E. Hammad and A. Farraj, "A physical-layer security approach for IoT against jamming interference attacks," in *Proc. IEEE Can. Conf. Electr. Comput. Eng. (CCECE)*, Sep. 2021, pp. 1–6.

[19] A. Tewari and B. B. Gupta, "Security, privacy and trust of different layers in Internet-of-Things (IoTs) framework," *Future Gener. Comput. Syst.*, vol. 108, pp. 909–920, Jul. 2020.

[20] N. F. Syed, Z. Baig, A. Ibrahim, and C. Valli, "Denial of service attack detection through machine learning for the IoT," *J. Inf. Telecommun.*, vol. 4, no. 4, pp. 482–503, Oct. 2020.

[21] B. Hettwer, S. Gehrer, and T. Güneysu, "Applications of machine learning techniques in side-channel attacks: A survey," *J. Cryptograph. Eng.*, vol. 10, no. 2, pp. 135–162, Jun. 2020.

[22] S. Sicari, A. Rizzardi, D. Miorandi, and A. Coen-Porisini, "REATO: REActing TO denial of service attacks in the Internet of Things," *Comput. Netw.*, vol. 137, pp. 37–48, Jun. 2018.

[23] V. Visoottiviseth, P. Sakarin, J. Thongwilai, and T. Choobanjong, "Signature-based and behavior-based attack detection with machine learning for home IoT devices," in *Proc. IEEE REGION 10 Conf. (TENCON)*, Nov. 2020, pp. 829–834.

[24] A. O. Sangodoyin, M. O. Akinsolu, P. Pillai, and V. Grout, "Detection and classification of DDoS flooding attacks on software-defined networks: A case study for the application of machine learning," *IEEE Access*, vol. 9, pp. 122495–122508, 2021.

[25] T. Gaber, A. El-Ghamry, and A. E. Hassanien, "Injection attack detection using machine learning for smart IoT applications," *Phys. Commun.*, vol. 52, Jun. 2022, Art. no. 101685.

[26] S. Modak, K. Majumder, and D. De, "Vulnerability of cloud: Analysis of xml signature wrapping attack and countermeasures," in *Proc. Int. Conf. Frontiers Comput. Syst.* Singapore: Springer, 2021, pp. 755–765.

[27] S. Alberternst, A. Anisimov, A. Antakli, B. Duppe, H. Hoffmann, M. Meiser, M. Muaz, D. Spieldenner, and I. Zinnikus, "Orchestrating heterogeneous devices and AI services as virtual sensors for secure cloud-based IoT applications," *Sensors*, vol. 21, no. 22, p. 7509, Nov. 2021.

[28] S. Abaimov and G. Bianchi, "CODDLE: Code-injection detection with deep learning," *IEEE Access*, vol. 7, pp. 128617–128627, 2019.

[29] A. Youssef, M. Abdelrazek, and C. Karmakar, "Use of ensemble learning to detect buffer overflow exploitation," *IEEE Access*, vol. 11, pp. 52009–52025, 2023.

[30] M. You, "An adaptive machine learning framework for access control decision making," Ph.D. thesis, Inst. Sustain. Ind. Liveable Cities (ISILC), Victoria Univ., Melbourne, VIC, Australia, 2022.

[31] S. Naaz, "Detection of phishing in Internet of Things using machine learning approach," *Int. J. Digit. Crime Forensics*, vol. 13, no. 2, pp. 1–15, Mar. 2021.

[32] Y. Liang, S. Samtani, B. Guo, and Z. Yu, "Behavioral biometrics for continuous authentication in the Internet-of-Things era: An artificial intelligence perspective," *IEEE Internet Things J.*, vol. 7, no. 9, pp. 9128–9143, Sep. 2020.

[33] E. Fernandes, J. Jung, and A. Prakash, "Security analysis of emerging smart home applications," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2016, pp. 636–654.

[34] M. Chan. (2017). *Why Cloud Comp, Uting is the Foundation of the Internet of Things*. Accessed: 2023-05-09. [Online]. Available: https://thorntech.com/cloud-computing-foundation-internet-things/

[35] J. M. Blanco, M. Ge, and T. Pitner, "Modeling inconsistent data for reasoners in web of things," *Proc. Comput. Sci.*, vol. 192, pp. 1265–1273, Jan. 2021.

[36] M. Aljabri, A. A. Alahmadi, R. M. A. Mohammad, F. Alhaidari, M. Aboulnour, D. M. Alomari, and S. Mirza, "Machine learning-based detection for unauthorized access to IoT devices," *J. Sensor Actuator Netw.*, vol. 12, no. 2, p. 27, Mar. 2023.

[37] G. Vojković, M. Milenković, and T. Katulić, "IoT and smart home data breach risks from the perspective of data protection and information security law," *Bus. Syst. Res., Int. J. Soc. Advancing Innov. Res. Economy*, vol. 11, no. 3, pp. 167–185, Nov. 2020.

[38] I. Ahmad, M. S. Niazy, R. A. Ziar, and S. Khan, "Survey on IoT: Security threats and applications," *J. Robot. Control (JRC)*, vol. 2, no. 1, pp. 42–46, 2021.

[39] C. Machado and A. A. Medeiros Fröhlich, "IoT data integrity verification for cyber-physical systems using blockchain," in *Proc. IEEE 21st Int. Symp. Real-Time Distrib. Comput. (ISORC)*, May 2018, pp. 83–90.

[40] I. Zualkernan, N. Ahmed, A. Elmeligy, A. Abdelnaby, and N. Sheta, "IoT sensor data consistency using deep learning," in *Proc. IEEE Int. Conf. Internet Things Intell. Syst. (IoTaIS)*, Nov. 2022, pp. 198–203.

[41] D. Zheng, A. Wu, Y. Zhang, and Q. Zhao, "Efficient and privacy-preserving medical data sharing in Internet of Things with limited computing power," *IEEE Access*, vol. 6, pp. 28019–28027, 2018.

[42] J. Sengupta, S. Ruj, and S. D. Bit, "End to end secure anonymous communication for secure directed diffusion in IoT," in *Proc. 20th Int. Conf. Distrib. Comput. Netw.*, Jan. 2019, pp. 445–450.

[43] P. Gope and B. Sikdar, "Lightweight and privacy-preserving two-factor authentication scheme for IoT devices," *IEEE Internet Things J.*, vol. 6, no. 1, pp. 580–589, Feb. 2019.

[44] K. Gai, K. R. Choo, M. Qiu, and L. Zhu, "Privacy-preserving content-oriented wireless communication in Internet-of-Things," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 3059–3067, Aug. 2018.

[45] C. Toma, C. Boja, M. Popa, M. Doinea, and C. Ciurea, "Viruses, exploits, malware and security issues on IoT devices," in *Proc. Int. Conf. Inf. Technol. Commun. Secur.* Cham, Switzerland: Springer, 2021, pp. 324–334.

[46] G. Ferronato, "IoT white worms: Design and application," M.S. thesis, Dept. Elect. Eng., Math., Comput. Sci., Univ. Twente, Enschede, The Netherlands, 2020.

[47] H. Kanaker, N. Abdel Karim, S. A. B. Awwad, N. H. A. Ismail, J. Zraqou, and A. M. F. Al ali, "Trojan horse infection detection in cloud based environment using machine learning," *Int. J. Interact. Mobile Technol. (iJIM)*, vol. 16, no. 24, pp. 81–106, Dec. 2022.

[48] M. Sulaiman, A. Khan, A. Negash Ali, G. Laouini, and F. Sameer Alshammari, "Quantitative analysis of worm transmission and insider risks in air-gapped networking using a novel machine learning approach," *IEEE Access*, vol. 11, pp. 111034–111052, 2023.

[49] J. Su, D. V. Vasconcellos, S. Prasad, D. Sgandurra, Y. Feng, and K. Sakurai, "Lightweight classification of IoT malware based on image recognition," in *Proc. IEEE 42nd Annu. Comput. Softw. Appl. Conf. (COMPSAC)*, vol. 2, Jul. 2018, pp. 664–669.

[50] H. Naeem, B. Guo, and M. R. Naeem, "A light-weight malware static visual analysis for IoT infrastructure," in *Proc. Int. Conf. Artif. Intell. Big Data (ICAIBD)*, May 2018, pp. 240–244.

[51] P. S. Dusane, "Logic bomb: An insider attack," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 3, pp. 3662–3665, Jun. 2020.

[52] J. Liu, C. Zhang, and Y. Fang, "EPIC: A differential privacy framework to defend smart homes against internet traffic analysis," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 1206–1217, Apr. 2018.

[53] U. Guin, A. Singh, M. Alam, J. Cañedo, and A. Skjellum, "A secure low-cost edge device authentication scheme for the Internet of Things," in *Proc. 31st Int. Conf. VLSI Design 17th Int. Conf. Embedded Syst. (VLSID)*, Jan. 2018, pp. 85–90.

[54] S. Rabhi, T. Abbes, and F. Zarai, "IoT routing attacks detection using machine learning algorithms," *Wireless Pers. Commun.*, vol. 128, no. 3, pp. 1839–1857, Feb. 2023.

[55] C. Pu and S. Hajjar, "Mitigating forwarding misbehaviors in RPL-based low power and lossy networks," in *Proc. 15th IEEE Annu. Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2018, pp. 1–6.

[56] C. Ioannou and V. Vassiliou, "Accurate detection of sinkhole attacks in IoT networks using local agents," in *Proc. Medit. Commun. Comput. Netw. Conf. (MedComNet)*, Jun. 2020, pp. 1–8.

[57] N. Jhanjhi, S. N. Brohi, and N. A. Malik, "Proposing a rank and wormhole attack detection framework using machine learning," in *Proc. 13th Int. Conf. Math., Actuarial Sci., Comput. Sci. Statist. (MACS)*, Dec. 2019, pp. 1–9.

[58] D. Airehrour, J. A. Gutierrez, and S. K. Ray, "SecTrust-RPL: A secure trust-aware RPL routing protocol for Internet of Things," *Future Gener. Comput. Syst.*, vol. 93, pp. 860–876, Apr. 2019.

[59] N. Sivasankari and S. Kamalakkannan, "Detection and prevention of man-in-the-middle attack in IoT network using regression modeling," *Adv. Eng. Softw.*, vol. 169, Jul. 2022, Art. no. 103126.

[60] K. M. Malik, A. Javed, H. Malik, and A. Irtaza, "A light-weight replay detection framework for voice controlled IoT devices," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 5, pp. 982–996, Aug. 2020.

[61] K. Rambus. (2022). *Industrial IoT: Threats and Countermeasures*. [Online]. Available: https://www.rambus.com/iot/industrial-iot/

[62] D. Yin, L. Zhang, and K. Yang, "A DDoS attack detection and mitigation with software-defined Internet of Things framework," *IEEE Access*, vol. 6, pp. 24694–24705, 2018.

[63] A. Qureshi, M. A. Qureshi, H. A. Haider, and R. Khawaja, "A review on machine learning techniques for secure IoT networks," in *Proc. IEEE 23rd Int. Multitopic Conf. (INMIC)*, Nov. 2020, pp. 1–6.

[64] L. S. Vailshery, "Number of IoT connected devices worldwide 2019–2021, with forecasts to 2030," *Retrieved*, vol. 8, p. 2021, Sep. 2022.

[65] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, Jul. 2015.

[66] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. Waltham, MA, USA: Morgan Kaufmann, 2011.

[67] I. H. Sarker, A. S. M. Kayes, S. Badsha, H. Alqahtani, P. Watters, and A. Ng, "Cybersecurity data science: An overview from machine learning perspective," *J. Big Data*, vol. 7, no. 1, pp. 1–29, Dec. 2020.

[68] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, vol. 2. Berlin, Germany: Springer, 2009.

[69] C. M. Bishop and N. M. Nasrabadi, *Pattern Recognition and Machine Learning*, vol. 4. Berlin, Germany: Springer, 2006.

[70] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[71] G. H. John and P. Langley, "Estimating continuous distributions in Bayesian classifiers," 2013, *arXiv:1302.4964*.

[72] S. Agrawal and J. Agrawal, "Survey on anomaly detection using data mining techniques," *Proc. Comput. Sci.*, vol. 60, pp. 708–713, Jan. 2015.

[73] M. Swarnkar and N. Hubballi, "OCPAD: One class naive Bayes classifier for payload based anomaly detection," *Exp. Syst. Appl.*, vol. 64, pp. 330–339, Dec. 2016.

[74] M. Usama, J. Qadir, A. Raza, H. Arif, K. A. Yau, Y. Elkhatib, A. Hussain, and A. Al-Fuqaha, "Unsupervised machine learning for networking: Techniques, applications and research challenges," *IEEE Access*, vol. 7, pp. 65579–65615, 2019.

[75] D. W. Aha, D. Kibler, and M. K. Albert, "Instance-based learning algorithms," *Mach. Learn.*, vol. 6, no. 1, pp. 37–66, Jan. 1991.

[76] S. Pokhrel, R. Abbas, and B. Aryal, "IoT security: Botnet detection in IoT using machine learning," 2021, *arXiv:2104.02231*.

[77] S. S. Keerthi, S. K. Shevade, C. Bhattacharyya, and K. R. K. Murthy, "Improvements to Platt's SMO algorithm for SVM classifier design," *Neural Comput.*, vol. 13, no. 3, pp. 637–649, Mar. 2001.

[78] Y. Liu and D. Pi, "A novel kernel SVM algorithm with game theory for network intrusion detection," *KSII Trans. Internet Inf. Syst. (TIIS)*, vol. 11, no. 8, pp. 4043–4060, 2017.

[79] E. M. Dovom, A. Azmoodeh, A. Dehghantanha, D. E. Newton, R. M. Parizi, and H. Karimipour, "Fuzzy pattern tree for edge malware detection and categorization in IoT," *J. Syst. Archit.*, vol. 97, pp. 1–7, Aug. 2019.

[80] H. Karimipour and V. Dinavahi, "On false data injection attack against dynamic state estimation on smart power grids," in *Proc. IEEE Int. Conf. Smart Energy Grid Eng. (SEGE)*, Aug. 2017, pp. 388–393.

[81] H.-S. Ham, H.-H. Kim, M.-S. Kim, and M.-J. Choi, "Linear SVM-based Android malware detection for reliable IoT services," *J. Appl. Math.*, vol. 2014, pp. 1–10, Jan. 2014.

[82] M. Goyal, I. Sahoo, and G. Geethakumari, "HTTP botnet detection in IoT devices using network traffic analysis," in *Proc. Int. Conf. Recent Adv. Energy-Efficient Comput. Commun. (ICRAECC)*, Mar. 2019, pp. 1–6.

[83] J. R. Quinlan, *C4. 5: Programs for Machine Learning*. Amsterdam, The Netherlands: Elsevier, 2014.

[84] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, Mar. 1986.

[85] L. Breiman, J. Friedman, R. Olshen, and C. Stone, "Classification and regression trees. Wadsworth int," *Group*, vol. 37, no. 15, pp. 237–251, 1984.

[86] G. Kim, S. Lee, and S. Kim, "A novel hybrid intrusion detection method integrating anomaly detection with misuse detection," *Exp. Syst. Appl.*, vol. 41, no. 4, pp. 1690–1700, Mar. 2014.

[87] M. Douiba, S. Benkirane, A. Guezzaz, and M. Azrour, "Anomaly detection model based on gradient boosting and decision tree for IoT environments security," *J. Reliable Intell. Environments*, vol. 9, no. 4, pp. 421–432, Dec. 2023.

[88] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.

[89] R. Doshi, N. Apthorpe, and N. Feamster, "Machine learning DDoS detection for consumer Internet of Things devices," in *Proc. IEEE Secur. Privacy Workshops (SPW)*, May 2018, pp. 29–35.

[90] Y. Chang, W. Li, and Z. Yang, "Network intrusion detection based on random forest and support vector machine," in *Proc. IEEE Int. Conf. Comput. Sci. Eng. (CSE) IEEE Int. Conf. Embedded Ubiquitous Comput. (EUC)*, vol. 1, Jul. 2017, pp. 635–638.

[91] Y. Meidan, M. Bohadana, A. Shabtai, M. Ochoa, N. O. Tippenhauer, J. D. Guarnizo, and Y. Elovici, "Detection of unauthorized IoT devices using machine learning techniques," 2017, *arXiv:1709.04647*.

[92] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, vol. 112. Berlin, Germany: Springer, 2013.

[93] M. Hasan, M. M. Islam, M. I. I. Zarif, and M. M. A. Hashem, "Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches," *Internet Things*, vol. 7, Sep. 2019, Art. no. 100059.

[94] A. Altaf, H. Abbas, F. Iqbal, and A. Derhab, "Trust models of Internet of Smart Things: A survey, open issues, and future directions," *J. Netw. Comput. Appl.*, vol. 137, pp. 93–111, Jul. 2019.

[95] E. Alpaydin, *Introduction to Machine Learning*, 2nd ed., Cambridge, MA, USA: MIT Press, 2010.

[96] R. U. Haque, A. T. Hasan, T. Nishat, and M. A. Adnan, "Privacy-preserving k-means clustering over blockchain-based encrypted iomt data," in *Advances in Blockchain Technology for Cyber Physical Systems*. Berlin, Germany: Springer, 2021, pp. 109–123.

[97] L. Rokach, *Data Mining and Knowledge Discovery Handbook*. Berlin, Germany: Springer, 2005.

[98] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "Network anomaly detection: Methods, systems and tools," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 303–336, 1st Quart., 2014.

[99] A. P. Muniyandi, R. Rajeswari, and R. Rajaram, "Network anomaly detection by cascading K-means clustering and C4.5 decision tree algorithm," *Proc. Eng.*, vol. 30, pp. 174–182, Jan. 2012.

[100] M. Xie, M. Huang, Y. Bai, and Z. Hu, "The anonymization protection algorithm based on fuzzy clustering for the ego of data in the Internet of Things," *J. Electr. Comput. Eng.*, vol. 2017, pp. 1–10, Jan. 2017.

[101] K. Teknomo, "K-means clustering tutorial," *Medicine*, vol. 100, no. 4, p. 3, 2006.

[102] S. Perveen, M. Shahbaz, K. Keshavjee, and A. Guergachi, "Metabolic syndrome and development of diabetes mellitus: Predictive modeling based on machine learning techniques," *IEEE Access*, vol. 7, pp. 1365–1375, 2019.

[103] Y. Hou, R. He, J. Dong, Y. Yang, and W. Ma, "IoT anomaly detection based on autoencoder and Bayesian Gaussian mixture model," *Electronics*, vol. 11, no. 20, p. 3287, Oct. 2022.

[104] I. H. Sarker, Y. B. Abushark, F. Alsolami, and A. I. Khan, "IntruDTree: A machine learning based cyber security intrusion detection model," *Symmetry*, vol. 12, no. 5, p. 754, May 2020.

[105] I. H. Sarker, Y. B. Abushark, and A. I. Khan, "ContextPCA: Predicting context-aware smartphone apps usage based on machine learning techniques," *Symmetry*, vol. 12, no. 4, p. 499, Apr. 2020.

[106] I. H. Sarker, H. Alqahtani, F. Alsolami, A. I. Khan, Y. B. Abushark, and M. K. Siddiqui, "Context pre-modeling: An empirical analysis for classification based user-centric context-aware predictive modeling," *J. Big Data*, vol. 7, no. 1, pp. 1–23, Dec. 2020.

[107] M. Mohy-eddine, A. Guezzaz, S. Benkirane, and M. Azrour, "An efficient network intrusion detection model for IoT security using K-NN classifier and feature selection," *Multimedia Tools Appl.*, vol. 82, no. 15, pp. 23615–23633, Jun. 2023.

[108] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[109] S. Zhao, W. Li, T. Zia, and A. Y. Zomaya, "A dimension reduction model and classifier for anomaly-based intrusion detection in Internet of Things," in *Proc. IEEE 15th Int. Conf Dependable, Autonomic Secure Comput., 15th Int. Conf Pervasive Intell. Comput., 3rd Int. Conf Big Data Intell. Comput. Cyber Sci. Technol. Congress(DASC/PiCom/DataCom/CyberSciTech)*, Nov. 2017, pp. 836–843.

[110] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*. Upper Saddle River, NJ, USA: Prentice-Hall, 2002.

[111] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *J. Artif. Intell. Res.*, vol. 4, pp. 237–285, May 1996.

[112] M. Mohammed, M. B. Khan, and E. B. M. Bashier, *Machine Learning: Algorithms and Applications*. Boca Raton, FL, USA: CRC Press, 2016.

[113] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Hoboken, NJ, USA: Wiley, 2014.

[114] A. S. Polydoros and L. Nalpantidis, "Survey of model-based reinforcement learning: Applications on robotics," *J. Intell. Robotic Syst.*, vol. 86, no. 2, pp. 153–173, May 2017.

[115] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, "Mastering the game of go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, Jan. 2016.

[116] S. M. Nagarajan, G. G. Deverajan, P. Chatterjee, W. Alnumay, and U. Ghosh, "Effective task scheduling algorithm with deep learning for Internet of Health Things (IoHT) in sustainable smart cities," *Sustain. Cities Soc.*, vol. 71, Aug. 2021, Art. no. 102945.

[117] O. Kayode and A. S. Tosun, "Deep Q-network for enhanced data privacy and security of IoT traffic," in *Proc. IEEE 6th World Forum Internet Things (WF-IoT)*, Jun. 2020, pp. 1–6.

[118] A. A. R. Melvin, G. J. W. Kathrine, S. S. Ilango, S. Vimal, S. Rho, N. N. Xiong, and Y. Nam, "Dynamic malware attack dataset leveraging virtual machine monitor audit data for the detection of intrusions in cloud," *Trans. Emerg. Telecommun. Technol.*, vol. 33, no. 4, p. 4287, Apr. 2022.

[119] E. Tsogbaatar, M. H. Bhuyan, Y. Taenaka, D. Fall, K. Gonchigsumlaa, E. Elmroth, and Y. Kadobayashi, "DeL-IoT: A deep ensemble learning approach to uncover anomalies in IoT," *Internet Things*, vol. 14, Jun. 2021, Art. no. 100391.

[120] S. Dwivedi, M. Vardhan, and S. Tripathi, "Distributed denial-of-service prediction on IoT framework by learning techniques," *Open Comput. Sci.*, vol. 10, no. 1, pp. 220–230, Aug. 2020.

[121] A. Verma and V. Ranga, "ELNIDS: Ensemble learning based network intrusion detection system for RPL based Internet of Things," in *Proc. 4th Int. Conf. Internet Things, Smart Innov. Usages (IoT-SIU)*, Apr. 2019, pp. 1–6.

[122] M. Ge, X. Fu, N. Syed, Z. Baig, G. Teo, and A. Robles-Kelly, "Deep learning-based intrusion detection for IoT networks," in *Proc. IEEE 24th Pacific Rim Int. Symp. Dependable Comput. (PRDC)*, Dec. 2019, p. 256.

[123] M. Hosseini and H. R. S. Borojeni, "A hybrid approach for anomaly detection in the Internet of Things," in *Proc. Int. Conf. Smart Cities Internet Things*, Sep. 2018, pp. 1–6.

[124] 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," *IEEE Access*, vol. 8, pp. 89337–89350, 2020.

[125] H. Alzahrani, M. Abulkhair, and E. Alkayal, "A multi-class neural network model for rapid detection of IoT botnet attacks," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 7, pp. 688–696, 2020.

[126] S. Fenanir, F. Semchedine, and A. Baadache, "A machine learning-based lightweight intrusion detection system for the Internet of Things," *Revue d'Intelligence Artificielle*, vol. 33, no. 3, pp. 203–211, Oct. 2019.

[127] A. Verma and V. Ranga, "Machine learning based intrusion detection systems for IoT applications," *Wireless Pers. Commun.*, vol. 111, no. 4, pp. 2287–2310, Apr. 2020.

[128] K. Chen, S. Zhang, Z. Li, Y. Zhang, Q. Deng, S. Ray, and Y. Jin, "Internet-of-Things security and vulnerabilities: Taxonomy, challenges, and practice," *J. Hardw. Syst. Secur.*, vol. 2, no. 2, pp. 97–110, Jun. 2018.

[129] A. Ghosh, A. Raha, and A. Mukherjee, "Energy-efficient IoT-health monitoring system using approximate computing," *Internet Things*, vol. 9, Mar. 2020, Art. no. 100166.

[130] E. M. Karanja, S. Masupe, and M. G. Jeffrey, "Analysis of Internet of Things malware using image texture features and machine learning techniques," *Internet Things*, vol. 9, Mar. 2020, Art. no. 100153.

[131] R. M. Shukla and S. Sengupta, "COP: An integrated communication, optimization, and prediction unit for smart plug-in electric vehicle charging," *Internet Things*, vol. 9, Mar. 2020, Art. no. 100148.

[132] E. Gelenbe, M. Nakip, D. Marek, and T. Czachorski, "Diffusion analysis improves scalability of IoT networks to mitigate the massive access problem," in *Proc. 29th Int. Symp. Model., Anal., Simul. Comput. Telecommun. Syst. (MASCOTS)*, Nov. 2021, pp. 1–8.

[133] C. K. Rath, A. K. Mandal, and A. Sarkar, "Microservice based scalable IoT architecture for device interoperability," *Comput. Standards Interfaces*, vol. 84, Mar. 2023, Art. no. 103697.

[134] M. Vishwakarma and N. Kesswani, "DIDS: A deep neural network based real-time intrusion detection system for IoT," *Decis. Analytics J.*, vol. 5, Dec. 2022, Art. no. 100142.

[135] N. Torres, P. Pinto, and S. I. Lopes, "Security vulnerabilities in LPWANs—An attack vector analysis for the IoT ecosystem," *Appl. Sci.*, vol. 11, no. 7, p. 3176, Apr. 2021.

[136] H. Tyagi, R. Kumar, and S. K. Pandey, "A detailed study on trust management techniques for security and privacy in IoT: Challenges, trends, and research directions," *High-Confidence Comput.*, vol. 3, no. 2, Jun. 2023, Art. no. 100127.

[137] H. Xiong, Z. Qu, X. Huang, and K.-H. Yeh, "Revocable and unbounded attribute-based encryption scheme with adaptive security for integrating digital twins in Internet of Things," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 10, pp. 3306–3317, Oct. 2023.

[138] B. Suresh and G. Shyama Chandra Prasad, "An energy efficient secure routing scheme using LEACH protocol in WSN for IoT networks," *Meas., Sensors*, vol. 30, Dec. 2023, Art. no. 100883.

[139] A. Falayi, Q. Wang, W. Liao, and W. Yu, "Survey of distributed and decentralized IoT securities: Approaches using deep learning and blockchain technology," *Future Internet*, vol. 15, no. 5, p. 178, May 2023.

[140] S. Von Solms and S. Furnell, "Human aspects of IoT security and privacy," in *Security and Privacy in the Internet of Things: Architectures, Techniques, and Applications*. Hoboken, NJ, USA: Wiley, 2021, pp. 31–55.

[141] P. Sanju, "Enhancing intrusion detection in IoT systems: A hybrid metaheuristics-deep learning approach with ensemble of recurrent neural networks," *J. Eng. Res.*, vol. 11, no. 4, pp. 356–361, Dec. 2023.

[142] R. Gasmi, S. Hammoudi, M. Lamri, and S. Harous, "Recent reinforcement learning and blockchain based security solutions for Internet of Things: Survey," *Wireless Pers. Commun.*, vol. 132, no. 2, pp. 1307–1345, Sep. 2023.

[143] V. Gugueoth, S. Safavat, and S. Shetty, "Security of Internet of Things (IoT) using federated learning and deep learning—Recent advancements, issues and prospects," *ICT Exp.*, vol. 9, no. 5, pp. 941–960, Oct. 2023.

[144] A. A. Laghari, A. A. Khan, R. Alkanhel, H. Elmannai, and S. Bourouis, "Lightweight-BIoV: Blockchain distributed ledger technology (BDLT) for Internet of Vehicles (IoVs)," *Electronics*, vol. 12, no. 3, p. 677, Jan. 2023.

[145] C. Zhonghua, S. B. Goyal, and A. S. Rajawat, "Smart contracts attribute-based access control model for security & privacy of IoT system using blockchain and edge computing," *J. Supercomput.*, vol. 80, no. 2, pp. 1396–1425, Jan. 2024.

[146] R. R. Irshad, S. Hussain, I. Hussain, J. A. Nasir, A. Zeb, K. M. Alalayah, A. A. Alattab, A. Yousif, and I. M. Alwayle, "IoT-enabled secure and scalable cloud architecture for multi-user systems: A hybrid post-quantum cryptographic and blockchain-based approach toward a trustworthy cloud computing," *IEEE Access*, vol. 11, pp. 105479–105498, 2023.

[147] M. G. Haricharan, S. P. Govind, and C. N. S. V. Kumar, "An enhanced network security using machine learning and behavioral analysis," in *Proc. Int. Conf. Advancement Technol. (ICONAT)*, Jan. 2023, pp. 1–5.

[148] K. Luo, "A distributed SDN-based intrusion detection system for IoT using optimized forests," *PLoS ONE*, vol. 18, no. 8, Aug. 2023, Art. no. e0290694.

[149] O. T. Modupe, A. A. Otitoola, O. J. Oladapo, O. O. Abiona, O. C. Oyeniran, A. O. Adewusi, A. M. Komolafe, and A. Obijuru, "Reviewing the transformational impact of edge computing on real-time data processing and analytic," *Comput. Sci. IT Res. J.*, vol. 5, no. 3, pp. 693–702, Mar. 2024.

[150] R. Yang, H. He, Y. Xu, B. Xin, Y. Wang, Y. Qu, and W. Zhang, "Efficient intrusion detection toward IoT networks using cloud–edge collaboration," *Comput. Netw.*, vol. 228, Jun. 2023, Art. no. 109724.

[151] A. Nazir, J. He, N. Zhu, A. Wajahat, F. Ullah, S. Qureshi, X. Ma, and M. S. Pathan, "Collaborative threat intelligence: Enhancing IoT security through blockchain and machine learning integration," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 36, no. 2, Feb. 2024, Art. no. 101939.

[152] M. Humayun, N. Tariq, M. Alfayad, M. Zakwan, G. Alwakid, and M. Assiri, "Securing the Internet of Things in artificial intelligence era: A comprehensive survey," *IEEE Access*, vol. 12, pp. 25469–25490, 2024.

[153] A. Salam, "Internet of Things for sustainability: Perspectives in privacy, cybersecurity, and future trends," in *Internet of Things for Sustainable Community Development: Wireless Communications, Sensing, and Systems*. Berlin, Germany: Springer, 2024, pp. 299–326.

ARUNKUMAR MUNISWAMY received the bachelor's degree from REC (Anna University Affiliated), Walajapet, Vellore, in 2005, and the master's degree from Vel Tech Rangarajan Dr. Sagunthala Research and Development Institute of Science and Technology, Avadi, Chennai, in 2012. He is currently a Research Scholar with the School of Computer Science Engineering and Information Systems, Vellore Institute of Technology, Vellore, Tamil Nadu, India. He has more than 14 years of teaching experience. He has published review papers in international conferences. His research interests include artificial intelligence, security, the Internet of Things, and machine learning.

R. RATHI received the B.Tech. degree from the University of Madras, Tamil Nadu, India, in 2003, the M.E. degree in computer science and engineering from Anna University, India, in 2006, and the Ph.D. degree from Vellore Institute of Technology University, Vellore, India. Currently, she is an Associate Professor with the School of Computer Science Engineering and Information Systems, Vellore Institute of Technology University. She has more than 15 years of teaching and research experience. She has published around 20 papers. Her research interests include rough sets, knowledge discovery databases, genetic algorithms, deep learning, machine learning, and federated learning.