International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT228654

A Literature Review on Machine Learning for Cyber Security Issues

Jay Kumar Jain[1], Akhilesh A. Waoo[1], Dipti Chauhan[2]
[1]Department of Computer Science and Engineering, AKS University, Satna, Madhya Pradesh, India
[2]Professor, Department of Computer Science and Engineering, PIEMR, Indore, Madhya Pradesh, India

Article Info

Publication Issue :
Volume 8, Issue 6
November-December-2022

Page Number : 374-385

Article History
Accepted: 20 Nov 2022
Published: 04 Dec 2022

ABSTRACT

Through the use of relevant data to build an algorithm, machine learning primarily aims to automate human help. A subset of artificial intelligence (AI), machine learning (ML) focuses on the development of systems that can learn from past data, recognize patterns, and reach logical conclusions with little to no human involvement. The concept of cyber security involves guarding against hostile attack on digital systems such computers, servers, mobile devices, networks, and the data they are connected to. Accounting for cyber security where machine learning is used and using machine learning to enable cyber security are the two main components of combining cyber security and ML. We may benefit from this union in a number of ways, including by giving machine learning models better security, enhancing the effectiveness of cyber security techniques, and supporting the efficient detection of zero day threats with minimal human involvement. In this review paper, we combine ML and cyber security to talk about two distinct notions. We also talk about the benefits, problems, and difficulties of combining ML and cyber security. In addition, we explore several attacks and present a thorough analysis of various tactics in two different categories. Finally, we offer a few suggestions for future research.

Keywords- Cyber-security issues, machine learning, algorithms, detection.

I. INTRODUCTION

Most of the devices we use in the modern era of computers are now part of an Internet of Things (IoT) ecosystem that is connected to the Internet. The Internet, a form of unsecure (open) communication, is the conduit through which these devices exchange and transfer their data [1]. Usually, these are sensitive pieces of information (i.e., healthcare data, insurance data, banking data, other finance related data, and social security numbers). Online attackers (hackers), for example, are continuously on the lookout for ways to fool around (for example, they can launch attacks like man-in-the-middle, replay credential guessing, impersonation, malware insertion, session key computation, and data manipulation) [2].

As a result, numerous researchers periodically recommend different security measures to lessen

these dangers. Security protocols or cyber security protocols can be divided into the following categories: authentication protocols, access control protocols, intrusion detection protocols, key management protocols, and security techniques using block chains. A summary of these protocols is provided below.

- Authentication protocols: Authentication is the process of determining whether a person or a device is who they claim to be. It is possible to execute it using credentials or factors that are directly related to the users or device (for example, username, password, smartcard, and biometrics). User to user, user to device, or device to device authentication are all options. User authentication protocols can also be broken down into three groups based on the factors that are available: one-factor user authentication protocol, two-factor user authentication protocol, and three-factor user authentication protocol.
- Protocols for access control: Restricting someone's or something's unauthorized access is done through access control (s). Once all steps of a user/device access control protocol have been followed, users or devices can securely access one another. Protocols for key management, intrusion detection, and block chain enabled security make up access control protocols. Below is a summary of these protocols.
- Authentication protocols: Authentication is the process of determining whether a person or a device is who they claim to be. It is possible to execute it using credentials or factors that are directly related to the users or device (for example, username, password, smartcard, and biometrics). User to user, user to device, or device to device authentication are all options. User authentication protocols can also be broken down into three groups based on the factors that are available: one-factor user authentication protocol, two-factor user authentication protocol, and three-factor user authentication protocol.
- Access control protocols: The process of limiting a person's or a device's unauthorized access (s). After all phases of a user/device access control protocol have been followed, users or devices can securely access other users or devices. The two types of access control protocols are: (1) user access control and (2) device access control. Device access control protocol can be used to control access to unauthorized devices, whereas user access control protocol can be used to control access to unauthorized users. Access control methods include certificate-based and certificate-less approaches. An entity (a client) is deemed to have permission to utilize a resource through the process of authorization, which is carried out by an authority (a server). In order for the server to identify the client making the access request, it is typically done in conjunction with authentication. It establishes who is permitted
- Protocols for detecting intrusions: An intrusion is anything or anyone that has malign purpose. This could be a hacker-controlled system that attacks the Internet or a malicious computer script. Typically, hackers attempt to introduce malware into online devices to degrade their performance or compromise their security (systems). We require a certain class of protocols that fall under the heading of "intrusion detection protocols" for the detection and mitigation of intrusion. The intrusion detection can be carried out in a variety of ways, including anomaly-based intrusion detection, hybrid intrusion detection, and signature-based and anomaly-based intrusion detection. Malware detection techniques based on machine learning or deep learning are becoming increasingly popular these days.
- Key management protocols: These protocols are used to securely manage keys between different entities, including some devices (such as Internet of Things (IoT) devices and smart vehicles) and certain people (smart home user, doctor traffic inspector). Typically, a trusted registration

authority registers all of the communication system's entities and then stores the secret credentials (i.e., secret keys) in their memory. For the objectives of creating new keys, storing them in devices, establishing keys, and revoking keys, we require a key management method. After establishing a shared secret key (also known as a session key), which may be accomplished through the crucial phases of an authenticated key agreement protocol, the devices/users can transmit their information in a secure manner [28].

- Security protocols that use block chain technology: Block chain technology is one of the era's rising technologies. Data is kept on a block chain in the form of specific blocks that are linked together using hash values. Block chain uses distributed ledger technology, also known as distributed ledger technology, to maintain data (DLT). The DLT is accessible to all legitimate participants (and occasionally miners) in the network. The data that we store on the blockchain is protected from a variety of potential cyberattacks. Consequently, the security mechanisms for blockchains can protect against a variety of cyberattacks. [3]. Computing systems learn from data and utilize algorithms to carry out tasks without being explicitly programmed. This process is known as machine learning (ML). AI's deep learning (DL) subfield is a form of machine learning (ML). A complex set of algorithms that are based on the human brain underlie deep learning (DL). This enables the processing of unstructured data, including text, images, and documents. ML describes a computer's capacity to reason and act independently of human intervention. However, DL often requires less constant human assistance. As a result, it performs image, video, and unstructured

Combining machine learning and cyber security can benefit us in a number of ways. For instance, improved cyber security techniques, increased machine learning model security, and more effective zero day attack detection with less human involvement. However, it might encounter a variety of issues and security obstacles, which should be handled with caution. We therefore need a review study in this particular field that addresses the "uniting of cyber security and machine learning," i.e., issues and challenges, various attacks, different protection schemes with a comparative study, and some future research directions on which other researchers should concentrate. Therefore, we made an effort to do comparable research in the study that was suggested [4,5].

II. RELATED WORK

Samson Ho and other people [6] The major objective of this paper is to propose a Convolutional Neural Network-based Intrusion Detection System (IDS) for enhancing internet security. The proposed IDS paradigm categorizes all network packet traffic into benign and malicious kinds in order to detect network intrusions. The Canadian Institute for Cybersecurity's CICIDS2017 dataset was used to train and validate the suggested model. All aspects of the model have been evaluated, including overall accuracy, attack detection rate, false alarm rate, and training overhead. Nine other well-known classifiers have been compared with the recommended model to see how effective they are.

Praneeth Narisetty and coworkers [7] On the basis of the most recent CICIDS2017 dataset, assist vector machines, ANNs, CNNs, Random Forests, and substantial learning estimations have all received moderate evaluations. Significant learning estimation fared worse than SVM, ANN, RF, and CNN. In the end, we will use this dataset to conduct port scope attacks similar to other attack kinds by combining AI and substantial learning calculations with Apache Hadoop and shimmer developments. This study determines which algorithm has the best accuracy rates for predicting the best outcomes to ascertain

whether or not a cyberattack occurred by predicting the four algorithms like SVM, ANN, RF, and CNN.

Emrah Tufan et al. [7], (Member, IEEE) In this study, network intrusion attempts are investigated using anomaly-based machine learning models, which offer superior security than the traditional misuse-based approaches. A data set acquired from a real-world, institutional production setting was used to build and apply two models: an ensemble learning model and a convolutional neural network model. The models were used with the UNSW-NB15 benchmarking data set to show their validity and dependability. To make the scope of the study modest, the sort of assault was restricted to probing attacks. The CNN model was marginally more accurate, according to the results, which showed high accuracy rates.

Iqbal H. Sarker et al. [8] introduce an intrusion detection tree ("IntruDTree") machine-learning-based security model in this article. This model takes into account the importance ranking of security features before creating a tree-based generalized intrusion detection model based on the important features that have been selected. This model is useful in terms of prediction accuracy for test cases that have not yet been seen since it reduces the computational complexity of the model by reducing the feature dimensions. The performance of our IntruDTree model was then evaluated utilizing cybersecurity datasets, and the scores for precision, recall, fscore, accuracy, and ROC were computed.

To assess the efficacy of the resulting security model, we also compare the outcome outcomes of the IntruDTree model with a number of conventionally well-liked machine learning techniques, including the naive Bayes classifier, logistic regression, support vector machines, and k-nearest neighbor.

Mohamed M. and others' [9], This article suggests using IDS with two layers. The first layer categorizes the network connection based on the service being used. Following that, a minimal set of features that improve the detection precision of malicious activity on that service are found. The second layer uses those features to categorize each network connection as an attack or regular activity using the pattern recognition technique. The normal behavior model and the attack behavior model are two multivariate normal statistical models that are produced during the training phase. The two multivariate normal statistical models are employed in the testing and operating phases to classify a network connection into attack or normal activity using a maximum likelihood estimate function. The experimental findings demonstrate the suggested IDS's advantage for network intrusion detection over comparable IDSs. It successfully achieves DR of 97.5%, 0.001 FAR, MCC 95.7%, and 99.8% overall accuracy using just four characteristics.

III. UNITING CYBER SECURITY AND MACHINE LEARNING

3.1 Machine learning in cyber security

Attacks such as replay, man-in-the-middle (MiTM), impersonation, credentials leakage, password guessing, session key leakage, unauthorized data update, malware injection, flooding, denial of service (DoS) and distributed denial of service (DDoS), among others, can be carried out against connected systems in the cyberspace. Therefore, in order to recognize and stop these attacks, we need some sort of security standard. Through the provided pre-processed dataset, the machine learning models (machine learning ML algorithms) can learn about various cyber-attacks in the offline/online mode. The machine learning algorithms identify any indication of intrusion (such as a cyberattack) in real time, or in online mode. Figure 1 shows the scenario of "machine learning in cyber security." In this case, a system that is connected to the Internet (such as a laptop, desktop, smartphone, or IoT device) can be used to carry out a variety of online operations, such as financial transactions, online access to healthcare data, social

security numbers, etc. Hackers are constantly looking for weaknesses in these systems, and when they find one, they launch an attack. Depending on the context, several ML techniques, including as supervised learning, unsupervised learning, reinforcement learning, and deep learning, can be utilized for the detection and mitigation of cyber-attacks. Whether supervised learning, unsupervised learning, reinforcement learning, or deep learning is the technique that best suits a system depends on the communication environment and resources that are accessible to it. Because cloud servers have good processing and storage capacity, it is possible to learn about (train) and forecast (test) cyberattacks using them.

3.2 Cyber security in machine learning

Figure 2 presents a scenario for "cyber security in machine learning," also known as machine learning (ML) security. For the analysis and forecasting of numerous events, ML models are employed. However, certain attacks, such as model poisoning disruption attacks and dataset poisoning attacks, can negatively impact the performance of ML models [6]. These attacks may cause machine learning (ML) models to forecast the associated phenomena incorrectly. The "dataset poisoning attack" involves the introduction of adversarial examples (updated values) into the dataset by an attacker, which leads the ML model to make incorrect predictions. The attacker's goal in the "model poisoning attack" is to corrupt the models by meddling with their internal operations and changing their settings. The attacker seeks to retrieve the model's useful information while simultaneously working to expose sensitive data during a "privacy breach attack." A privacy breach includes a membership inference attack. Additionally, in a "runtime disruption attack," the attacker subverts the ML workflow by assaulting the model's execution process, which has an impact on the accuracy of the prediction outcomes. Therefore, in order to defend against these attacks, there is a need for specific cyber security methods (such as encryption techniques, signature generation and verification techniques, and hashing processes). The ML models and the related datasets are secured under the use of these cyber security procedures, and the predicted results are accurate.

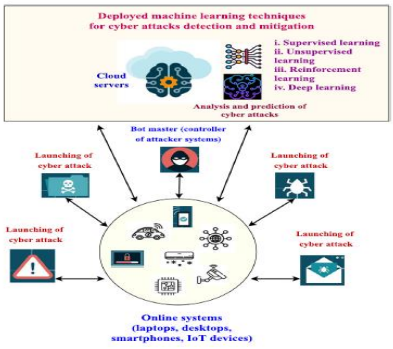

Fig. 1. Scenario of machine learning in cyber security.

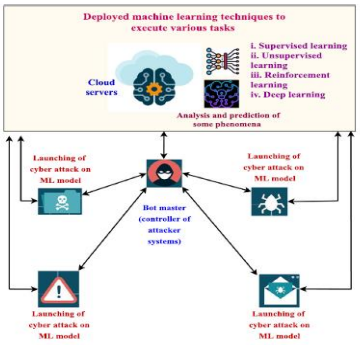

Fig. 2. Scenario of cyber security in machine learning.

IV. ADVANTAGES OF UNITING CYBER SECURITY AND MACHINE LEARNING

Machine learning and cyber security are both crucial for one another and can enhance each other's performance. The following are some advantages of their union.

• ML models' complete evidence of security: The ML models are susceptible to a variety of attacks, as was previously addressed. The prevalence of these attacks may have an impact on how well the ML models function, perform, and predict. However, these unpleasant occurrences can be prevented by implementing specific cyber security measures. The functioning, performance, and input datasets of the ML models are safeguarded under the use of cyber security procedures, and we obtain accurate predictions and outcomes [7].

• Enhanced effectiveness of cyber security methods: When we use ML algorithms to cyber security plans (such as intrusion detection systems), their efficacy is increased (i.e., improved accuracy and detection rate with less false positive rate). According to the communication environment and the associated systems, ML techniques including supervised learning, unsupervised learning, reinforcement learning, and deep learning algorithms can be applied.

• Good zero-day attack detection: ML-based cyber security systems that identify infiltration appear to be particularly effective at detecting zero-day attack (i.e., unknown malware attacks). They use certain deployed ML models to carry out the detection, which is why it occurs. The ML mode works by gathering and matching specific features; if the features of a program match those of a malicious program, then that program can be regarded as malicious. The ML models are capable of carrying out this detecting operation automatically. Thus, by combining cyber security with machine learning, it is possible to identify zero day threats effectively.

• Limited need for human intervention: In ML-based systems, deployed ML models handle the majority of the tasks. When we combine ML and cyber security, most of the jobs that these systems are used for are completed either entirely or largely without the assistance of humans.

• Rapid scanning and mitigation: Because they use specific ML algorithms, ML-based intrusion detection systems are particularly effective at detecting the presence of threats. As a result, the combination of machine learning with cyber security systems performs intrusion screening very quickly and also offers quick response in case of any sign of intrusion. The selection of an appropriate ML algorithm is the only thing we need to be concerned with.

IV. OVERVIEW OF VARIOUS THREATS AND ATTACKS

We describe the following attacks in detail in this section since they can happen in different computing environments.

• Eavesdropping is a passive form of attack sometimes referred to as sniffing or snooping attack. In this attack, a foe tries to overhear the communicating parties' private discussion.

• The attack is passive in nature, according to traffic analysis. In this attack, adversary A intercepts the discussion in progress before examining the messages to gather data about the communication's nature, pattern, and behavior as well as its location and timing. The info that was intercepted also aids A in carrying out related cyberattacks.

• Replay attack: In this attack, A purposefully retransmits the intercepted messages that were previously sent. A does this to deceive or mislead the recipient and coerces the lawful users into acting in accordance with A's wishes.

• Man-in-the-middle (MiTM) attack: This active attack involves establishing separate connections with interacting entities and relaying the messages to both ends. In these circumstances, the two communicating entities believe they are speaking directly to one another. Therefore, without being noticed, a may intercept, remove, change, or insert new information for transmission [8].

• Impersonation attack: This type of attack is similarly current in nature. A impersonates one of the network's legitimate parties by determining its identity, and then sends modified or some brand-new messages on that party's behalf to the other legal party.

• Denial-of-Service (DoS) assault: In a DoS attack, A floods the victim's computational resources with a large number of fictitious requests (i.e., HTTP flood messages). As a result, the valid user's service request cannot be handled. In this case, the network's services are inaccessible to the genuine user. Another type of DoS assault is referred to as a distributed denial-of-service (DDoS) attack, in which A utilizes a network of devices (a botnet) to send several requests to the victim's workstation at once, quickly using up all of the system's processing capabilities. DoS or DDoS attack can be carried out using many flooding

techniques, such as SYN flooding, HTTP flooding, UDP flooding, etc.

• Malware attack: These attack are carried out by having harmful scripts executed on the victim's computer. Malware that has been introduced or installed is a file or a piece of code that carries out unauthorized operations on a system, such as data theft, illegal drive or data encryption, data manipulation, or data deletion. Keyloggers, spyware, viruses, ransomware, worms, trojan horses, and other sorts of malware are examples [6].

• Scripting attack: These attack involve the release of data from an online database that is kept up by a web server (i.e., online banking database). For instance, "password cracking," "SQL injection attack," and "cross-site scripting (XSS) attack" can be used to obtain sensitive data from the system, such as passwords, credit card information, and debit card details.

• Privileged insider attack: Any privileged user of the system who has access to the registration data of various users and devices can carry out this attack. Due to the insider's access to the sensitive data, this attack is far more difficult to fight against and has a more negative effect.

• Actual theft of smart devices: In the modern world, the majority of computing environments are run using smart devices, such as smart household appliances, smart healthcare devices, and smart manufacturing machines. No physical security is used when deploying the smart devices. If an adversary A physically steals these smart devices, they can be utilized to harvest sensitive data by employing power analysis attack. After sensitive data has been extracted, illegal operations like computing session keys without authorization can be carried out [9].

• Birthday attack: This sort of cryptographic attack uses the birthday problem's mathematical foundations, which can be found in probability theory. Birthday attacks can be used for malicious objectives like credential guessing (passwords). This attack, which is predicated on a fixed degree of permutations and a higher likelihood of collisions between random attack attempts, is detailed in the birthday paradox. The possibility that any paired individuals in a group of n randomly chosen individuals will have a birth date is addressed by the birthday paradox (also known as the birthday dilemma). The birthday attack, a well-known cryptographic attack that employs this probabilistic tactic to lessen the difficulty of breaking a hash function, was inspired by the mathematics underlying this problem [10].

• Dictionary attack: A dictionary attack on a cryptographic system is a specific kind of malicious brute force attack. The attacker tries to get past the system's security by methodically entering every word in a dictionary as a password. It's also possible to use a dictionary attack to determine the key required to decrypt an encrypted message or document. A library of phrases or keywords that is maintained current is used by the attacker to try to bypass encryption or get access. Automatic word insertion into the target can be done using dictionary terms or numerical sequences. By using ineffective password strategies, such as changing passwords to ones that include consecutive numbers, symbols, or letters, dictionary attack are made simpler. It functions as a password since some people use common words. Systems that employ multi-word passwords are often resistant to these attack. Furthermore, it is challenging for an attacker to crack passwords made consisting of a mix of random digits, uppercase, and lowercase characters [10].

• Stolen verifier attack: In this harmful act, an attacker first attempts to steal some devices (such as smart IoT devices), after which they launch a power analysis assault on their memory units to steal sensitive data (such as secret passwords and keys) from them. In order to launch further potential network attacks, such as unauthorized session key computation, password guessing, MiTM, and impersonation attack, the attacker eavesdrops on part of the exchanged messages.

• Attack on the computation of the session key without authorization: In this malicious behavior, a hacker attempts to compute the session key that is established between the network's legitimate organizations. The attacker employs a variety of techniques to complete this task, including physical device theft, insider access, and stolen verifier attacks. It is generally advised to compute the session keys using both short term secrets (i.e., random secret nonce values) and long term secrets (i.e., pseudo identities, secret keys). This technique provides unique keys to various entities during several sessions. Unfortunately, if one session key is disclosed to the attacker, the other session keys will still be protected, and the remaining portion of the connection will still be secure.

• Attacks on machine learning (ML) models can be roughly categorized into four categories: (a) dataset poisoning, (b) model poisoning, (c) privacy breach attack, and (d) runtime disruption attack [11].

• Dataset poising attack: In this assault, A uses a variety of techniques to access the training and test data in order to interfere with the ML task's regular operation. A can assault the data server from which raw data must be extracted by using adversarial examples. The compromise of the data sources enables the insertion of false data, which may change how the ML model operates. This further modifies the ML-based system's output [12].

• Model poisoning attack: In a model poisoning attack, a parameter change is made by A, who then interferes with the classifier to produce faulty output. The classifier modifies the parameters used to create the ML model. A can alter the sensitivity limits, rate of accession, and lead to under- or over-fitting, which further impacts how normally ML tasks are carried out [13].

• Privacy violation: Using a variety of techniques, the internal workings of the model and the user's sensitive data may be compromised. The ML task's training and deployment phases can result in data leakage because of unprotected files and a lack of encryption mechanisms. Additionally, it makes it possible for the user to alter the model. As the confidentiality of the sensitive data may be compromised, it raises the privacy risks connected with the data [14]. The many privacy-preserving strategies to safeguard the model's privacy were described by paper not et al. [15,16]. Additionally, they talked about how to "randomize the behavior of the model" [17] in order to use noise creation to give differential privacy

Runtime interference attack: A foe This task is used by A to put off or complete the ongoing ML task. A typically attacks the server during the deployment process. The continuing ML process is then attempted to be remotely disrupted by A. As a result, there is a disruption in the ML task's regular operation, which wastes time and resources. Through various attack including phishing, denial-of-service (DoS) attack, and SQL injection attack, A penetrates the run time server by identifying the weak points (vulnerabilities). The decentralization of the ML work space will be able to reduce the impact of this attack. In order to further divide and implement "distributed machine learning," which helps to safeguard the accuracy and privacy of user data and the related information, blockchain-based mechanisms can be used.

V. ISSUES AND CHALLENGES OF UNITING OF CYBER SECURITY AND MACHINE LEARNING

Although there are several benefits to combining machine learning and cyber security. It also has various problems and difficulties that must be treated with extreme caution. The list below discusses a few of them.

• Problems with compatibility: The combination of machine learning and cyber security involves a variety of machine learning and security approaches, including convolutional neural networks (CNNs), clustering, classification, and signature generation and verification algorithms. Additionally, the data that serves as the primary input for the analytic process

originates from a variety of sources, including IoT devices. Different communication methods are used to operate these Internet of Things devices. Issues with compatibility may arise during the merging of these various algorithms. Therefore, we must be extremely picky about which algorithm and scheme complement each other. Therefore, compatibility-related concerns need to be handled with extreme caution [18].

• Overloading: As was said before, we use a variety of methods that combine machine learning with cyber security. We need additional resources in order to run these algorithms. The system will not operate properly in any other case. As a result, combining and using different algorithms may overwhelm the system, which could therefore impair how well the system actually functions. For instance, we are unable to use the entire system's resources for security-related operations. For the completion of ML-related tasks, we additionally require some resources. As a result, we should choose the algorithms carefully and in accordance with the communication environment's resources. For instance, we would choose to use the Advanced Encryption Standard (AES) algorithm—a symmetric-key-based encryption—instead of any public key cryptographic algorithm for the secure communication of IoT because it costs less to compute, communicate, and store data than public key cryptographic algorithms. In that case, we can also assign system resources for the accomplishment of crucial tasks.

• Accuracy: When combining machine learning and cyber security, we employ a variety of ML processes, or models, to make predictions about certain physical events (i.e., chances of roadside accident in the intelligent transportation system). The ML models rely on certain datasets to function, therefore errors in either the dataset or the ML model's settings might cause serious problems. For instance, the accuracy attained is not entirely accurate [19].

• Security system flaws: When combining ML and cyber security, we may employ a variety of cyber security systems. If these mechanisms have any weaknesses, the system's security may suffer as a result. The majority of the time, hackers look for zero-day vulnerabilities to later attack. The system's sensitive data may be exposed, altered, or rendered unavailable in such circumstances. As a result, security protocol designers should use extreme caution while creating new security protocols.

Through specific procedures, such as the Automated Validation of Internet Security Protocols and Applications (AVISPA) [20], the security of the newly constructed protocol can be verified formally. This tests the security of the protocol against replay and man-in-the-middle attacks. In addition, the Burrows-Abadi-Needham (BAN) logic test [21] can be used to determine whether there is a chance for "safe mutual authentication among the communicating organizations." In addition to these, we can analyze the formal security of a security protocol using the Real-or-Random (ROR) model implementation [22], which highlights the potential for an attack on the planned authentication, access control, or key management protocol that involves unauthorized session key computation. In this approach, the security of the designed protocol may be assessed and examined.

VI. FUTURE RESEARCH

Through specific procedures, such as the Automated Validation of Internet Security Protocols and Applications (AVISPA) [20], the security of the newly constructed protocol can be verified formally. This tests the security of the protocol against replay and man-in-the-middle attacks. In addition, the Burrows-Abadi-Needham (BAN) logic test [21] can be used to determine whether there is a chance for "safe mutual authentication among the communicating organizations." In addition to these, we can analyze the formal security of a security protocol using the Real-or-Random (ROR) model implementation [22], which highlights the potential for an attack on the

planned authentication, access control, or key management protocol that involves unauthorized session key computation. In this approach, the security of the designed protocol may be assessed and examined [23]. Therefore, new security protocols are needed that have more security and functionality characteristics and can withstand zero day vulnerabilities as well [24].

• The compatibility of various tools and mechanisms The "uniting of cyber security and ML" makes use of a variety of methods and tools, or numerous security techniques. as hashing techniques, machine learning algorithms like CNNs and clustering, signature creation and verification algorithms, and encryption algorithms). They also call for various hardware and configurations. These mechanisms and tools' compatibility under such conditions can give rise to some problems [25].

• Performance and overloading: We use a number of the previously mentioned techniques to combine machine learning and cyber security. To run these various algorithms, we need some extra resources. Otherwise, the tasks won't be completed properly. As a result, the system may become overloaded by the combination and use of numerous algorithms, which could hinder the system's actual operation. Therefore, whether in ML or security, we should carefully select the algorithms and work to develop new, lightweight algorithms that consume fewer system resources [26].

• Increased system accuracy: Since ML models rely on certain datasets to function, errors in either the dataset or the ML model's configuration might lead to issues. For instance, the obtained accuracy may not be entirely accurate or the algorithm may produce incorrect predictions. Therefore, it is important for researchers to try to find solutions to these problems. By developing new techniques, mistakes in datasets can be found and the systems' accuracy can be increased [27].

VII. CONCLUSION

Machine learning is well known and used extensively in a variety of fields. Natural language processing (NLP), which can be used to understand what a person or a piece of text is saying, and image processing for recognition are two of the most well-liked applications. In several ways, cybersecurity is unique from other machine learning use cases. Utilizing machine learning for cybersecurity has its own requirements and obstacles. We will go through three particular difficulties with using ML in cybersecurity as well as three typical but more difficult difficulties. Only machine learning can classify complex events and scenarios at scale to enable enterprises to address the challenge of cybersecurity now and in the years to come. This is because more devices and dangers are coming online every day, while human security resources are in short supply. By combining cyber security and machine learning, we revealed the details of two distinct concepts: "cyber security in machine learning" and "machine learning in cyber security." The benefits, problems, and challenges of combining ML with cyber security were then reviewed. In addition, we described several attacks and offered a comparison of various strategies in two separate categories. Finally, some recommendations for future research are given.

VIII. REFERENCES

[1]. Chauhan, D., and J. K. Jain. "A Journey from IoT to IoEA Journey from IoT to IoE." International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075.

[2]. Z. Lv, L. Qiao, J. Li, H. Song, Deep-learning-enabled security issues in the internet of things, IEEE Internet Things J. 8 (12) (2021) 9531–9538.

[3]. Y. Wang, J. Yu, B. Yan, G. Wang, Z. Shan, BSV-PAGS: Blockchain based special vehicles priority access guarantee scheme, Comput. Commun. 161 (2020) 28–40.

[4]. N. Magaia, R. Fonseca, K. Muhammad, A.H.F.N. Segundo, A.V. Lira Neto, V.H.C. de Albuquerque, Industrial internet-of-things security enhanced with deep learning approaches for smart cities, IEEE Internet Things J. 8 (8) (2021) 6393–6405.

[5]. S.A. Parah, J.A. Kaw, P. Bellavista, N.A. Loan, G.M. Bhat, K. Muhammad, V.H.C. de Albuquerque, Efficient security and authentication for edge-based internet of medical things, IEEE Internet Things J. 8 (21) (2021) 15652–15662.

[6]. Ho, Samson, et al. "A novel intrusion detection model for detecting known and innovative cyberattacks using convolutional neural network." IEEE Open Journal of the Computer Society 2 (2021): 14-25.

[7]. Praneeth Narisetty, Pavan Narra "A MACHINE LEARNING APPROACH FOR DETECTING CYBERATTACKS IN NETWORKS", International Journal of Emerging Technologies and Innovative Research (www.jetir.org | UGC and issn Approved), ISSN:2349-5162, Vol.9, Issue 6, page no. ppg26-g31, June-2022, Available at : http://www.jetir.org/papers/JETIR2206605.pdf

[8]. Tufan, Emrah, Cihangir Tezcan, and Cengiz Acartürk. "Anomaly-based intrusion detection by machine learning: A case study on probing attacks to an institutional network." IEEE Access 9 (2021): 50078-50092.

[9]. Sarker, Iqbal H., et al. "Intrudtree: a machine learning based cyber security intrusion detection model." Symmetry 12.5 (2020): 754.

[10]. Abdeldayem, Mohamed M. "Intrusion Detection System Based on Pattern Recognition." Arabian Journal for Science and Engineering (2022): 1-9.

[11]. Y. Sun, A.K. Bashir, U. Tariq, F. Xiao, Effective malware detection scheme based on classified behavior graph in IIoT, Ad Hoc Netw. 120 (2021) 102558.

[12]. J. Yang, Z. Bian, J. Liu, B. Jiang, W. Lu, X. Gao, H. Song, Noreference quality assessment for screen content images using visual edge model and AdaBoosting neural network, IEEE Trans. Image Process. 30 (2021) 6801–6814.

[13]. Y. Zhao, J. Yang, Y. Bao, H. Song, Trustworthy authorization method for security in industrial internet of things, Ad Hoc Netw. 121 (C) (2021).

[14]. T.S. Messerges, E.A. Dabbish, R.H. Sloan, Examining smart-card security under the threat of power analysis attacks, IEEE Trans. Comput. 51 (5) (2002) 541–552.

[15]. M.R.K. Soltanian, I.S. Amiri, Chapter 3 - problem solving, investigating ideas, and solutions, in: M.R.K. Soltanian, I.S. Amiri (Eds.), Theoretical and Experimental Methods for Defending Against DDOS Attacks, Syngress, 2016, pp. 33–45.

[16]. T. Lei, Z. Qin, Z. Wang, Q. Li, D. Ye, EveDroid: Event-aware android malware detection against model degrading for IoT devices, IEEE Internet Things J. 6 (4) (2019) 6668–6680.

[17]. J. Steinhardt, P.W. Koh, P. Liang, Certified defenses for data poisoning attacks, in: 31st International Conference on Neural Information Processing Systems, in: NIPS'17, Curran Associates Inc. Long Beach, California, USA, 2017, pp. 3520–3532.

[18]. M. Aladag, F.O. Catak, E. Gul, preventing data poisoning attacks by using generative models, in: 1st International Informatics and Software Engineering Conference, UBMYK, Ankara, Turkey, 2019, pp. 1–5, http://dx.doi.org/10.1109/UBMYK48245.2019.8965459.

[19]. C. Huang, S. Chen, Y. Zhang, W. Zhou, J.J.P.C. Rodrigues, V.H.C. de Albuquerque, a robust approach for privacy data protection: IoT security assurance using generative adversarial imitation learning, IEEE Internet Things J.

(2021) 1, http://dx.doi.org/10.1109/JIOT.2021. 3128531.

[20]. N. Papernot, P. McDaniel, X. Wu, S. Jha, A. Swami, Distillation as a defense to adversarial perturbations against deep neural networks, in: 2016 IEEE Symposium on Security and Privacy, 2016, pp. 582–597, http://dx.doi.org/10.1109/SP.2016.41.

[21]. N. Papernot, A marauder's map of security and privacy in machine learning, in: 11th ACM Workshop on Artificial Intelligence and Security, Toronto, Canada, 2018.

[22]. S. Pirbhulal, W. Wu, K. Muhammad, I. Mehmood, G. Li, V.H.C. de Albuquerque, Mobility enabled security for optimizing IoT based intelligent applications, IEEE Netw. 34 (2) (2020) 72–77.

[23]. Chauhan, Dipti, Jay Kumar Jain, and Sanjay Sharma. "An end-to-end header compression for multihop IPv6 tunnels with varying bandwidth." 2016 Fifth international conference on eco-friendly computing and communication systems (ICECCS). IEEE, 2016.

[24]. Jain, Jay Kumar, Devendra Kumar Jain, and Anuradha Gupta. "Performance analysis of node-disjoint multipath routing for mobile ad-hoc networks based on QOS." International Journal of Computer Science and Information Technologies 3.5 (2012): 5000-5004.

[25]. Waoo, A., and Sanjay Sharma. "Threshold Sensitive Stable Election Multi-path Energy Aware Hierarchical Protocol for Clustered Heterogeneous Wireless Sensor Networks." International Journal of Recent Trends in Engineering & Research 3.09 (2017): 158-16.

[26]. Jain, Jay Kumar, and Sanjay Sharma. "Performance Evaluation of Hybrid Multipath Progressive Routing Protocol for MANETs." International Journal of Computer Applications 71.18 (2013).

[27]. Jain, Jay Kumar, and Akhilesh A. Waoo. "An Analytical Study of Energy Efficient Routing Approaches in Wireless Sensor Network." THEETAS 2022: Proceedings of The International Conference on Emerging Trends in Artificial Intelligence and Smart Systems, THEETAS 2022, 16-17 April 2022, Jabalpur, India. European Alliance for Innovation, 2022.

[28]. J. K. Jain, C. S. Dangi and D. Chauhan, "An Efficient Multipath Productive Routing Protocol for Mobile Ad-hoc Networks," 2020 IEEE International Conference for Innovation in Technology (INOCON), 2020, pp. 1-5, doi: 10.1109/INOCON50539.2020.9298291.

Cite this article as :

Jay Kumar Jain, Akhilesh A. Waoo, Dipti Chauhan, "A Literature Review on Machine Learning for Cyber Security Issues", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8, Issue 6, pp.374-385, November-December-2022. Available at doi : https://doi.org/10.32628/CSEIT228654

Journal URL : https://ijsrcseit.com/CSEIT228654