University of Ibadan
Journal of Science and Logics in ICT Research

Machine Learning in Cyber Security Operations

[1]✉ Nureni A. Azeez and [2]Isiekwene C. Chinyere.

University of Lagos, Nigeria
nurayhn1@gmail.com; isiekwenechioma@gmail.com.

Abstract

The defense of computational devices as well as computer networks against information leaks, theft, and damage to their electronic data, software, hardware, or other components, as well as against interruption or misrepresenting the services they offer, is defined as cyber security by securitystudio.com. In recent years, there has been an unparalleled increase in public interest in machine learning (ML) research. People's learning and working styles are changing as the Internet and social life become more intertwined, yet this also exposes them to major security risks. Protecting confidential data, networks, and computer-connected systems against illegal cyberattacks is a difficult challenge. Effective cyber security is crucial for this. To solve this issue, recent technologies like machine learning and deep learning are combined with cyberattacks. The write-up covers machine learning technology in cyber security, explores the benefits and limitations of employing them, and offers recommendations for future research. The world of today is highly network-interconnected due to the prevalence of both small personal devices (like smartphones) and large computing devices or services (like cloud computing or online banking). As a result, millions of data bytes are generated, processed, exchanged, shared, and used every minute to produce results in specific applications. As a result, protecting user privacy, machine (device) security, and data in cyberspace has become a top priority for private citizens, corporate entities, and national governments. Machine learning (ML) has often been used in cybersecurity in recent years, including for biometric-based user authentication and intrusion or virus detection. But ML algorithms are vulnerable to intrusions during both the training and testing phases, which often lead to noticeable performance decreases and security vulnerabilities. Comparatively little studies have been conducted to ascertain the type, extent, and defense mechanisms of ML methods' vulnerabilities against security threats. Systematizing recent cybersecurity-related initiatives leveraging ML is vital to garner the interest of researchers, scientists, and engineers.

Keywords: *Machine Learning, Cyber Security, Application, Cyberattacks, Detection*

1. INTRODUCTION

Cyber security can defend systems connected to the internet, including hardware, software, and data, against cyberattacks. Computers, networks, programs, and data are all secured by a variety of technologies and procedures called "cybersecurity" to prevent assaults and unauthorized users, modification, or destruction [1].

The cybersecurity sector uses cutting-edge technology like machine learning (ML) to exploit security capabilities as attacks become more sophisticated [2]. The goal of Machine learning in Cyber security is to detect the minutest behavior of malware and ransomware attacks before they enter the system.

Faster response in time is enabled by applying automation to threat hunting and there is a better response in form of recommendations [3]. There are three types of security controls (operational security, management security, and physical security) but the focus is on operational security [4]. Its operations include internet of things (IoT) security, network security, application security, and critical infrastructure security [8].

Azeez N. A. and Isiekwene C. C., (2024). Machine Learning in Cyber Security. *University of Ibadan Journal of Science and Logics in ICT Research (UIJSLICTR),* Vol. 11 No. 2, pp. 57 - 70

©U IJSLICTR Vol. 11, No. 2, June 2024

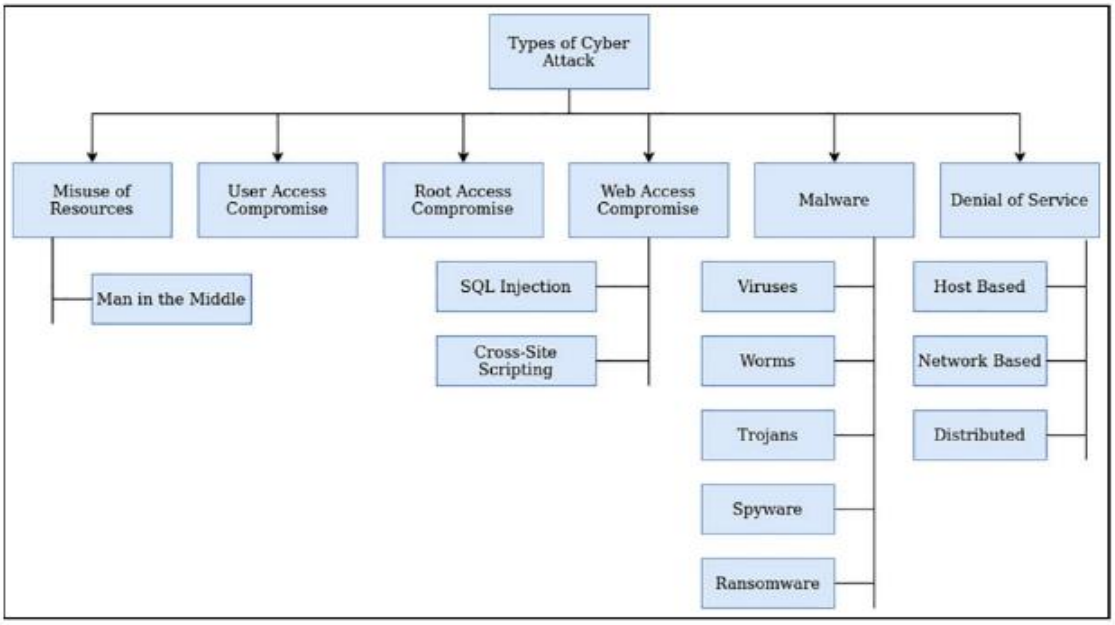


Fig 1: Cyber attach taxonomy [11]

1.1 Cyber Security

Cyber security operations are categorized into five steps, namely; (i) Identification of sensitive or critical information (product research, intellectual property, financial data, customer information, and employee information), (ii) threat and (iii)vulnerability analysis, (iv)risk assessment, and (v)application of effective countermeasures are all steps in the process [27]. The use of Machine learning in cyber security operations has become more popular due to the positive impact experienced over the years but the issue of cybercrime has increased as much as the impact made [5]. This has become a major concern, hence the use of artificial intelligence in cyber security has become crucial as there is a great need for smart automated systems to detect crime or one that looks like it with the use of advanced systems machine-trained software agents capable of making decisions for the safety of the cyberspace [6].

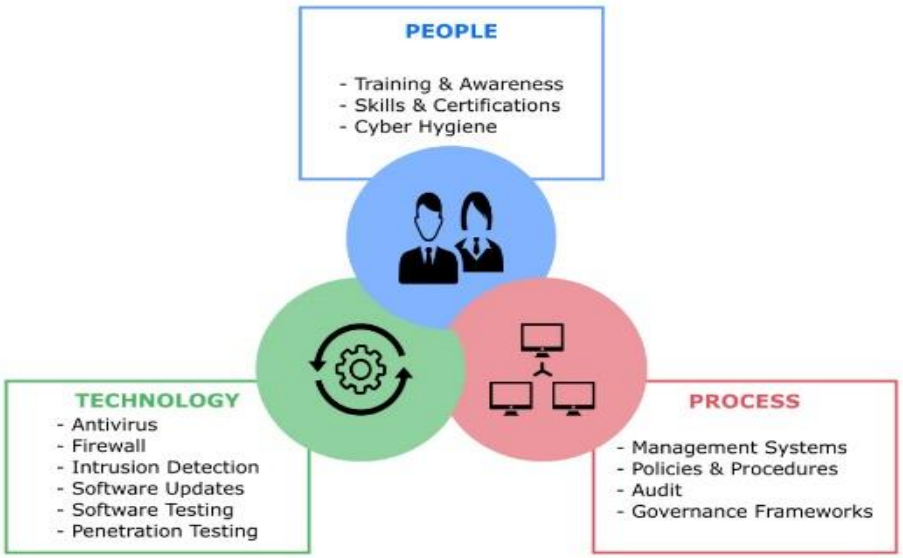


Fig 2: Cyber security framework (Github.com) [12]

1.2 Classification

The capacity of a system to automatically learn from data, enhance performance based on previous experiences and make predictions is known as machine learning [18]. Numerous algorithms used in machine learning process vast amounts of data. These algorithms get training data, which they use to develop the model and complete a certain task [7].

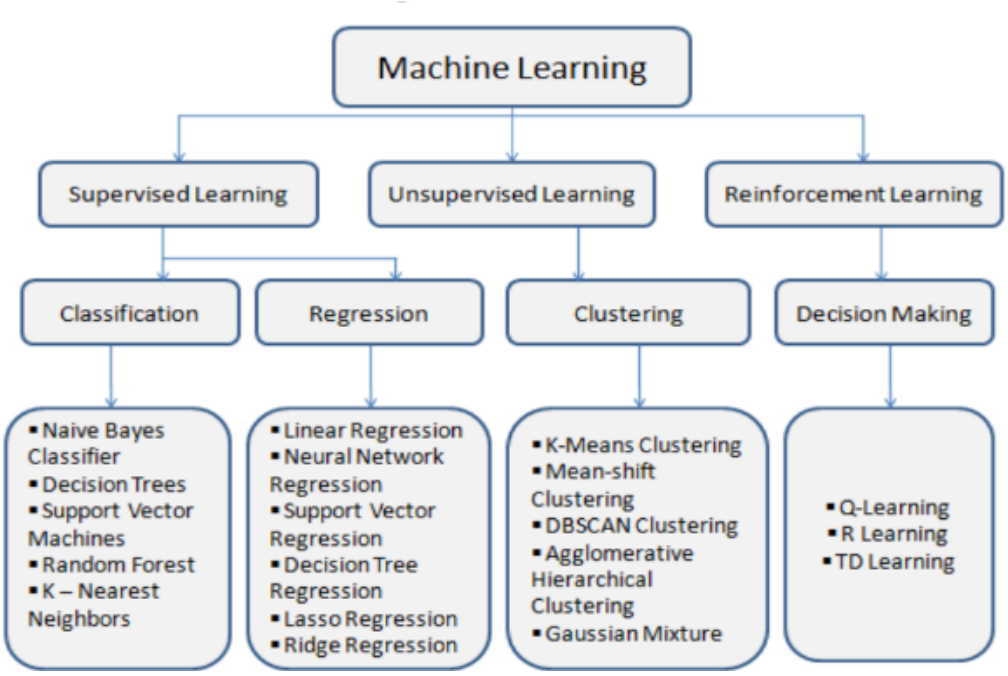


Fig 3: Types of ML with its corresponding algorithm [25]

These machine-learning methods support the resolution of a variety of business issues, including regression, classification, forecasting, clustering, associations, etc. The four basic categories of machine learning are determined by the learning processes and methods, which are:

1. **Supervised Machine Learning**: "Supervised learning, often known as supervised machine learning, is a subset of machine learning and artificial intelligence. It stands out for the way it uses labeled datasets to teach computers to appropriately categorize data or anticipate events. As input data are fed into the model, the weights are adjusted until the model is adequately fitted, which happens as part of the cross-validation process." With the use of supervised learning, organizations may address a variety of scaled real-world concerns, such as sorting spam in a separate category from your email [15]
2. **Unsupervised Machine Learning**: In contrast to supervised learning, unsupervised learning uses unlabeled data. It takes these data and derives patterns to assist with clustering or association problems. This is especially useful when subject matter specialists are not aware of common features in a data set. Techniques for hierarchical clustering, k-means, and Gaussian mixture models are often employed [15].
3. S**emi-Supervised Machine Learning**: Only a portion of the incoming data is labeled in semi-supervised learning. Unsupervised and semi-supervised learning may be more appealing alternatives because using domain expertise to accurately categorize data for supervised learning can be time-consuming and expensive.
4. **Reinforcement Learning**: a subfield of machine learning that examines how intelligent agents should act in a specific environment to maximize the idea of cumulative reward. One of the three primary machine learning paradigms is reinforcement learning, along with supervised learning and unsupervised learning. Contrary to supervised learning, reinforcement learning does not need the

explicit correction of undesirable behavior or the presentation of labeled input/output pairings. Instead, a focus on creating a balance between the use of current knowledge and its exploration (of uncharted territory).

2. Related Works

This section explores the historical adaptation of these three applications to more basic, established ML methods. Machine Learning (ML) research was dominated by a large set of decades-old algorithms before the significant, primarily deep learning-driven advancements of the last 5–10 years. This is referred to as "traditional machine learning" in this context. [22].

Usually, these techniques fall into one of two groups: supervised learning or unsupervised learning. Unsupervised learning examines unlabeled datasets to find underlying patterns in the data as opposed to supervised learning, which uses labeled datasets to train a model to categorize incoming inputs.

Comparison research of deep learning methods for cyber security intrusion detection, and the datasets that were employed. In particular, an overview of deep learning-based intrusion detection systems [24]. The dataset is crucial for intrusion detection, so about 35 well-known cyber datasets were described and classified into seven groups: datasets based on network traffic, datasets based on electrical networks, datasets based on the internet, datasets based on virtual private networks, datasets based on Android apps, datasets based on IoT traffic, and datasets based on internet-connected devices.

Recurrent neural networks, deep neural networks, restricted Boltzmann machines, deep belief networks, convolutional neural networks, deep Boltzmann machines, and deep auto-encoders are the seven deep learning models that we investigate. We examine the performance of each model for binary and multiclass classification under two brand-new real traffic datasets, the CSE-CIC-IDS2018 dataset, and the Bot-IoT dataset. Additionally, accuracy, false alarm rate, and detection rate are three of the most crucial performance measures we employ for assessing the effectiveness of various techniques.

The study introduced an intrusion detection tree ("IntruDTree") machine-learning-based security model that first considers the ranking of security elements according to their relevance before developing a tree-based generalized intrusion detection model based on the chosen essential features. This model decreases the computational complexity of the model by lowering the feature dimensions, making it effective in terms of prediction accuracy for test cases that have not yet been seen.

Finally, tests were run using cybersecurity datasets to test the performance of our IntruDTree model, and the precision, recall, F-score, accuracy, and ROC scores were calculated. We also contrast the IntruDTree model's performance with that of several well-established and well-liked machine learning techniques, including the naive Bayes classifier, logistic regression, support vector machine, and k-nearest neighbor to assess the efficacy of the resulting security model [17].

Zaib *et. al* [33] in the year 2020 is of the opinion that several industries are playing a crucial role in the development of smart cities, including intelligent transportation, cyber security, smart grids (SGs), and UAV-assisted next-generation connectivity (5G and B5G), among others. Big data analysis and the efficient application of AI, ML, and DRL-based approaches that can increase their efficiency and scalability in a smart city project have a significant impact on all the sectors mentioned above.

Attacks need to be identified before they can be prevented. IDS data sets have been constructed to simulate different attack types, and IDS systems have been designed to identify attack traffic. Several commonly used data sets, including CSE-CIC IDS-2018, UNSW-NB15, ISCX-2012, NSL-KDD, and CIDDS-001, were employed in this investigation [16].

Likewise, Thanh & Vijay 2021[29] surveyed Deep Reinforcement Learning (DRL) strategies created for cyber security is presented. Several important topics were

discussed, such as DRL-based security solutions for cyber-physical systems, autonomous intrusion detection methods, and multi-agent DRL-based game theory simulations for cyberattack defensive strategies

Miloud *et. al* [23] proposed that the study introduces a brand-new machine learning (ML)-based security framework that adapts automatically to the evolving security needs of the IoT domain. To reduce various vulnerabilities, this framework makes use of both Software Defined Networking (SDN) and Network Function Virtualization (NFV) enablers. This AI framework combines a monitoring agent with an AI-based reaction agent that analyses network patterns using ML-Models and detects intrusions into IoT devices based on anomalies to accomplish its objectives, the framework makes use of supervised learning, a distributed data mining system, and neural networks.

Hamed *et. al* [14] used the Bayesian theory, Naive Bayes classifier, Decision Tree, Random Decision Forest, Random Tree, Decision Table, and Artificial Neural Network are among the well-known machine learning classification algorithms in their study to deliver intelligent services in the realm of cyber-security.

“CrowdStrike's models are trained on the rich telemetry of the CrowdStrike Security Cloud, which correlates trillions of data points across CrowdStrike's asset graph, intelligence graph, and patented Threat Graph to deliver unmatched exposure and continuously improve threat intelligence across an organization's attack surface [10]

Wang, [31] insists that Machine learning can assist in generating policy suggestions for security equipment, such as firewalls. Instead of having users to manually navigate among competing access control lists for different devices and network segments, machine learning may deliver accurate suggestions that are automatically implemented.

Gordon & Matthew [13] said that with machine learning, businesses are becoming more proactive rather than waiting for cyberattacks to occur. Penetration testing simulates a cyberattack to find openings in a company's systems, networks, and firewalls. The issue is that online threats are evolving. It is nearly impossible to detect malicious data or code using techniques like steganography. Cyberattacks can also take on different shapes to escape detection, and fresh cyberattacks may exploit undiscovered system flaws. The cybersecurity sector is developing new roles and best practices to consider these possibilities. Before attacks occur, ethical hackers identify software flaws and patch them, but deception technology detects cyberattacks in their early phases.

Accuracy and efficiency are essential quality components to find the finest algorithms and models to identify IoT attacks in real-time or almost real-time [26]. We put forward six ideas for machine learning and IoT security research. This comprehensive literature evaluation of the most recent papers in IoT security identified a few key research trends that will direct future studies in this area. Given the exponential growth of large-scale IoT risks, it is imperative to develop models that can include cutting-edge big data and machine learning techniques and technologies.

An extensive examination of the most recent (from 2013 to 2018) works on machine learning in cybersecurity, describing the fundamentals of cyberattacks and corresponding defences, the fundamentals of the most well-known ML algorithms, and suggested ML and data mining schemes for cybersecurity in terms of features, dimensionality reduction, and classification/detection techniques [11].

Vivek *et. al* [30] proposed that Software Information and Event Management (SIEM) systems' job is to ensure an enterprise's cybersecurity. At the SIEM level, the system reports on any risky actions on the system as well as any intrusion attempts made by malicious users. However, many of these warnings are untrue and shouldn't be ignored to focus on the system's more pressing problems, such as intrusion detection, and vulnerable ports. Machine learning can efficiently assist us in studying the system across all safety parameters to detect all risks on the system and categorize them by the seriousness of the alert and the frequency at which that particular alert is coming to the system.

Furthermore, (Yuantian Miao et al, 2022) [32] proposed that to create an effective defense, these three factors namely: detection, disruption, and isolation countermeasures are suggested.

Additionally, the outcomes show that the suggested IoT architecture based on the extreme gradient boosting (XGBoost) machine learning technology can effectively display network hacks as well as all GIS flaws with various alarms. Additionally, to improve decision-making regarding the GIS state, the flaws in GIS and fake data caused by cyberattacks are identified and provided on the dashboard of the suggested IoT platform with high accuracy and clearer visualization [20]

The difficulties of using deep learning to solve security issues in 5G heterogeneous networks are examined. The outcomes demonstrate that the modulation recognition problem can be successfully solved by the deep learning model and that the modulation mode of the convolutional neural network can successfully recognize the modulation signals used in the experiment. Deep learning hence provides a significant advantage in the modulation recognition problem. Furthermore, when compared to the conventional algorithm, the unsupervised beamforming algorithm based on deep learning proposed in this research can significantly reduce the computational complexity for various numbers of transmitting antennas, demonstrating its superiority over the conventional algorithm [34].

Lastly, Binbin *et al*, [8] think that it is appropriate to apply crucial machine learning models to detect web shells and provide security solutions for IoT networks to create a secure IoT system. Future ensemble techniques will be utilized to enhance the performance of these machine learning models, such as random forest (RF), highly randomized trees (ET), and voting

Table 1: Application for Prevention: A Timeline of Machine Learning Developments for Three Major Cybersecurity Tasks [22]

	Pre-1990s	1990s	2000s	2010s
SPAM DETECTION	**1978:** First spam email	Spam continues to worsen due to growth in email **1996:** First spam blockers	**2002:** Machine learning methods first proposed for spam detection **2003:** First attempts to regulate spam in the United States	Machine learning spam detection widely embedded in email services Emergence of deep learning-based classifiers
INTRUSION DETECTION	**1980:** First intrusion detection systems **1986:** Anomaly detection systems combine expert rules and statistical analysis	**Early 1990s:** Neural networks for anomaly detection first proposed **1999:** DARPA creates datasets to study intrusion detection systems	Machine learning further studied as a possible tool for misuse-based and anomaly-based intrusion detection	**Late 2010s:** Emergence of large-scale, cloud-based intrusion detection systems Deep learning studied for intrusion detection
MALWARE DETECTION	**Early 1980s:** First viruses found "in the wild" **Late 1980s:** First antivirus companies founded	**Early 1990s:** First polymorphic viruses **1996:** IBM begins studying machine learning for malware detection	**Early 2000s:** First metamorphic viruses Wide number of traditional machine learning methods studied to detect malware	Rise of "next-gen" antivirus detection Emergence of ML-focused antivirus companies

Elsevier Science Direct journal concludes that Data sets were normalized, and classification was carried out using traditional machine learning techniques such as support vector machines (SVM), K-Nearest Neighbor (KNN), and decision trees (DT). As a result, several of the investigations mentioned in the literature have produced more fruitful outcomes. This work is seen as helpful for creating IDS systems based on artificial intelligence using strategies like machine learning [19]

2.1 ***Cybersecurity Benefits of Machine Learning*** [10]

Applying machine learning to cybersecurity-related issues has various advantages. These consist of:

1. Synthesize massive amounts of data quickly: One of the main issues analysts have is the requirement to quickly synthesize intelligence gathered from throughout their attack surface, which is often generated far more quickly than their teams can manually evaluate. Large volumes of historical and dynamic intelligence may be quickly analyzed by machine learning, allowing data operationalization teams from diverse sources in close to real time.
2. Turn on scaled expert intelligence by allowing models to learn from their changing sample population, which may include analyst-labeled detections or analyst-reviewed alerts, through consistent training cycles. Additionally, it gives models the ability to understand and apply expert-generated ground truth, preventing recurrent false positives.
3. Automate tedious, manual operations: By applying machine learning to particular jobs, security teams can be relieved of tedious, repetitive duties. The machines will be able to transfer time and resources to complex, strategic projects as a result, scaling their response to incoming alerts.
4. Increase analyst productivity: Machine learning may augment analyst expertise with up-to-date, real-time intelligence, enabling analysts in threat hunting and security operations to effectively allocate resources to solve their organization's most urgent vulnerabilities and look into time-sensitive issues, detections with ML alerts.

2.2 ***Challenges of Machine Learning in Cybersecurity*** [31]

1. **The first is the substantially higher accuracy standards:** For instance, it might be unpleasant, but it probably won't have a life-or-death effect if you're only conducting image processing and the system misidentifies a dog as a cat. The impact of the incorrect classification can be serious if a machine learning system mistakes a fraudulent data packet for a valid one and this leads to an attack against a hospital and its gadgets.

2. **Access to a lot of training data, particularly labeled data, is a second challenge:** For machine learning to produce more accurate models and forecasts, a lot of data must be collected. It is much more difficult to obtain malware samples than it is data for image processing or NLP. There is a lack of attack information, and a lot of security risk information is sensitive and unavailable due to privacy issues.

3. The **third challenge is the reality of The ground truth**: cybersecurity may not always be available or fixed, unlike images. The cybersecurity environment is dynamic and constantly shifting. The world's malware is constantly evolving, and no single malware database can be said to contain all of it. What is the reality that we should consider?

4. **A lack of talent:** For ML to be useful in any field, we need to combine domain experience with ML knowledge. It is difficult to locate professionals who are knowledgeable in both ML and security; rather, either ML or security alone lacks skill. We discovered that, even though they don't share the same language, employ distinct methodologies, and take different approaches, it is crucial to ensure that ML data scientists and security experts collaborate. They must learn how to cooperate. Applying ML to cybersecurity effectively depends on cooperation between these two groups.

5. **Making and ensuring the ML we use for cybersecurity is secure on its own** is even more important given the crucial role cybersecurity plays in every business. Academic research has been done in this area, therefore we are happy to see and support the commercial drive toward safeguarding ML models and data. Palo Alto Networks is promoting innovation and taking every precaution to guarantee the security of our ML.

6. **Explainability of machine learning models** is the sixth and last challenge. For us to behave appropriately, having a thorough knowledge of the machine learning outcomes is essential.

3. Architecture/conceptual design

Machine Learning in Cybersecurity operations is an interesting topic, and it has been influenced by the computerization of many different application domains, including e-commerce platforms such as banking, business, medicine, and many other crucial fields [21]. A critical problem is to recognize the different network assaults, especially those that have never been encountered before. The architectural framework in Fig 2 shows that the machine is built using a bottom-up clustering technique; it gathers information, patterns, and sequences before gradually combining them into sizable clusters. According to such procedures, clusters are created by joining existing clusters that are close together. Using the KNN rule, the target knowledge is searched for and compared to a set of predefined rules and sequences and the algorithm compares the target data at first with the network's historical behavior and a set of trained tagged data that contain information on damaging data to the system. The main goal here is to create a system that examines common network trends and behavior and gradually learns to distinguish between regular risks and typical data [28].

This network system (fig 2) was built for 'Host-based intrusion detection', we frequently keep an eye on host logs. It can record incursions in terms of frequently occurring patterns, very effective attacks, or network node vulnerabilities, these nodes include all types and forms of computational devices.

Here, we create a misuse detection version using the MLP algorithm. A multilayer perceptron is a feedforward variant of an artificial neural network that examines the known facts and produces outputs that are hard and fast accurate and correct. In a directed network, an MLP consists of a few layers of nodes, each of which is inextricably related to the node below it. Backpropagation is a supervised learning technique that is used by MLP to train the network.

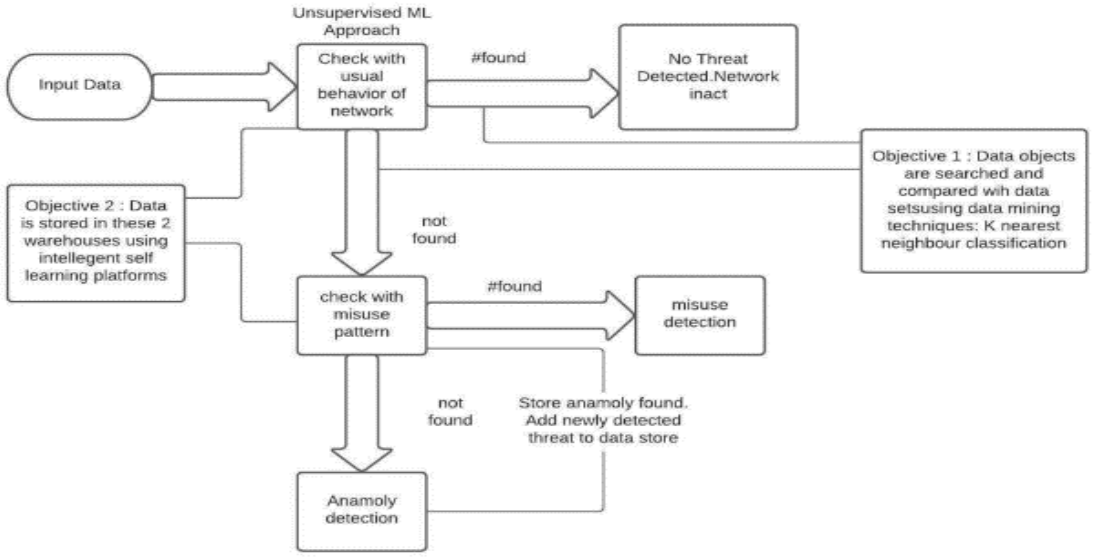


Fig 4: Architectural Diagram [30]

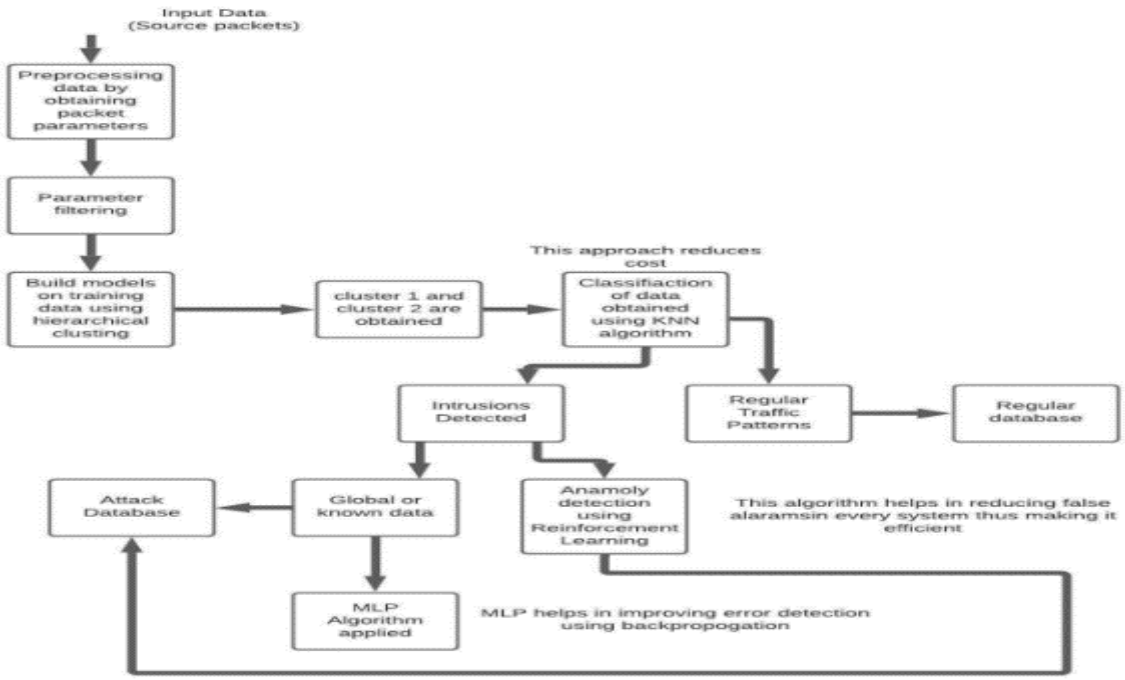


Fig 5: Workflow of the Architectural Diagram (Vivek et al, 2021) [30]

There is an excessive level of false alarms in anomaly detection. To deal with such issues, reinforcement learning is used, where the network is trained to make decisions and identify potential risks. This device (fig 3) uses a reinforcement signal that is sent to the fusion center using the environment to alter the weights defining each agent's selection capability and the weights expressing their trust in making decisions in person. The computer no longer wants to waste resources responding to a false threat thanks to this algorithm's reduction in the number of false alarms.

3.1 Establishing Solutions for Machine Learning to Boost Cybersecurity

A developer must comprehend the function that machine learning and deep learning play in cybersecurity. You may create cutting-edge, reliable security software by using machine learning ML model operations. But first, you must be familiar with the procedures and tools that will aid in the creation of ML models. Additionally, effective model management is necessary for success [1]. One can manage the model registry and obtain good results from it by doing this.

1. Closing the Skills Gap In Cybersecurity:

Different types of businesses are becoming increasingly concerned about cybersecurity. In the current insecure digital environment, no one can feel secure or unaffected by attacks. This has forced businesses to hire more cybersecurity specialists to improve their ability to defend against complex attacks.

2. Supports Task Automation:

Daily responsibilities that never seem to end are handled by security personnel and business owners. The primary drawback is that the majority of these duties are repetitive. They are being compelled by this to investigate alternate options like automation, which developers like you can assist them in putting into practice. Each firm can benefit from machine learning. As a developer, you ought to encourage your clients to use it for task automation for this reason. A firm benefits from automating operations with machine learning and developing models to streamline procedures because:

I. Simple malware detection
II. Analyze threats to a specific vulnerability quickly
III. Facilitate the work of security personnel.

IV. Accelerates the identification and reaction to threats

3. Facilitates the detection and classification of threats:

You are aware of the advantages of software security testing as a developer. It enables you to advertise flawless, high-caliber software. Businesses might gain from automating threat detection and classification using machine learning (ML). This is a crucial stage in any network's security. Large data sets are ideal for machine and deep learning analysis. It can benefit a company by:

i. Recognize harmful behavior and act immediately to address it.
ii. Utilize signs from its database to find persistent security concerns.

4. Inhibits phishing:

One of the prevalent attack methods being utilized by cybercriminals is phishing. By educating employees about phishing, businesses may stop it from happening. The knowledge they gain from training can then be applied to spot phishing emails, links, and websites.

5. Endpoint Security:
Keeping an organization's endpoints contained is one of the best ways to keep it safe. A company may be exposed to a sizable number of threats because of infected hosts, endpoints, and devices. A corporation requires machine learning since it can aid with endpoint protection.

6. Contributes to Network Risk Scoring:
Network analysis aids organizations in future assault prevention. A business can dedicate additional resources to secure the network's weak points if it is aware of them. The best course of action is to assess previous dangers and pinpoint the openings that intruders used to access the network.

7. Encourages Human Contact:
The way that people interact with technology is a crucial consideration when trying to secure a company network. In actuality, technology will never be able to fully replace people. It can only improve human performance and increase production in less time and at a cheaper cost. (https://www.codemotion.com/magazine/backend/cybersecurity/automating-cybersecurity-with-machine-learning/

4.0 APPLICATION

Here, the cybersecurity model's preventive phase comprises defenders' attempts to find and patch vulnerabilities to thwart potential threats. It has long been an aim to create tools that can automatically find and address new vulnerabilities, but machine learning has only recently proven to be a practical way to do this.

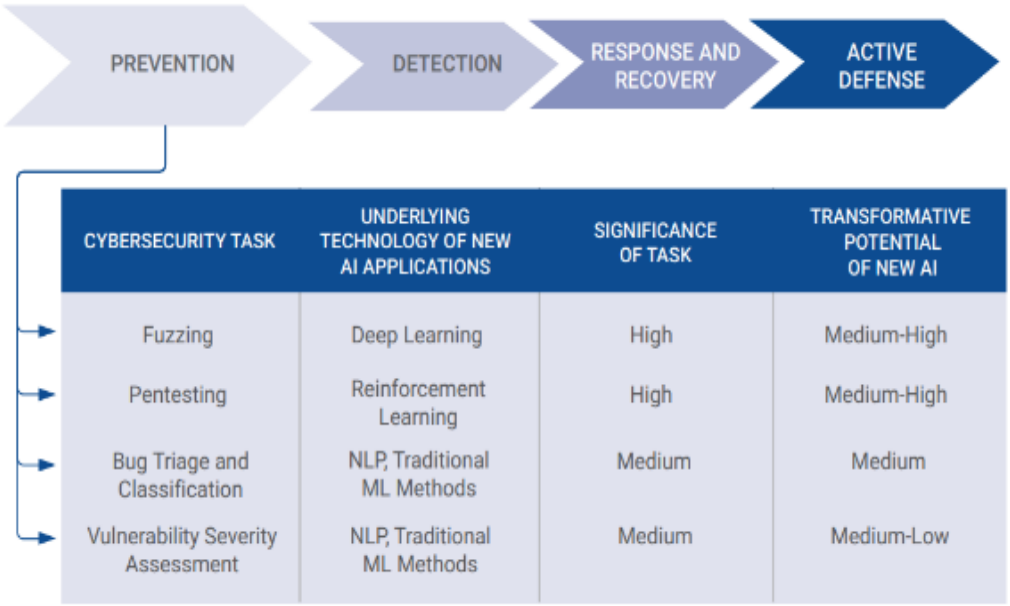


Fig 6: Machine Learning and AI Application for Prevention of Vulnerabilities [22]

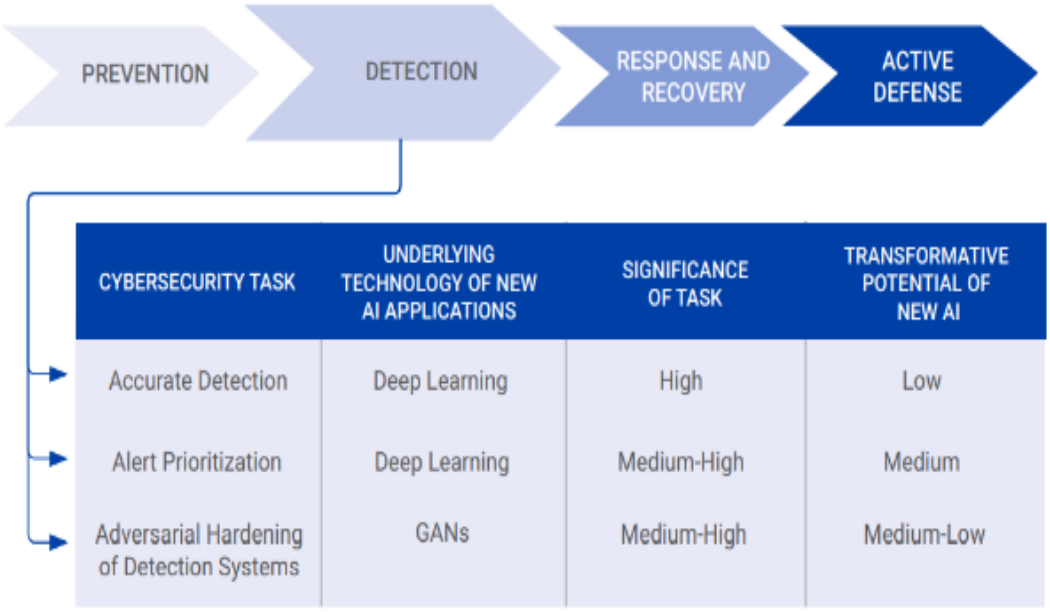


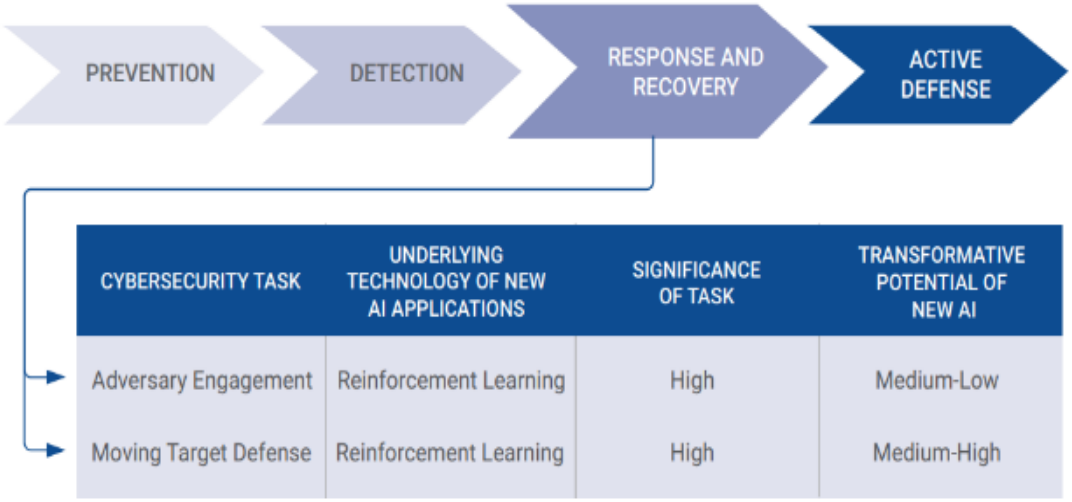
Fig 7: Machine Learning and AI Application for Detection of vulnerabilities [22]

Due to the cyber threat environment, it is necessary to continuously track and correlate the ever-changing external and internal data points across the infrastructure and users of a business. It is just not practical to manage this amount of information with just a small group of people. This is where machine learning excels since it can rapidly analyze huge data sets to find patterns and predict threats. By automating the analysis, cyber teams may immediately spot threats and pinpoint situations that need further in-depth human study.

By continuously observing network behavior for anomalies, machine learning detects threats. Machine learning engines quickly process massive amounts of data to uncover noteworthy situations. These methods enable the detection of unknown malware, insider risks, and policy violations. The primary application area where deep learning and more recent ML techniques are seen to be potentially disruptive forces is detection, at least among many public-facing sites. Unfortunately, machine learning has not yet produced the game-changing innovations that many had hoped for. Even if sufficiently large models, especially at sufficiently large scales, do tend to perform marginally better than simpler models, these gains are occasionally offset by the increasing number of risks that the majority of organizations must contend with.

The main line is that many cybersecurity organizations still heavily rely on simpler models today, despite the crucial role that deep learning has played in the Machine Learning advancements of the last half-decade. Online "bad neighborhoods" can be anticipated by machine learning to assist in preventing users from connecting to dangerous websites. To automatically detect attack infrastructures prepared for existing and emerging threats, machine learning monitors Internet traffic [9].

While ML-driven detection systems are the subject of extensive research, more ambitious suggestions envision AI systems that could one day move around networks on their own, repairing holes and engaging in dynamic defense against attackers.



CYBERSECURITY TASK	UNDERLYING TECHNOLOGY OF NEW AI APPLICATIONS	SIGNIFICANCE OF TASK	TRANSFORMATIVE POTENTIAL OF NEW AI
Adversary Engagement	Reinforcement Learning	High	Medium-Low
Moving Target Defense	Reinforcement Learning	High	Medium-High

Fig 8: Machine Learning and AI Application for Recovery and Response in time of vulnerabilities [22]

5. PROPOSED ARCHITECTURE

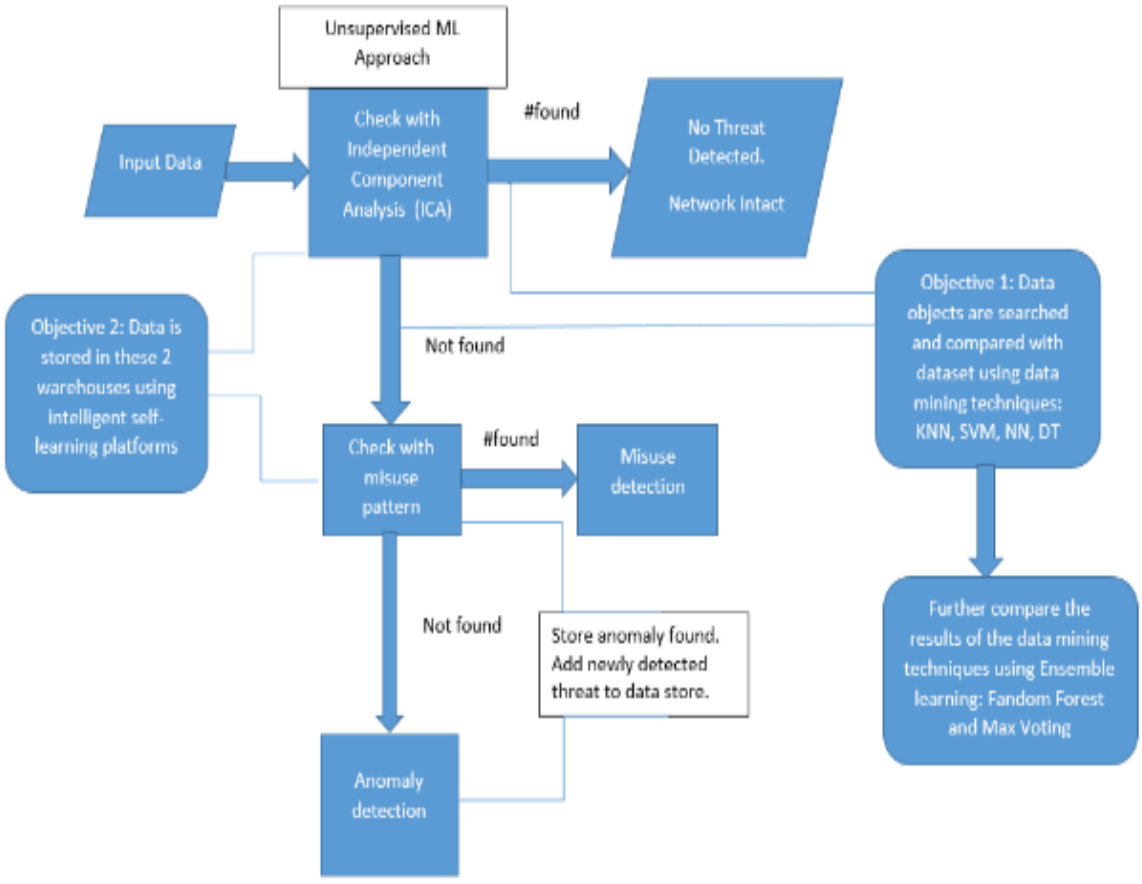


Fig 9: Proposed Architecture for Cyber-attack/security using Machine and Ensemble Learning

Using the KNN rule, searching for and comparing the target information to a set of established rules and sequences [30]. This paper is an improvement on Vivek *et al,* [30] algorithm that relates the target data at first with the network's historical behavior and a set of trained tagged data that contain information on damaging data to the system. The main goal here is to create a system that examines common network trends and behavior and gradually learns to differentiate between regular risks and typical data by using other machine learning algorithms alongside KNN such as SVM, Neural Networks, Decision Trees, and Ensemble learning such as Random Forest and Max voting to further analyze the results of the machine learning. We frequently check host logs for host-based intrusion detection. It can record incursions in terms of characteristics, very effective attacks, or device vulnerabilities.

6. Conclusion

Here, it is concluded that the fact that Machine Language has a significant impact on cyber security operations, the systems used in carrying out these operations are still susceptible to various sorts of attack that do not apply to other types of detection systems and this is a major drawback of Machine Learning-based detection that is occasionally disregarded in popular coverage. Attackers frequently find "adversarial examples" i.e. slightly altered inputs that dramatically alter a model's response despite being undetectable to a human, because the process by which many Machine Learning ML systems reach decisions can frequently be poorly understood and highly sensitive to small changes that a human analyst would view as trivial.

The usage of Machine Learning models also creates additional attack vectors; in addition to maintaining the model's security, defenders must ensure that their data is not contaminated and that the (usually open-source) algorithms and statistical software are secure. A model architecture is proposed using more machine learning algorithms and ensemble learning methods to check the performance of the data set in Fig 9.

References

[1] Apriorit. (2022, January 27). *Implementing Artificial Intelligence and Machine Learning in Cybersecurity Solutions*. Retrieved from apriorot.com: https://www.apriorit.com/dev-blog/474-ai-ml_cybersecurity#:~:text=Machine%20learning%20in%20cybersecurity%20is,on%20supervised%20and%20unsupervised%20learning

[2] Azeez, N.A.; Salaudeen, B.B.; Misra, S.; Damasevicius, R.; Maskeliunas, R. Identifying phishing attacks in communication networks using URL consistency features. Int. J. Electron. Secur. Digit. Forensics 2020, 12, 200–213.

[3] Azeez, N.A.; Odufuwa, O.E.; Misra, S.; Oluranti, J.; Damaševičius, R. Windows PE Malware Detection Using Ensemble Learning. Informatics 2021, 8, 10. https://doi.org/10.3390/informatics8010010

[4] Azeez, N.A.; Salaudeen, B.B.; Misra, S.; Damasevicius, R.; Maskeliunas, R. Identifying phishing attacks in communication networks using URL consistency features. Int. J. Electron. Secur. Digit. Forensics 2020, 12, 200–213.

[5] Azeez, N.A, Vyver, C.V "Towards a Dependable Access Framework for E-Health", 2017 International Conference on Computational Science and Computational Intelligence (CSCI), pp.1695-1701, 2017.

[6] Azeez, N.A; Vyver, C.V "Dynamic Patient-Regulated Access Control Framework for Electronic Health Information", 2017 International Conference on Computational Science and Computational Intelligence (CSCI), pp.1684-1690, 2017.

[7] Azeez, N.A; Ade, J; Misra, S; Adewumi, A; Vyver, C.V; Ahuja, R "Identifying Phishing Through Web Content and Addressed Bar-Based Features",

[8] Binbin Yong et al. (2020, August 18). *Ensemble machine learning approaches for webshell detection in Internet of things environments.* Retrieved from Wiley Online Library: https://doi.org/10.1002/ett.4085

[9] Cisco. (n.d.). *Machine learning security*. Retrieved from Cisco.com: https://www.cisco.com/c/en/us/products/security/machine-learning-security.html

[10] CrowdStrike. (2022, September 14). *MACHINE LEARNING (ML) & CYBERSECURITY.* Retrieved from Crowdstrike: https://www.crowdstrike.com/cybersecurity-101/machine-learning-cybersecurity/

[11] Dasgupta et al. (2020). Machine learning in cybersecurity: a comprehensive survey. *Journal of Defense Modeling and Simulation, Research Gate*, 1-50.

[12] Github.com

[13] Gordon , G., & Matthew , U. (2022, Nov 17). *Machine Learning in Cybersecurity: How It Works and Companies to Know.* Retrieved from BuiltIn: https://builtin.com/artificial-intelligence/machine-learning-cybersecurity

[14] Hamed , A., Iqbal , S. H., Asra , K., Syed Md. , H. M., Sheikh, I., & Sohrab , H. (2020). Cyber Intrusion Detection Using Machine Learning Classification Techniques. *Springer*, 1235.

[15] IBM cloud hub. (2020, August 19). *IBM cloud education.* Retrieved from ibm.com: https://www.ibm.com/cloud/learn/supervised_learning#:~:text=Supervised%20learning%2C%20also%20known%20as,data%20or%20predict%20outcomes%20accurately.

[16] Ilhan , F. K., Fatih , E., & Abdulkadir , S. (2021). Machine learning methods for cyber security intrusion detection: Datasets and comparative study. *Elsevier, ScienceDirect*, 188.

[17] Iqbal et al. (2020). IntruDTree: A Machine Learning Based Cyber Security Intrusion Detection Model. *MDPI Symmetry*, 12(5).

[18] javatpoint. (2021). *Types of machine learning*. Retrieved from javatpoint: https://www.javatpoint.com/

[19] Kilincer, I. F., Ertam, F., & Sengur, A. (2021). Machine learning methods for cyber security intrusion detection: Datasets and comparative study. *Science Direct*, 188.

[20] Mahmoud Elsisi et al. (2021). Towards Secured Online Monitoring for Digitalized GIS Against Cyber-Attacks Based on IoT and Machine Learning. *IEEE* , vol 9.

[21] MAMIDALA, R. (2000). AI in Cyber security.

[22] Micah Musser, A. G. (2021). *Machine Learning and Cyber Security (Hype and Reality).* George Town: Centre for Security and Emerging Technologies, CSET.GEORGETOWN.EDU.

[23] Miloud et al. (2020). A Machine Learning Security Framework for Iot Systems. *IEEE*, 114066 - 114077.

[24] Mohamed et al. (2020). Journal of Information Security and Applications. *Elsevier*, 50.

[25] Rajbanshi, S. (2021, March 30). Analytics vidhya. *Everything you need to know about Machine Learning*.

[26] Rasheed, A., & Izzat , A. (2021). Machine learning approaches to IoT security: A systematic literature review. *Elsevier, Science Direct*, Volume 14

[27] Sanjay, T. (2020). Machine Learning in Cyber security.

[28] Sreekesh, D. a. (2016). A two-tier network based intrusion detection system architecture using machine learning approach. *International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, 42-47

[29] Thanh , T. N., & Vijay , R. J. (2021). Deep Reinforcement Learning for Cyber Security. *IEEE*, 1-17.

[30] Vivek et al. (2021). A Machine Learning Framework for Cyber security Operations. *International Journal of Innovative Science and Research Technology*, 5.

[31] Wang, M. (2022, September 7). *the-future-of-machine-learning-in-cybersecurity.* Retrieved from BrandPost Sponsored by Palo Alto Networks : https://www.cio.com/article/406441/the-future-of-machine-learning-in-cybersecurity.html

[32] Yuantian Miao et al. (2022). Machine Learning–based Cyber Attacks Targeting on Controlled Information: A Survey. *ACM Digital Library*, 1-36

[33] Zaib et al. (2020). Applications of Artificial Intelligence and Machine learning in smart cities. *Elsevier ScienceDirect*, 313-323

[34] Zhihan Lv et al. (2021). Deep Learning for Security Problems in 5G Heterogeneous Networks. *IEEE Xplore*, 67-73.