Journal of Theoretical and Applied Information Technology
15th April 2023. Vol.101. No 7

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

CYBER SECURITY THREATS DETECTION AND PROTECTION USING MACHINE LEARNING TECHNIQUES IN IOT

MS. PRAGATI RANA1*, DR. B P PATIL2**

1*Research Scholar at Pimpri Chinchwad College of Engineering, Nigdi, Pune and

Assistant Professor at Army Institute of Technology, Dighi, Pune

2Principal, Army Institute of Technology, Dighi, Pune

Corresponding Author Email ID: 1*pragati.rana@gmail.com, 2** bp_patil@rediffmail.com

ABSTRACT

Recently, technology has enhanced itself to the 4th Industrial Revolution, with the Internet of Things (IoT), Edge computing, Computer safety, and along with Cyber-attacks are rapidly evolving. The quick proliferation of Internet of Things (IoT) devices and web in many shapes produces more data, posing cyber security risks. Detection and protection of cybersecurity threats is a significant concern in IoT. Machine Learning (ML) methods are widely regarded as one of the most promising solutions to address cyber security threats and provide security. Machine Learning (ML) methods are crucial in various cyber security applications. This study examines the literature on Cyber security threat detection and protection in IoT such as detection of spam, malware and intrusion over the previous ten years using machine-learning methods. The scope of Systematic Literature Review includes an in depth examination of the majority of ML trending methods in cyber security threat detection and protection in IoT. In recent years, increased Machine Learning techniques are used to solve four major cyber security issues namely identification of Intrusion, Android malware, Spam and Malware.

Keywords: *Cyber security, Threat Detection, Security Risks, Machine Learning, IOT*

1. INTRODUCTION

Internet of Things (IoT) devices and its usages are getting tremendously important in modern life. These devices found almost everywhere, including homes, workplaces, commercial complexes, educational institutes, airports, and many other areas and they provide safe and on-demand services. IoT devices make it easier for stakeholders to collaborate and understand business requisites& results. In addition to that, IoT analysis and processing of data improves industrial infrastructure productiveness and efficacy. IoT systems implement helpful technological advances in many fields. Several companies and dealers take up principles for the protection of connected device from malicious attacks. More privacy and safety concerns are report; when more devices are, connect with private networks and internet. The rigidity of the safety proneness corresponding with these devices are report by a number of real-world examples [1].

Although IoT provides excellent flexibility and scalability, its large size may indicate a safety disaster. The hazard to the individual and the network, global infrastructure's cyber security increases as per the number of devices a person connect. All devices are rapidly evolving over the global IoT network; however, they are prune to assaults and regard as weak areas. Hence, cyber security framework of the IoT verifies whether the mechanisms used securely and kept up well.

The IoT has created huge differences in end-users daily lives as a nascent technology and transformation. Individuals are carrying on their livings, studies and works in an IoT network, utilizing smart environments (at houses and in cities), e-Health, and transportation systems. For organisations or institutions, futuristic automation and industrial production, knowledge exchange and data management, smart, self-modifying mechanisms are getting increasingly desired [2]. IoT may cooperate along Wireless Sensor Networks, Radio Frequency

Identification, things, and networks in anyway, at any time, and everywhere due to significant advancements in telecommunication systems. In IoT development, cyber security is an unavoidable issue that must be solve. If the problem not dealt with properly, hackers will use the flaws and frailties of devices to misinterpret data or crush the system over the global IoT network. The Internet of Things' assaults and failures may outweigh its advantages. Traditional safety protocols and mechanisms are also ineffective due to inadequate scalability, integrity, and interoperability in existing devices. As a result, new technologies must be develop to confront safety, privacy, and dependability requisites of the Internet of Things [3] - [5].

ML is a concept that is related to AI, which is a new age area of knowledge that utilizes statistics, data mining, pattern recognition, and portending analysis to discover the models, make predictions, and decide from data[6]. This technology aids in the extraction of meaningful data from large and diverse data sources. It is data-driven and well suited to complicated and large-scale data sources, such as Big Data. To analyze an expanding dataset, like data from sensors or linked bodies, machine learning is used [7]. General goal of machine learning is evaluating data structure and incorporating the data structure into models that everyone can understand and use [8].The research focuses on machine learning (ML) and how it can be improve in terms of design, analysis, and implementation to respond to complicated tasks or issues [9]. ML plays a significant part in finding solution too many issues and help the network expand deprive of creating unpredicted problems. ML is a good way to improve the safety of the devices connected, with identification of harmful code assaults, maintaining privacy to avoid unauthorized locating, monitoring, performing power analysis, and detecting intrusion systems among other things [10]. The procedure of accessing individual information in order to disorder by exhibiting itself as a lawful user is termed as Phishing or brand spoofing. Exhibiting internet pages as legal and acting deceptive to obtain individual data is an example of phishing [11-13].

Worms, Trojan horses, and viruses are three main types of malicious software. A virus is a computer program that harms computer operations without the awareness of the user. A virus can harm computer files and the operating systems. In 1981, the 1st computer virus to disseminate via a floppy drive is termed as Elk Cloner [14]. A worm is a program, consistently duplicates itself and destroy the materials on the system or web. Unalike viruses and worms, the Trojan horse never replicates itself; instead, it appears as a lawful program that is activate in response to a certain operation or action [15, 16]. Spam email messages that are unwanted and unsolicited are another cyber security risk. When an email read, it consumes a lot of time, fills the mailbox, and becomes the source of Java applet execution. Spam calls, text, and video messaging are all examples of spam on mobile devices and networks [17-19]. Spam messages on Twitter, in the form of text, and spam films on YouTube, in the form of videos, are widely disseminated by spammers.

2. LITERATURE REVIEW

Lee et al. [20] developed IoT cyber security technologies & four-fold cyber risk management frameworks namely ecosystem, infrastructure, risk assessment, and performance layer. IoT cyber risks are recognise calibrated, given importance via Cyber risk assessment layer**.** The objective of IoT cybersecurity is to decrease cyber security risks for the firm and users by safeguarding IoT assets, privacy.

Al-Omari et al. [21] presented a smart tree-based system to anticipating and finding cyber-attacks that were efficient and effective. The main phases in machine learning were follow within the model, like data rescaling and encoding. The result shows that the introduced method gave outstanding efficiency and effectiveness.

Farooq et al. [22] gave several instances of how Machine Learning analytics may be use to improve cyber security monitoring and examining the optimal algorithms for regular cyber menaces. Machine learning-based analysis is a great way to produce context obtained from learning security occurrences and common behavioral guidelines, resulting I n a low number of false-positive security warnings.

Mohan et al. [23] presented a cybersecurity framework for Personal Medical Devices (PMDs). IoT allows the patient to move around more freely while also allowing improved monitoring of his medical status. The PMDs become part of the IoT for medical devices that

provide almost seamless communication capabilities.

Kozik et al. [24] presented a study showing cyber-related threats must be considered remarkably determinant points incorporated into the strategic investigation of infrastructure disorganizations, outcome assessment, and evaluation of system reliance. Challenges related to cyber security of Critical Infrastructures (CI) are review in this paper.

Rashid et al. [25] investigated an assault and abnormality identification technique formulated upon machine learning techniques (LR, SVM, DT, RF, ANN, KNN) for countering as well as reducing IoT cyber security threats in posh cities. However, as the number of intelligent city networks grows, so does the possibility of cyber-attacks and threats. Intelligent city IoT devices are attach with sensors connected to enormous cloud servers, disclosing them to harmful attacks and threats. Subsequently, it is desperate to formulate the procedure to stop corresponding attacks and safeguard IoT devices from crash.

Jenna et al. [26] gave out various security risks, and new cyber-attacks categorization in IoT-based health care infrastructure. Due to the complexity of the environment and nature of the deployed devices, IoT-based health care suffers from many security concerns that differ from other areas of methodology, motivations, and effects.

Kure et al. [27] presented a unified strategy that includes a clearest theory for property desperation, Machine-learning coordinators for risk prediction and the Comprehensive Assessment Model (CAM) to evaluate the efficacy of prevailing controls. Results reveal that machine learning classifiers perform exceptionally well in predicting various risk kinds, such as repudiation of service, cyber spying, crime wares.

Hiller et al. [28] discussed how variety of outlooks could influence the cybersecurity risk and suggested an ideology to envision the impression of law and policy on safety. Cyber security and defense against cyber threats are continuous issues; they require endured attentiveness from the public and private sectors. Apple, Facebook, and Twitter have all recently admitted to being attacked and have taken new security steps to protect their networks.

Mittal et al. [29] described the Cyber-Twitter framework, provided customer cyber security intelligence warnings utilizing openly accessible information from the Twitter. A Security Vulnerability Concept Extractor (SVCE) used for bringing out terms related to security vulnerabilities.

USING MACHINE LEARNING TO SECURE IOT

It is critical to talk about the aspects of IoT networks that make them vulnerable to security threats. An Internet of Things (IoT) network is a heterogeneous network made up of various types of devices that communicate with one another using various protocols. On the one hand, such a diverse network enables cross-platform communication between various interfaces, but on the other, it puts the network's security at risk. Furthermore, the IoT networks are less secure due to the widespread deployment and connectivity of devices. There are billions of devices connected to the internet, and the sensitive data from those devices is process on a cloud server that is easily breach. Despite the existence of edge computing, there is a requirement for standard protocols to secure.

3. IOT PRIVACY AND CHALLENGES

Accessing hazardous applications to IoT systems, sensitive data and cyber security concerns have increased as the result of lack of device updates and password changes. Such poor safety practices increase data flouts and other threats. Because of weak safety protocols and frameworks, most security experts regard IoT to be a sensitive area for cyber-attacks. Regardless of developing numerous safety measures to secure IoT devices from cyber-attacks, safety regulations are not formulated well [30].Hence, end-users never able to use precautionary action to obviate data breaches. Hackers created various types of malware, which infects IoT devices, since 2008 and devised, a number of spam schemes to lure people divulging important data [31].This resulted in high-profile hacks regularly compromising the privacy of corporate workstations and individual devices. Device manufacturers and safety professionals create a productive protective mechanism for prevention or neutralization of cyber threats if they suitably identify cyber threats.

For example, the Internet of Things is vulnerable to different authentication flows, which remains one of many, most significant security issues.

The authentication utilized is confine to protect a single threat, like Denial of Service (DoS) or replay assaults. Due to dangerous applications and natural diversity of information collected in the IoT domain, safety of the information is one of the most expose domains in IoT validation. Man in the middle is consider as the most prevalent assault, where the purpose of third-party hijack communication channel is to take off the specifications of the noticeable nodes engaged in network exchange. "Man in the middle" strongly oblige the bank server identify the transaction right since the antagonist never need to know the originality of the presumed victim [32].

Internet-enabled products have set off a subject for cybercriminals. As IoT industry, grow the amount of possible threats also increase which impacts on productivity, device safety, and privacy. Research findings showed that 90% of end users were not aware of IoT cybersecurity [33].

IoT Security

Organizations design efficient training programs to enhance safety consciousness that direct the consumers to generate strong passwords and bring them up to date regularly. Moreover, users are instruct to upgrade safety patches continuously and requested to stay away from spam emails, 3rdparty applications, and any other things that could compromise IoT security [34].

Threats to Validity

A threat is the likelihood of an action that compromises network security and has a negative effect. Active threats are define as intrusions in which the message's content is alter. While passive threats are intrusions, in which a perpetrator reads and records the messages for later use in a malicious manner. IoT networks, as previously mentioned, connect numerous devices together using wired or wireless technology. Therefore, a hacker can either physically accessed IoT device and steal data from the network, or they can attack the communication link and leak sensitive data by listening in. Additionally, the IoT devices lack the computational power necessary to support the sophisticated security framework. Thus, there is a requirement of security solution.

4. PROBLEM STATEMENT OF CYBER SECURITY ATTACKS &THREATS

Unauthorized access, deletion, and alterations of information are the feasible flouts and safety infringements on a computer or mobile device [35]. There are various ways to define cyber security. Attackers known to be innovative and adapt new techniques faster than protectors who identify and defend against penetrations, intrusions, and attacks [36]. According to Cisco's annual, report from 2018, more than half of all attacks resulted in losses of $500 million or more [37]. The purpose of cyber security is to secure confidential details, administration information and business reports from unlawful insights and exploitation .In addition, cyber security encloses safeguarding of software, tools, and equipment and ascertaining as well as guaranteeing the solitude and uprightness of documents secured from a variety of threats and assaults [38].

5. MAJOR CYBER SECURITY ISSUES

There are four vital cyber security issues, namely Intrusion, Spam and Malware detection, Android malware detection. The major cyber security issues are illustrate in figure.1.

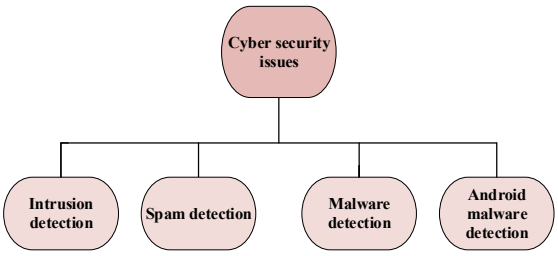

Figure.1 Major Cyber security issues

Intrusion Detection System

In modern life, networks are crucial, and cyber security has emerged as an experimenting field. The condition of software and hardware functioning in the network monitored by intrusion detection systems, which can be a vital cyber security tool. Still the prevailing intrusion identification system faces troubles in enhancing precision of detection, lowering false alarm rates, and detecting mysterious attacks despite decades of development. Numerous researchers have focused on developing intrusion detection systems that use machine-learning methods to overcome the issues mentioned in above. With high accuracy, machine-learning systems automatically detect the main differences

between ordinary and unusual data. Moreover, machine-learning techniques have robust universality to find the mysterious assaults. Anti-virus firewalls and intrusion detection systems are the general cyber security apparatuses. These methods safeguard the web from internal and external assaults. Among them, Intrusion detection systems are a kind of detection system that supervises the condition of software and hardware running in a network and helps to protect cyber security.

In 1980, the first intrusion detection scheme was present. Many mature Intrusion detection systems products have arisen since then. However, many Intrusion detection systems still have excessive false alarm rate, creating numerous alerts for small non-threatening circumstances, that increases the encumbrance for safety analysts and causes vigorously deleterious attacks that to be disregard. Therefore, several scholars have concentrated on developing intrusion detection system with elevated detection rates and lower false alarm rates. The next issue accompanied by traditional intrusion detection systems is their inability to find unknown assaults. As network domains changes rapidly, attack variations and unknown assaults appear constantly. Hence, Intrusion detection systems that can identify unknown threats must be develop. To solve the for mentioned issues, researchers have beg unto concentrate on developing Intrusion detection systems utilizing machine learning techniques, which is an artificial intelligence technique for automatically extracting valuable data from large datasets. When enough training data are available, machine learning-based Intrusion detection systems achieves reasonable sensing levels, and machine learning models have adequate universality to find assault variations and unknown assaults.Furthermore, machine learning-based IDSs never depend much on domain knowledge, making them easier to design and build[39].

Spam detection

Spam detection is an overseen machine learning issue. That indicates a person needs to furnish his machine-learning model accompanied by a collection of examples of spam, ham messages, and let it find the applicable model which discrete two varied types. Nearly all email providers have their own extensive data sets of labeled emails. For Spam detection and preventing spam, filtering email can be one of the most vital and noteworthy techniques. Many machine learning methodologies have used for this objective (Naïve Bayes, Random forest). There are many IoT-based social media platforms and applications. Because of the arrival of IoT, spamming issues increased at higher scale. Scholars developed a variety of spam identification methodologies to identify and filter spam and spammers. Spam detection includes social network spam detection. K.Lee et al noticed that Spammers utilize social systems for executing phishing assaults, spreading malware, as well as nurturing complementary websites [40]. A social argosy made to discover spammers in social networks like Twitter and Facebook in order to defend social systems from these attacks. The developed solution is about the Support Vector Machine, and it possess an elevated clarity and low positive rate. A social argosy speaks for a lawful user and a correlating bot, which collects legal, spam profiles, and reinforce them into the SVM classifier. To evaluate the capabilities of developed machine learning framework, the scholars looked into Myspace and Twitter. Many legal accounts were generate in both social networks and details were accumulate for months. Misleading spam profiles, specifically click traps, friend infiltrators, duplicate spammers, promoters and phishers signed out into various groupings. The data (Myspace: 388 legitimate profiles, 627 deceptive spam profiles; Twitter: 104 legitimate profiles, 61 spammers, 107 promoters' profiles) were fed into the SVM. The results show that spam precision is 70% and 82 percent for Myspace and Twitter [40].

Malware detection

Malware is a software that can be establish on a computer with the intent of disrupting its operation and causing damage to its electronic data. Viruses, worms, ransomware, adware, spyware, advertising, and Trojan horse are regard as the important kinds of malware. Malware disrupts natural operation of computers. Along the increasing use of computers and mobile devices, the cybercriminals find it simple to deal with the uprightness of information. The accessibility of computer and network resources are disrupt by Malware. Machine learning methodologies are attune to recognize malware [41].

Android system security mechanism

Google Play finds malicious software in Android application market by establishing the most up-to-date machine learning modules and technologies, efficiently recognizing bogus and spiteful software, and considerably boosting the safety of Google Play application software. China uses several 3rd party application stores instead of Google play, owing to policies, languages, business strategies and further purposes. Malicious software outspread blindly and varied safety threats appear one after the other, because the 3rd party app stores are not stringent in the application audit process. Hence, effective malware detection is a vital issue to be resolved in the android application store. Machine learning has long been attune to find malware on computers, but there is still a new application domain in mobile device using machine learning categorizing algorithms to categorize standard software and malware. The identification step includes: (1) Collect sufficient regular samples and malicious samples, analyse the sample software. Then extract the characteristics; (2) reduce the characteristics by feature selection; (3) split up the sample set into training and test sets, weigh up the functioning of various classifier algorithms, and select the best classifier; (4) utilize the best classifier to find hidden samples to decide if they are malicious software. The detection techniques are splitted into static and dynamic analysis based on the distinction between feature extraction and operation mode.

It is possible to the efficiently find malware by putting in machine learning for the identification of Android malware. Combining static and dynamic analysis enhances detection accuracy while also increases efficiency. However, because of the newest kind sand variations of Android malware, the version of Android system are constantly upgrade, therefore corresponding sensing technology should constantly upgraded, and enhanced [42].

6. MACHINE LEARNING IN CYBER SECURITY

Machine learning techniques are critical for recognizing and identifying IoT cyber; security threats. The significant issues in cyber security are Intrusion, malware and spam detection. The intrusion detection system aids to discover unlawful penetration or unofficial access with malicious intent. The Malware detection refers to the procedure of inspecting the computer and files to find out malware. Because it employs a variety of methods and approaches, it is effective at detecting malware. It is quite a complex procedure. The best thing is that malware identification and deletion take less than 50 seconds. Managing business email is an essential part in spam detection. With the amount of spam continuously rising, a spam detection tool helps to enlarge user productivity by removing unwanted messages and enhance system performance by keeping unwanted traffic off email servers.

For the major cyber security challenges, researchers use machine learning techniques: SVM, Nave Bayes, k-NN, RNN, and k-means [26].

7. CHALLENGES OF USING ML TECHNIQUES FORCYBER SECURITY

In the domain of Cyber security, machine-learning methodologies are widely used and has many challenges in it. Machine learning techniques require extensive data and high-performance resources while instructing the models. Using numerous GPUs (Graphics Processing Units) as one of the solutions can be neither energy-efficient nor cost-effective. Additionally, machine-learning techniques never designed to find cybercrime. There can be a necessity for powerful as well as strong machine learning techniques that are exclusively made for dealing with security assaults as well as controlling deleterious inputs.A noteworthy thing is that a single machine learning model will never be able to find out all types of security threats. There should be a specialised machine-learning model created to deal with a particular type of cyber-attack. Another challenging task is to prevent an attack from occurring at an early stage. ML techniques should be able to find real-time and zero-day assaults in a flash of time [26].

8. MACHINE LEARNING APPLICATIONS IN CYBER SECURITY ATTACKS DETECTION

ML strategies for IoT protection have advanced considerably in recent years. This segment is broken down into two sub-divisions. They are Machine-learning techniques and Machine learning based IoT security technologies [43].

Machine Learning (ML) techniques for Cyber security

Machine learning is a real m of AI, which unites a set of methodologies along with algorithms to supply intuitions for computers and smart devices. Machine-learning methodologies,

specifically supervised, unsupervised, and reinforcement learning has generally been accept in the network security environment. Used to precisely find and outline the exclusive security regulations that should impose the data plane. The ultimatum is to calibrate various frameworks corresponding safety protocols to alleviate a particular sort of assault, by naming network traffic or characterizing access control guidelines. Actually, many machine-learning techniques used to counter act IoT attacks. In particular, neural networks are attune to find network intrusion, DOS assaults and K-NN in identification of malware.

1) **Supervised Learning:** The general machine-learning method is supervise learning, which involves utilising a qualified data set and a learning algorithm to grade the output depending on the input. Classification and regression learning are two types of supervised learning.

2) **Unsupervised Learning:** Dissimilar to supervised learning, the model in unsupervised learning concealed, which means the data does not need to be label. Corresponding models attempt to notice a concurrence among the data and categorize it into several groups.

3) **Reinforcement Learning:** Reinforcement learning concentrates on problem study and methodologies, which attempt to revamp its model. It possess a distinctive model training method, it used trial, error and reward functions. It keeps track of the output's upshot and uses the advantage to calculate a value known as a "value function". The model perceives the correctness of its conclusion based on this value and adjusts itself accordingly[44].Machine learning techniques specifically, supervised, unsupervised, and reinforcement learning can be used to find novel assaults in IoT devices as well as developing a solid defence strategy.Fig.2 demonstrates IoT device security is achieved using various machine learning algorithms. Two examples of supervised learning are classification and regression learning. In unsupervised learning, there are no output data for those input variables. The majority of the data are unmarked, so the machine tries to find relationships between this data collection. They are group together as different class clusters. Additionally, by performing actions that maximise overall feedback, reinforcement learning enables the machine to learn from interactions with its environment in a manner similar to that of humans. Depending on how the assigned mission turned out, the feedback might be worth more. In reinforcement learning, there are no present behaviours for any specific task, and the system uses trial and error methods. The agent may discover and use the best strategy from its knowledge through trial and error in order to obtain the maximum reward [43].

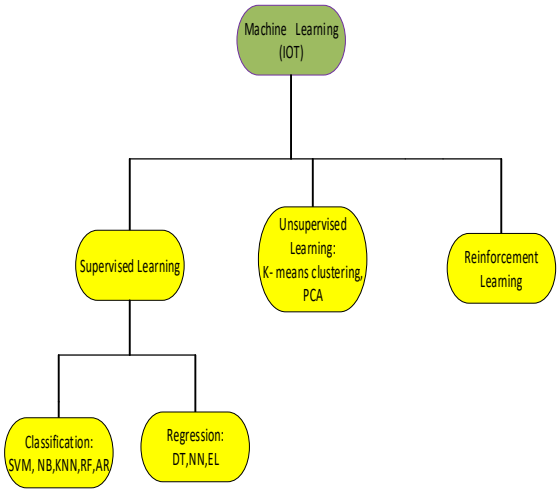

Figure.2. Machine learning and its classification

SUPPORT VECTOR MACHINE

Support Vector Machine generally used as successful machine learning technology for cyber security tasks, particularly Intrusion identification System. Support Vector Machine distinguishes and divides two data classes grounded on the annotation to the margin on either side of the hyperplane.Picture.3illustratessymbolized representation of Support Vector Machine. The fidelity of categorizing a data point is amplify, with increase in boundary and interspace among hyperplanes. The data points on the border of the hyperplane, known as support vector points. SVM is categorize into 2 significant types. Based on the kernel function, SVM is termed as linear or non-linear. Based on the detection type, SVM is termed as one-class and multi-class. SVM requires more time for training and processing memory. To learn about the dynamic user's behaviour, SVM requires training at various time intervals.

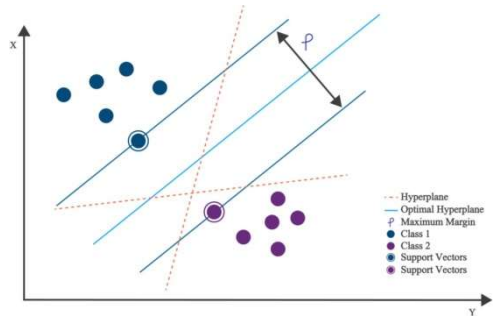

Figure.3. Support vector machine

Random Forest

Random forest is a supervised machine learning mechanism, which is, used widely for sorting as well as relapsing issues. It creates decision trees from various samples, classifies data using the majority vote, and determines relapse using the median. Due to the demand for memory space and computing power, random forest performance is sufficient for IoT. Random forest's efficacy with modern IoT data has proven, confirming that it is still the best algorithm for IoT cyber security.

NAIVE BAYES

Naïve Bayes is a sort of classifier based on Bayes' theorem, which deteriorates the constrained prospect of examining an issue. In cyber security; however, this independent state does not holds to all types of attacks. Several features in a dataset, such as those in KDD'99, are extremely reliant on one another. Hidden Naïve Bayes is an upgraded version that can handle problems with 99.6% accuracy. With discrete type attributes, the NB classifier performs well. This classifier considers more straightforward and possess a quicker detection speed.

K - Nearest Neighbor (KNN)

K-Nearest Neighbor was the simplified machine-learning algorithm grounded on supervised learning techniques, proposed by Fix and Hodges in 1951. In cyber security issues like intrusion and malware detection, K-NN was the second highest performing algorithm, closely behind Random forest with F1-scores within 3% of Random forest and accuracies of 91.48% and 89.80%, respectively. Researchers concluded Random forests and KNN's could all adequately classify and differentiate between normal and attacked data, and that most machine learning algorithms had average accuracy rates of approximately 90% in the area of IoT.

Machine learning-based solution for IoT based detection

Machine learning based safety solutions for IoT have developed as an emerging research area over the last few years, attracting the interest of today's researchers to contribute more to it. These solutions were researched using the three primary structural layers namely, the layers of Physical Perception, Network, and Web/Application. Traditional authentication mechanisms for securing the physical surface are insufficient due to the precise approach, value to find unwanted signals that cause false alarms. Therefore, ML-based learning techniques could be use as a physical layer authentication option [43].

Summary Of The Studies Classified Based On Machine Learning Algorithm

Table 1. Machine Learning Algorithms Classification

Author	year	Description	Dataset	Machine learning algorithms
Shafiq et al. [45]	2020	A novel Framework model and a hybrid algorithm was present to tackle the challenges of ML algorithms for cyber-attacks, uses the BoT-IoT identification dataset to assess whether ML algorithms are effectual.	Bot-IoT dataset	NB, Bayes Net, DT, RF

Roldán et al.[46]	2020	To detect many forms of IoT security attacks in real time, a smart architecture that combines CEP and Machine Learning (ML) was present. In particular, a smart architecture was proficient of easily governing event patterns whose states based on values secured by machine learning methodologies.	MQTT regular traffic packets	Support Vector Regression
Li et al. [47]	2020	This research examined the performance of DAS-CIDS in the domains of identification and false alarm depletion, utilizing both datasets in real network layouts	KDD99	KNN, SVM, RF, DT
Dovom et al.[48]	2019	In IoT, converts the programs' OpCodes into a vector space and uses unclear and fast fuzzy pattern tree algorithms for malware ident cation and categorization.	IoT, Vx-Heaven, Kaggle and Ransomware	Fuzzy Pattern Tree (FPT)
Soe et al. [49]	2020	Proposes the Correlated Set thresholding on gain-ratio (CST-GR) as a Novel Function selection methodology for building a delicate IDS based on machines utilising a fresh Feature selection algorithm.	Bot-IoT dataset from Cyber Range Lab	Logistic Model Tree, RF
Rashid et al. [50]	2020	This study explored an assault and deviation detection technique based on the algorithms for defending &mitigating IoT cyber risks in a smart city.	UNSW-NB15, CICIDS2017	SVM, DT, RF, KNN
Yakub Kayode Saheed et al. [51]	2022	This study focuses on machine learning supervised algorithm based IDS for IOT Attack detection.	UNSW-NB15	XGBoost, CatBoost, KNN, SVM, QDA and NB classifier
Tarek Gaber et al. [52]	2022	In this study, an intrusion detection is one way to detect injection attacks in IOT applications using Machine Learning algorithms.	Public dataset, AWID	SVM, Random Forest, and Decision Tree

Aldaej et al. [53]	2022	The proposed technique mitigates cybersecurity vulnerabilities while making the NoD protected and secure. For validation purposes, the suggested technique is test against a challenging dataset.	NS-KDD, KDD Cup 99	Random Forest, Decision Tree, Logistic Regression, Naïve Bayesian, SVM, MLP
Aslam et al. [54]	2022	The proposed framework utilizes machine learning algorithms in an adaptive multilayered feed-forwarding scheme to successfully detect the DDoS attacks by examining the static features of the inspected network traffic	environment-specific dataset	SVM, Naïve Bayesian, Random Forest, K-Nearest Neighbor, Logistic Regression, Ensemble Voting
Javed et al. [55]	2022	This paper presents an intelligent APT detection and classification system to secure I-IoT. After pre-processing, several machine-learning algorithms are apply to detect and classify complex APT signatures accurately.	KDDCup99	Decision Tree, Random Forest, SVM, Logistic Regression, Gaussian Naïve Bayesian, Extreme Gradient Boosting, and AdaBoosting

9. STRENGHTS OF MACHINE LEARNING IN IOT

When performing ML training with a sizable initial dataset, accurate results can obtained before subjecting the algorithm to classification tasks. Although there is many data from different types of devices available in IoT networks, there is not enough security-related data to be useful. Additionally, there is the problem of training each algorithm using some sensitive data. Therefore, a crowdsourcing platform must be create to generate various datasets for various security tasks. For the ML algorithms to be easily train, these datasets should include all authentication types and attack patterns. Testing classifiers on that dataset will also aid in establishing the standards for them. Additionally, patterns for new attacks should continuously monitored and added to datasets. Furthermore, only high-level data can used to train ML algorithms. However, IoT networks house most of the heterogeneous device data, along with some low-level data. This low-level data may be corrupted or noisy, which could have an impact on the ML model during training. Therefore, data that can sent to the ML model for real-time training should be filter.

10. CONCLUSION

This study provides a thorough evaluation of machine learning strategies for detecting and protecting cybersecurity threats in the IoT. Large data are constantly being create because of their faster development in various domains, necessitating higher attention to the privacy and security. Machine-learning techniques plays a significant part in several applications of the cyber safety systems. The literature of Cyber security threat detection and protection in IoT such as Intrusion, Spam and Malware detection over the previous ten years by using machine-learning techniques are examine. If these threats are successful, IoT performance are harm in several ways including, providing incorrect information. Traditional methods used to improve IoT security owing to the quicker advancement of cyber threats in the past. The existing literature on machine learning algorithms for detecting and defending cyber security threats in IoT systems are summarised and categorised. However, the SLR (Systematic Literature Review) confirms that ML techniques area propitious method for ensuring security and privacy in IoT domains.

11. FUTURE DIRECTIONS

There are some existing grey areas that need to be looking into and solutions found if the IoT

domain is to change the world in the upcoming years. Following is a summary of some of the areas that make up the IoT domain's future directions:

Intelligent Decision Making

Many Internet of Things (IoT) devices have created and put into use in various aspects of our lives up to this point, but there are still obstacles that must overcome before these devices can soon make decisions that are more intelligent. By incorporating artificial intelligence and machine learning into the IoT domain, intelligent IoT devices have the potential to transform a variety of decision-making processes.

Edge Computing

IoT is primary flaw is that it grows its device count behind the firewall of the network. IoT device security demands a lot more focus. The requirement to include security components between the network connection that connects to the devices and the software applications. It has been propose that edge computing could provide a remedy for the current IoT devices' slow data processing behaviour. All smart devices should process data more quickly to reduce communication latency between IoT devices. For the development of IoT, edge computing data processing is anticipate to increase.

Block chain Integration

Decentralization and self-governance are becoming more prevalent in a wide range of business, governmental, and consumer practises. The current ecosystems are vulnerable to exploitation because of the single points of failure, and DDoS attacks could bring down entire systems. All information sharing and communication between devices can based on an autonomous system by integrating the IoT environment with block chain technology. Block chain technology might offer time-stamped contractual handshakes that are approve between devices and secure documented transactions. According to IDC, up to 20% more high-quality products will be deliver by 2021 because one-third of retailers and manufacturers will be using block chain to track goods in advance of regulatory changes.

Better Security

IoT advancements will bring about more security issues. Finding novel ways to integrate security throughout the entire IoT ecosystem will require research. This means that security should prioritised at all levels, from the sensor/actuator level to the backend analytical engines.

REFERENCES

[1] Tawalbeh, Lo'ai, FadiMuheidat, MaisTawalbeh, and MuhannadQuwaider. "IoT Privacy and security: Challenges and solutions." Applied Sciences 10, no. 12 (2020): 4102.

[2] L. Xu, W. He, and S. Li, "Internet of Things in industries: a survey," IEEE Trans. Ind. Informat., vol. 10, no. 4, pp. 2233-2243, 2014.

[3] L. Atzori, A. Iera, and G. Morabito, "The internet of things: a survey," Comput. Netw. vol. 54, no. 15, pp. 2787–2805, 2010.

[4] D. Bandyopadhyay, and J. Sen, "Internet of things: applications and challenges in technology and standardization," Wireless Pers. Commun., vol.58, no.1, pp.49-69, 2011.

[5] R. Roman, J. Zhou, and J. Lopez, "On the features and challenges of security and privacy in distributed internet of things," Comput. Netw. vol. 57, no. 10, pp. 2266-2279, 2013

[6] Michalski, Ryszard S., Jaime G. Carbonell, and Tom M. Mitchell, eds. Machine learning: An artificial intelligence approach. Springer Science & Business Media, 2013.

[7] Ayodele, TaiwoOladipupo. "Machine learning overview." New Advances in Machine Learning. InTech, 2010.

[8] Thrun, Sebastian, and Lorien Pratt, eds. learning to learn. Springer Science & Business Media, 2012.

[9] Langley, Pat, and Herbert A. Simon. "Applications of machine learning and rule induction." Communications of the ACM 38.11 (1995): 54-64.

[10] Messaoud, Seifeddine, Abbas Bradai, Syed HashimRazaBukhari, Pham Tran AnhQuang, Olfa Ben Ahmed, and Mohamed Atri. "A survey on machine learning in internet of things: algorithms, strategies, and applications." Internet of Things 12 (2020): 100314.

[11] 11.T. N. Jagatic, N. A. Johnson, M. Jakobsson, and F. Menczer, "Social phishing," Communications of the ACM, vol. 50, no. 10, pp. 94-100, 2007

[12] R. M. Mohammad, F. Thabtah, and L. McCluskey, "Tutorial and critical analysis of phishing websites methods," Computer Science Review, vol. 17, pp. 1-24, 2015.

[13] E. H. Spafford, "Computer viruses as artificial life," Artificial life, vol. 1, no. 3, pp. 249-265, 1994.
[14] G. B. Shelly and M. E. Vermaat, "Discovering ComputersFundamentals 2011 Edition," 2010.
[15] G. B. Shelly, T. J. Cashman, and M. E. Vermaat, "Discovering computers," SalembaInfotek. Jakarta, 2012.
[16] H. Drucker, D. Wu, and V. N. Vapnik, "Support vector machines for spam categorization," IEEE Transactions on Neural networks, vol. 10, no. 5, pp. 1048-1054, 1999.
[17] N. Jindal and B. Liu, "Review spam detection," in Proceedings of the 16th international conference on World Wide Web, 2007: ACM, pp. 1189-1190.
[18] M. A. Shafi'I et al., "A review on mobile SMS spam filtering techniques," IEEE Access, vol. 5, pp. 15650-15666, 2017.
[19] D. D. Arifin and M. A. Bijaksana, "Enhancing spam detection on mobile phone Short Message Service (SMS) performance using FPgrowth and Naive Bayes Classifier," in 2016 IEEE Asia Pacific Conference on Wireless and Mobile (APWiMob), 2016: IEEE, pp. 80- 84.
[20] Lee, In. "Internet of Things (IoT) cybersecurity: Literature review and IoT cyber risk management." Future Internet 12, no. 9 (2020): 157.
[21] Al-Omari, Mohammad, MajdiRawashdeh, FadiQutaishat, Alshira'H. Mohammad and NedalAbabneh. "An Intelligent Tree-Based Intrusion Detection Model for Cyber Security." Journal of Network and Systems Management 29, no. 2 (2021): 1-18.
[22] Farooq, Hafiz M., and Naif M. Otaibi. "Optimal machine learning algorithms for cyber threat detection." In 2018 UKSim-AMSS 20th International Conference on Computer Modelling and Simulation (UKSim), pp. 32-37. IEEE, 2018.
[23] Mohan, Apurva. "Cyber security for personal medical devices internet of things." In 2014 IEEE international conference on distributed computing in sensor systems, pp. 372-374. IEEE, 2014.
[24] Kozik, Rafał, and MichałChoraś. "Current cyber security threats and challenges in critical infrastructures protection." In 2013 Second International Conference on Informatics & Applications (ICIA), pp. 93-97. IEEE, 2013.
[25] Rashid, MdMamunur, JoarderKamruzzaman, Mohammad Mehedi Hassan, Tasadduq Imam, and Steven Gordon. "Cyberattacks detection in iot-based smart city applications using machine learning techniques." International Journal of environmental research and public health 17, no. 24 (2020): 9347.
[26] Djenna, Amir, and DiamelEddineSaïdouni. "Cyber-attacks classification in IoT-based-healthcare infrastructure." In 2018 2nd Cyber Security in Networking Conference (CSNet), pp. 1-4. IEEE, 2018.
[27] Kure, Halima Ibrahim, Shareeful Islam, MustansarGhazanfar, AsadRaza, and Maruf Pasha. "Asset criticality and risk prediction for an effective cybersecurity risk management of cyber-physical system." Neural Computing and Applications (2021): 1-22.
[28] Hiller, Janine S., and Roberta S. Russell. "The challenge and imperative of private sector cybersecurity: An international comparison." Computer Law & Security Review 29, no. 3 (2013): 236-245.
[29] Mittal, Sudip, Prajit Kumar Das, VarishMulwad, Anupam Joshi, and Tim Finin. "Cybertwitter: Using twitter to generate alerts for cybersecurity threats and vulnerabilities." In 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 860-867. IEEE, 2016.
[30] Conti, M.; Dehghantanha, A.; Franke, K.; Watson, S. Internet of Things security and forensics: Challenges and opportunities. Future Gener. Comput. Syst. 2018, 78, 544–546. [CrossRef]
[31] Monther, A.A.; Tawalbeh, L. Security techniques for intelligent spam sensing and anomaly detection in online social platforms. Int. J. Electr. Comput. Eng. 2020, 10, 2088–8708.
[32] Conti, M.; Dragoni, N.; Lesyk, V. A survey of man in the middle attacks. IEEE Commun. Surv. Tutor. 2016, 18, 2027–2051. Available online: https://ieeexplore.ieee.org/abstract/document/7442758 (accessed on 10 April 2020). [CrossRef]
[33] Gemalto. Securing the IoT-Building Trust in IoT Devices and Data. 2020. Available online: https: //www.gemalto.com/:https://www.gemalto.co

m/iot/iot-security. (Accessed on 17 February 2020).
[34] Estrada, D.; Tawalbeh, L.; Vinaja, R. How Secure Having IoT Devices in Our Home. J. Inf. Secur. 2020, 11. [CrossRef]
[35] Shaukat, Kamran, SuhuaiLuo, Vijay Varadharajan, Ibrahim A. Hameed, and Min Xu. "A survey on machine learning techniques for cyber security in the last decade." IEEE Access 8 (2020): 222310-222354.
[36] Cisco 2018 Annual Cybersecurity Report," 2018. Accessed: December 25, 2019. [Online]. Available: https://www.cisco.com/c/m/en_au/products/security/offers/annualcybersecurity-report-2018.html
[37] E. A. Fischer, "Creating a national framework for cybersecurity: An analysis of issues and options," 2005: LIBRARY OF CONGRESS WASHINGTON DC CONGRESSIONAL RESEARCH SERVICE.
[38] J. M. Torres, C. I. Comesaña, and P. J. García-Nieto, "Machine learning techniques applied to cybersecurity," International Journal of Machine Learning and Cybernetics, pp. 1-14, 2019.
[39] Liu, Hongyu; Lang, Bo (2019). Machine Learning and Deep Learning Methods for Intrusion Detection Systems: A Survey. Applied Sciences, 9(20), 4396–. Doi: 10.3390/app9204396
[40] K. Lee, J. Caverlee, and S. Webb, "Uncovering social spammers: social honeypots + machine learning", SIGIR'10, July 19-23, 2010, Geneva, Switzerlan
[41] Kamran Shaukat; SuhuaiLuo; ShanChen; Dongxi Liu; (2020). Cyber Threat Detection Using Machine Learning Techniques: A Performance Evaluation Perspective. 2020 International Conference on Cyber Warfare and Security (ICCWS), (), –. doi:10.1109/iccws48432.2020.9292388
[42] Qing-Fei, Wang; Xiang, Fang (2018). [IEEE 2018 4th Annual International Conference on Network and Information Systems for Computers (ICNISC) - Wuhan, China (2018.4.19-2018.4.21)] 2018 4th Annual International Conference on Network and Information Systems for Computers (ICNISC) - Android Malware Detection Based on Machine Learning. , (), 434–436. doi:10.1109/ICNISC.2018.00094
[43] Haji, SaadHikmat, and Siddeeq Y. Ameen. "Attack and anomaly detection in iot networks using machine learning techniques: A review." Asian journal of research in computer science 9, no. 2 (2021): 30-46.
[44] M. Bagaa, T. Taleb, J. B. Bernabe and A. Skarmeta, "A Machine Learning Security Framework for Iot Systems," in *IEEE Access*, vol. 8, pp. 114066-114077, 2020, doi: 10.1109/ACCESS.2020.2996214.
[45] Shafiq, M.; Tian, Z.; Sun, Y.; Du, X.; Guizani, M. Selection of effective machine learning algorithm and Bot-IoT attacks traffic identification for internet of things in smart city. Futur. Gener. Comput. Syst. 2020, 107, 433–442.
[46] Roldán, J.; Boubeta-Puig, J.; Luis Martínez, J.; Ortiz, G. Integrating complex event processing and machine learning: An intelligent architecture for detecting IoT security attacks. Expert Syst. Appl. 2020, 149, 113251.
[47] Li, W.; Meng, W.; Au, M.H. Enhancing collaborative intrusion detection via disagreement-based semi-supervised learning in IoT environments. J. Netw. Comput. Appl. 2020, 161, 102631.
[48] Dovom, E.M.; Azmoodeh, A.; Dehghantanha, A.; Newton, D.E.; Parizi, R.M.; Karimipour, H. Fuzzy pattern tree for edge malware detection and categorization in IoT. J. Syst. Archit. 2019, 97, 1–7.
[49] Soe, Y.N.; Feng, Y.; Santosa, P.I.; Hartanto, R.; Sakurai, K. Towards a lightweight detection system for cyber attacks in the IoT environment using corresponding features. Electronics 2020, 9, 144.
[50] Rashid, M.M.; Kamruzzaman, J.; Hassan, M.M.; Imam, T.; Gordon, S. Cyberattacks detection in iot-based smart city applications using machine learning techniques. Int. J. Environ. Res. Public Health 2020, 17, 9347.
[51] Yakub Kayode Saheed, Aremu Idris Abiodun, Sanjay Misra, Monica Kristiansen Holone, Ricardo Colomo-Palacios, A machine learning-based intrusion detection for detecting internet of things network attacks, Alexandria Engineering Journal, Volume 61, Issue 12, 2022, ISSN 1110-0168, https://doi.org/10.1016/j.aej.2022.02.063.
[52] Tarek Gaber, Amir El-Ghamry, Aboul Ella Hassanien, Injection attack detection using machine learning for smart IoT applications, Physical Communication, Volume 52, 2022, 101685, ISSN 1874-4907,

https://doi.org/10.1016/j.phycom.2022.101685.

[53] Aldaej, Abdulaziz, Tariq Ahamed Ahanger, Mohammed Atiquzzaman, Imdad Ullah, and Muhammad Yousufudin. 2022. "Smart Cybersecurity Framework for IoT-Empowered Drones: Machine Learning Perspective" *Sensors* 22, no. 7: 2630. https://doi.org/10.3390/s22072630

[54] Aslam M, Ye D, Tariq A, Asad M, Hanif M, Ndzi D, Chelloug SA, Elaziz MA, Al-Qaness MAA, Jilani SF. Adaptive Machine Learning Based Distributed Denial-of-Services Attacks Detection and Mitigation System for SDN-Enabled IoT. *Sensors*. 2022; 22(7):2697. https://doi.org/10.3390/s22072697

[55] Javed, Safdar Hussain, Maaz Bin Ahmad, Muhammad Asif, Sultan H. Almotiri, Khalid Masood, and Mohammad A. Al Ghamdi. 2022. "An Intelligent System to Detect Advanced Persistent Threats in Industrial Internet of Things (I-IoT)" *Electronics* 11, no. 5: 742. https://doi.org/10.3390/electronics11050742