For cited: *Kadhim, A. A. (2022). Uniting Cyber Security and Machine Learning; Advantages, Challenges and Future Research. Journal of Current Research on Engineering, Science and Technology, 8 (2), 129-140.*

Journal of Current Research on Engineering, Science and Technology (JoCREST)
ISSN: 2651-2521

Year: 2022
Volume: 8
Issue: 2

Crossref doi: 10.26579/jocrest.8.2.10

Review

Uniting Cyber Security and Machine Learning; Advantages, Challenges and Future Research

Ali Awad KADHİM[1]

Keywords
Cyber security, machine learning, advantages, challenges, future learning.

Article History
Received
20 Sep, 2022
Accepted
31 Dec, 2022

Abstract
Accounting for cyber security, where machine learning is applied, and using ML to facilitate cyber security are two essential parts of the combination of cyber security with ML. Cybersecurity systems may identify trends using machine learning, react to adapt, prevent similar attacks, and respond to changing activity. It can help cybersecurity businesses become more competitive in preventing threats and providing the necessary defences against continuing attacks. It can reduce wasted time on repetitive tasks and help businesses use their resources more effectively. Cybersecurity is improved by machine learning because it becomes more proactive, less expensive, and less sophisticated. Only when machine learning is accompanied by data that completely captures the environment will it be able to perform such jobs. As they say, trash in, waste out.

1. Introduction

In the modern era of computers, most gadgets are interconnected via the Internet of Things (IoT). Through the Internet, direct communication that is open and insecure, these gadgets exchange and transmits data. This kind of confidential material is typical (i.e., healthcare data, also insurance data, and social security numbers). The negative actors, such as the online criminals (hackers), are continuously searching for ways to manipulate the objects (for instance, they can carry out assaults like replaying, man-in-the-middle, impression, code guess, key exchange calculation, software insertion, and data manipulation)(Butun et al., 2019). As a result, many researchers occasionally suggest various security methods to reduce these threats. Rules must be followed; among the groups within which cyber security protocols can be alienated are user access procedures, intrusion protection protocols, critical management protocols, and namecoin security procedures. The following is a summary of these procedures.

Checking somebody's or a device's validity is the procedure of authentication. It is possible to execute it using passwords or factors directly related to the users or device (for example, username, password, smartcard, and fingerprints). Users-to-user, consumer-to-device, or device-to-device identification are all options.

[1] Corresponding Author. ORCID: 0000-0001-9000-0597. Asst. Lectuer, Middle Technical University, Institute of Medical Technology Al-Mansour

According to the scoring systems, strong authentication methods can also be divided into three groups: one-factor user authentication protocols, multifactor authentication methods, and multifactor authentication procedures. Limiting illegal access to someone or something using access control is a method (s). After completing all consumer user access protocol steps, users or devices can securely access other customers or devices (Lv et al., 2020).

In order to handle keys securely among the many entities, including some devices (such as Internet of Things smart plans and smart cars) and certain users (smart home users, doctors, traffic inspectors). The registration of each entity in the communication network is often handled by a reliable registration authority, which then memorizes the secret credentials (i.e., secret keys). We require a key management method for creating new keys, storing them in devices, establishing keys, and revoking them. After establishing a shared private key (also known as a key pair), which may be accomplished through the crucial phases of an authenticated agreement of keys protocol, the smart objects can safely transmit their data.

One of the modern era's growing technologies is blockchain. Data is kept on a blockchain in the form of particular blocks connected to use hash values. Data is preserved in the blockchain in the format of a public ledger, also known as a distributed ledger (DLT). The DLT is accessible to all legitimate network participants (and occasionally miners). The data we store on the blockchain is protected from various potential cyberattacks. Consequently, blockchain-enabled security protocols can prevent numerous cyberattacks(Wang et al., 2020).

Computing systems study through data and utilize algorithms to carry out tasks without even being pattern recognition. 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 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. Nevertheless, DL often requires less constant human assistance. This allows it to evaluate unstructured data, such as photographs, videos, and films, more effectively than conventional ML algorithms (Magaia et al. (2020).

We can benefit from machine learning and cyber security in several ways. For instance, improved cyber security techniques increased machine learning model security and more effective zero-day attack detection with less human involvement. Nevertheless, this could experience several problems and security challenges that must be managed carefully. As a result, in this specific field, we require a review study that addresses issues and challenges, obstacles in various safety systems with a comparison study, as well as some future areas for study with the other researchers must include the "uniting of cyber security and machine learning," i.e., problems and difficulties. Therefore, we attempted to conduct such research in the presented design (Parah et al., 2020).

2. Literature Review

The development of machines2 that autonomously learn to make judgments is the aim of machine learning (ML). Training is the first step in the learning process. A machine learning model is created by instructing a computer to use a certain ML algorithm to examine some "existing" (training) data. One such model includes a function that makes judgments based on "future" data and includes all the knowledge gained during the training stage. An ML model's effectiveness must be evaluated before deployment in a real-world setting. To achieve this, the ML model processes some "validation" data, and the forecasts are either examined by people or contrasted with an established truth (Apruzzese et al., 2022). Therefore, we may define a machine learning technique as "the process of constructing a machine learning model by employing ML algorithms on some training examples." Fig. 1 roughly shows an example of how the training and testing phases would go.

Figure 1

Fig. 1. advancing machine learning A ML model is created after gathering data for training and using an ML algorithm to analyze it. Such an ML model needs to be validated to use some data. The ML model can be implemented in operation if the result of such an evaluation is notable.

Kumar et al. (2019) proposed a system that improved penetration testing for IoT. by fusing the benefits of blockchain with ML models. They put it into practice step-by-step using "clustering, categorization, and blockchain." Utilizing classification and clustering techniques, ML was developed to extract the infection data, which was then saved over the public blockchain.

Lei et al. (2019).). Proposed "EveDroid, a modular and occurrence malware detection software", is the name of the security system. Their approach, which differs from existing ones, employs event groups directly to characterize the actions of apps, which could capture a greater semantic level for detection methods.

Nguyen et al. (2018) investigated a method for detecting Linux IoT botnets. The "PSI graph and CNN classification" pairing provided the basis for the identification.

Dinakarrao et al. (2019) argued that They used multiple approaches to find the incursions, using a runtime infection detection (HaRM) that used "Hardware Productivity Counter (HPC)" data to find malicious and good apps. Su et al. (2018) investigated a simple method for identifying DDoS malware in IoT systems. They used a lightweight convolution neural network to classify the malware images they had collected. Only recent name names have been considered for the comparative study portion. Several performance metrics are employed, including clarity, accuracy, correctness, and F1-score. Kumar et al. (2019) concluded that we could

calculate these variables using parameters like true positive, false acceptance, true negative, and false negative. A "normal programme" identified by intrusion detection as normal is known as a "true negative." In contrast, a "normal programme" identified by an intrusion detection system as malicious is known as a "false positive." Similar to this, if an intrusion detection system identifies a "malicious programme" as such, it is referred to as a "true positive (TP)"; however, if an intrusion detection system identifies a "malicious programme" as such, it is referred to as a "negative result (Lv et al., 2020).

Machine learning and cyber security are both crucial to one another and can enhance efficiency. ML models are susceptible to a variety of attacks. These assaults may impact the functioning, effectiveness, and forecasts of the ML models. Nevertheless, implementing specific cyber security measures can prevent these undesirable incidents. The functioning, performance, and input data of the ML models are secure underneath the application of cyber security procedures, and they obtain correct estimates and outcomes (Yang et al., 2021). Moreover, ML techniques, such as supervised methods, unsupervised learning, strengthening, and deep algorithms, can be employed depending on the communications and the interconnected systems when we employ ML algorithms in cyber security strategies (i.e., intrusion prevention systems) that enhance their performance results (i.e., started to improve accuracy and positive predictive value with less false negative rate). The cyber security techniques that identify intrusions using ML models appear to be particularly good at identifying zero-day assaults (i.e., unknown malware attacks). We use certain deployed ML models to carry out the detection, which is why it occurs. ML models gather and compare specific features; if a project's features match another program's characteristics, that programme may be judged malevolent. The ML models are capable of carrying out this detecting operation autonomously. Therefore, combining cyber security and machine learning makes it possible to detect zero-day threats efficiently (Magaia et al. (2020).

When we combine ML and cyber security, almost all of the jobs used for these systems are completed without any need for human interference or minimal human engagement. As they employ specific ML algorithms, ML-based intrusion detection systems are particularly effective at detecting the existence of attacks. As a result, combining deep learning with cyber security systems speeds up the scanning of breaches and offers a quick response in any indication of an attack. The selection of an appropriate ML algorithm is the sole factor we must consider. (Apruzzese et al., 2022)

Although there are several benefits to combining machine learning and cyber security, it also has some problems and difficulties that must be managed cautiously. Combining machine learning with cyber security involves using a variety of machine learning and security approaches, such as convolutional neural networks (CNNs), clustering, categorization, and signatures creation and verification algorithms (Butun, et al 2019) Additionally, data—the primary input for the analysis process—comes from various sources, including IoT devices. Various communication methods are used to operate these Internet of Things devices. There might be compatibility problems when combining these various

algorithms. As a result, we must be extremely picky about which algorithm and framework work well together. As a result, compatibility-related issues need to be treated with extreme care (Yang et al., 2020). We employ several methods, as was already said, to combine machine learning with cyber security. We require additional resources to execute such algorithms. Then, the system will need to be fixed. As a result, combining and using different algorithms may overwhelm the system, impairing how well the system functions. For instance, we can only dedicate some of the system's resources to security-related operations. For the completion of ML-related activities, we additionally require some resources. So, we should pick the algorithms carefully and follow the capabilities of the communications (Lv et al., 2020).

For secure IoT communications, Instead of any general populace encryption method, we would prefer to use based on the cryptography, like the Advanced Encryption Standards (AES) algorithm. because AES has lower computational, communication, and storage space. In that case, we can assign system resources to critical activities (Guimaraes et al., 2018).

We combine machine learning and cyber security by using various ML techniques, or machine learning (ML) models, to forecast some physical events (i.e., chances of a roadside accident in the intelligent transportation system). The ML models depend on certain datasets for operation, and if either the data or the ML model's settings contain an error, serious problems may result. For instance, the accuracy attained is not entirely accurate (Guimaraes et al., 2018)

Moreover Variety of cyber security approaches to integrate ML with cyber security. These mechanisms could compromise the system's security if they exhibit some faults. The majority of the time, hackers look for zero-day flaws to attack. Sensitive data stored in the system may be made public, altered, or inaccessible in such circumstances. As a result, security protocol developers should exercise extreme caution while creating new security protocols. Using specific procedures, such as the Automatic Verification of Internet Security Protocols and Applications, the security of the newly designed protocol can be evaluated (Armando et al., 2005)., It uses formal security requirements to assess the protocol's security against replay and guy assaults. Additionally, It detects the potential for "safe mutual authentication among the communicative organizations." In addition to all these, the Real-or-Random model can be used to examine the formal security of a security protocol (Abdalla et al., 2005). an implementation that alerts the intended authentication, security systems, or key management protocol to the danger of an unauthorized session key calculation attack. In this approach, the security of the proposed protocol may be assessed and examined (Magaia et al. (2020).

3. Methodology

Cyberattacks are developing swiftly due to the expansion of the Internet, and cyber security might be higher. This survey report includes a brief instructional overview of each ML/DL technique and a key review of the literature on ML and DL methods for network theory and vulnerability scanning. The publications that represent each approach were indexed, examined, and summarised based on the temporal or climatic relationships between them. We discuss some of the communication

networks widely utilised in machine learning and deep learning, discuss challenges in applying ML/DL to cybersecurity, and suggest future study areas. Network intrusion detection data are essential for system development and testing (Yang et al., 2020).

With the need for a data set, the ML and DL algorithms can be used, and obtaining just one dataset is easier and more important. There are many problems with the present public dataset, such as inconsistent data, outdated information, etc. These problems have severely restricted the development of research in this area. Rapid network information changes make it difficult to train and use DL and ML models. And others, 2019). Consequently, training needs to be given to models quickly and completely. As a result, lifetime learning and learning algorithms will be the subject of future study.

4. Result and Discussion

Replay, impersonating, credential leakage, public key leakage, unauthorized data update, flood, service disapproval, and scattered services refusal are only a few of the threats that are susceptible to cyber world devices. Therefore, we require security measures to detect and thwart these threats. The supplied, which was prior to the dataset, allows the machine learning algorithms (ML algorithms) to understand about various cyberattacks in the online and offline forms. In real-time or fully online, the ML algorithms identify any indication of infiltration (such as a cyberattack) (Magaia et al. (2020).

Figure 2 shows the scenario of "machine learning in cyber security." We have Web devices (such as laptops, desktop computers, smartphones, and IoT devices) that may be used for various online functions, including financial transactions, access to healthcare information, social security numbers, and more. Hackers are always looking for weaknesses in these systems, and once they find one, they begin their attacks. Various machine learning (ML) approaches can be applied to identify and counteract cyberattacks, including supervised, uncontrolled, reinforcement, and deep learning. The learning method that works best for a system—supervised learning, unsupervised, supervised learning, or deep learning—depends on the communications, the resources that are accessible, and the system itself. The internet servers' strong storage and processing capabilities allow for the understanding (training) and forecasting (testing) of cyber threats.

Figure 2. "machine learning in cyber security".

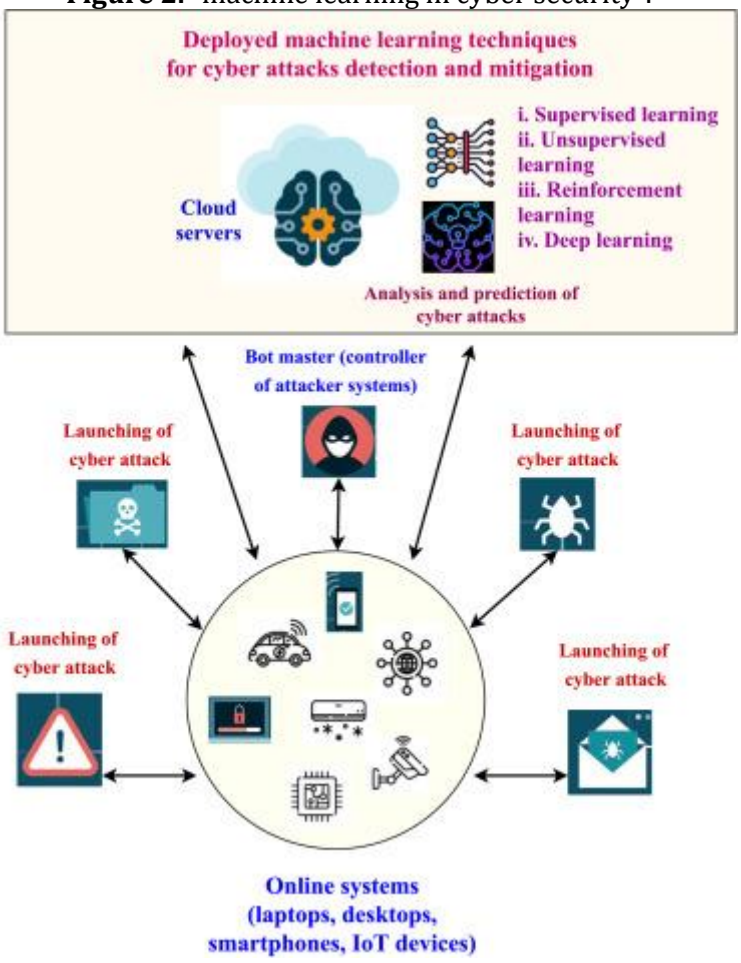

Figure 3: "Scenario of cyber security in machine learning"

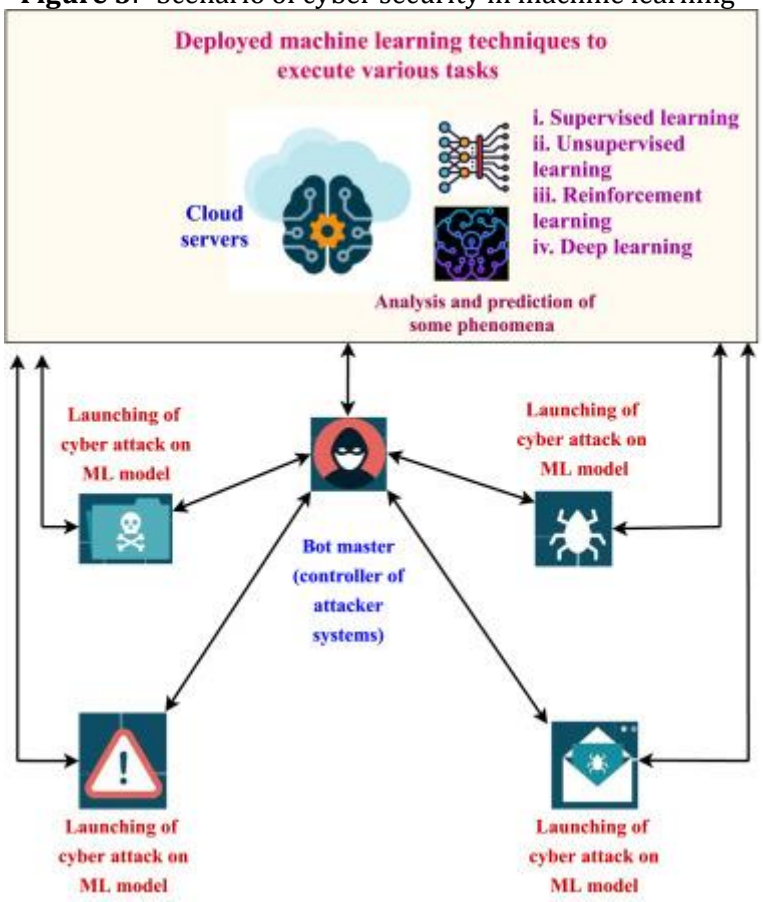

Figure 3 presents a scenario for "cyber security in machine learning," also known as machine learning (ML) security. For the analysis and forecasting of numerous events, ML models are employed. However, the launch of some assaults, such as the dataset poison attack, model poisoned attack, privacy breach attack, member inference strike, real-time interruption attack, etc., might impact the accuracy of ML models. (Sun et al. 2021). These attacks may incorrectly cause machine learning (ML) models to forecast related events. The "dataset poison attack" involves the introduction of adversarial attacks (updated values) into the dataset by an attacker, which leads the ML model to make incorrect predictions. The "model poisoning assault" also aims to degrade the models by tampering with their internal operations and changing their settings. In a "privacy breach assault," the attacker seeks to retrieve the model's valuable information while also working to expose sensitive data. An example of a privacy violation is a member inference attack.

Additionally, in a "runtime disturbance attack," the adversary subverts the ML workflow to influence the accuracy of the prediction by assaulting the model's execution procedure. Therefore, to defend against such assaults, it needs some cyber security mechanisms (such as encryption methods, signature creation and validation methods, and hash processes). The ML models and the related datasets are secured using these cybersecurity procedures, and the predicted results are accurate (Magaia et al. 2020).

Below are some potential future study areas related to the "unification of machine learning and cyber security."

The privacy of the transferred and stored data is very important. To keep the data private, various security mechanisms have been suggested. These procedures, however, fall short when there is a design fault or when a zero-day assault occurs. Thus, there is room for improvement as online attackers (hackers) develop and deploy cutting-edge techniques to circumvent the system's security. Therefore, There is a need for new security procedures that have improved security and usefulness and can resist zero-day vulnerabilities (Wang et al., 2020).

The "uniting of cyber security and ML" employs a wide range of methodologies and instruments (i.e., numerous security approaches). such as hashing techniques, machine learning algorithms like grouping, categorization, and CNNs, signature creation and validation algorithms, and data encryption). Additionally, they need various hardware and setups. In these conditions, certain problems might be connected to the interoperability of various mechanisms and instruments (Guimaraes et al., 2018).

We combine ML and cyber security using a variety of techniques. We require additional resources in order to execute these numerous algorithms. Otherwise, the tasks must be appropriately carried out. As a result, combining and using different algorithms may overwhelm the computer and further impair its functionality. As a result, we should choose our methods carefully and work to develop new, resource-efficient light ML or security algorithms. The ML models rely on specific datasets to function, and difficulties may result if there is an issue with both the data and the ML model's settings. (Wang et al., 2020). For instance, the accuracy must be 100% accurate, or the algorithm may anticipate anything incorrectly. Researchers, therefore, should try to resolve these issues; new techniques can be developed to find faults in the datasets or raise the system's accuracy.

5. Conclusion

Machine learning can be used by cybersecurity system to analyse trends and learn more for them in order to stop numerous attacks and respond to changing activity. It can help cybersecurity organizations be so much more proactive in preventing assaults and creating the necessary circumstances for continuing assaults. It can reduce the amount of time spent on repeated tasks and let businesses use their resources more effectively. Cybersecurity could be considerably enhanced by computer vision by making it more efficient, quick, and affordable. It can only complete these tasks, though, if the machine learning is supplemented with data that accurately reflects the surroundings. One rapidly growing area of computer science with many applications is machine learning (ML). ML algorithms can be divided into supervised, unsupervised, and supervised learning categories. The pre-classified are well-processes utilized in machine learning algorithms. Different categories of supervised learning algorithms include extrapolation and categorization. Many domains already use machine learning (ML), a leading technique for existing and future data systems.

Nevertheless, the use of ML in cybersecurity is still in its infancy, demonstrating a significant gap between theory and application. Such a mismatch stems from the current state of the art, which makes it impossible to pinpoint the function of ML in cybersecurity. The true capacity of ML will only be realized if its advantages and disadvantages are widely acknowledged. This paper represents the first attempt to give any interest in this area with just an interest in the matter a comprehensive knowledge of the function of ML in the entire cybersecurity sector. It highlights the benefits of ML over human-driven detection techniques and the testing of the system duties that ML can handle.

Additionally, we clarify several inherent issues that affect actual ML implementations in cybersecurity. Finally, we demonstrate how multiple stakeholders can participate in ML in cybersecurity research, which is vital for the field's further advancement. Two actual case studies demonstrating industry implementations of ML as a defence against cyber threats are added as a complement to our efforts.

References

Abdalla, M., Fouque, P. A., & Pointcheval, D. (2005, January). Password-based authenticated key exchange in the three-party setting. In *International workshop on public key cryptography* (pp. 65-84). Springer, Berlin, Heidelberg.

Abdalla, M., Fouque, P. A., & Pointcheval, D. (2005, January). Password-based authenticated key exchange in the three-party setting. In *International workshop on public key cryptography* (pp. 65-84). Springer, Berlin, Heidelberg.

Apruzzese, G., Laskov, P., de Oca, E. M., Mallouli, W., Rapa, L. B., Grammatopoulos, A. V., & Franco, F. D. (2022). The Role of Machine Learning in Cybersecurity. *Digital Threats: Research and Practice*.

Armando, A., Basin, D., Boichut, Y., Chevalier, Y., Compagna, L., Cuéllar, J., ... & Vigneron, L. (2005, July). The AVISPA tool for the automated validation of internet security protocols and applications. In *International conference on computer aided verification* (pp. 281-285). Springer, Berlin, Heidelberg.

Butun, I., Österberg, P., & Song, H. (2019). Security of the Internet of Things: Vulnerabilities, attacks, and countermeasures. *IEEE Communications Surveys & Tutorials*, *22*(1), 616-644.

Coulter, R., Han, Q. L., Pan, L., Zhang, J., & Xiang, Y. (2019). Data-driven cyber security in perspective—Intelligent traffic analysis. *IEEE transactions on cybernetics*, *50*(7), 3081-3093.

Dinakarrao, S. M. P., Sayadi, H., Makrani, H. M., Nowzari, C., Rafatirad, S., & Homayoun, H. (2019, March). Lightweight node-level malware detection and network-level malware confinement in iot networks. In *2019 Design, Automation & Test in Europe Conference & Exhibition (DATE)* (pp. 776-781). IEEE.

Guimaraes, R. R., Passos, L. A., Holanda Filho, R., de Albuquerque, V. H. C., Rodrigues, J. J., Komarov, M. M., & Papa, J. P. (2018). Intelligent network security monitoring based on optimum-path forest clustering. *Ieee Network*, *33*(2), 126-131.

Kumar, R., Zhang, X., Wang, W., Khan, R. U., Kumar, J., & Sharif, A. (2019). A multimodal malware detection technique for Android IoT devices using various features. *IEEE access*, *7*, 64411-64430.

Lei, T., Qin, Z., Wang, Z., Li, Q., & Ye, D. (2019). EveDroid: Event-aware Android malware detection against model degrading for IoT devices. *IEEE Internet of Things Journal*, *6*(4), 6668-6680.

Lv, Z., Qiao, L., Li, J., & Song, H. (2020). Deep-learning-enabled security issues in the internet of things. *IEEE Internet of Things Journal*, *8*(12), 9531-9538.

Magaia, N., Fonseca, R., Muhammad, K., Segundo, A. H. F. N., Neto, A. V. L., & de Albuquerque, V. H. C. (2020). Industrial internet-of-things security enhanced with deep learning approaches for smart cities. *IEEE Internet of Things Journal*, *8*(8), 6393-6405.

Messerges, T. S., Dabbish, E. A., & Sloan, R. H. (2002). Examining smart-card security under the threat of power analysis attacks. *IEEE transactions on computers*, *51*(5), 541-552.

Nguyen, H. T., Ngo, Q. D., & Le, V. H. (2018, September). IoT botnet detection approach based on PSI graph and DGCNN classifier. In *2018 IEEE international conference on information communication and signal processing (ICICSP)* (pp. 118-122). IEEE.

Parah, S. A., Kaw, J. A., Bellavista, P., Loan, N. A., Bhat, G. M., Muhammad, K., & de Albuquerque, V. H. C. (2020). Efficient security and authentication for edge-based internet of medical things. *IEEE Internet of Things Journal*, *8*(21), 15652-15662.

Soltanian, M., & Amiri, I. (2016). Problem Solving, Investigating Ideas, and Solutions. *Theoretical and Experimental Methods for Defending Against DDOS Attacks*, 33-45.

Su, J., Vasconcellos, D. V., Prasad, S., Sgandurra, D., Feng, Y., & Sakurai, K. (2018, July). Lightweight classification of IoT malware based on image recognition. In *2018 IEEE 42Nd annual computer software and applications conference (COMPSAC)* (Vol. 2, pp. 664-669). IEEE.

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

Wang, Y., Yu, J., Yan, B., Wang, G., & Shan, Z. (2020). BSV-PAGS: Blockchain-based special vehicles priority access guarantee scheme. *Computer Communications*, *161*, 28-40.

Yang, J., Bian, Z., Liu, J., Jiang, B., Lu, W., Gao, X., & Song, H. (2021). No-reference quality assessment for screen content images using visual edge model and

adaboosting neural network. *IEEE Transactions on Image Processing*, *30*, 6801-6814.

Yang, J., Han, Y., Wang, Y., Jiang, B., Lv, Z., & Song, H. (2020). Optimization of real-time traffic network assignment based on IoT data using DBN and clustering model in smart city. *Future Generation Computer Systems*, *108*, 976-986.