A Literature Review on Machine Learning for Cyber Security Issues

Bhojraj Singh 1, Amit Chaudhary 2, Karan Chauhan 3

Department of Computer Science and Engineering, Galgotias University, Greater Noida, Uttar Pradesh India

# ABSTRACT

Machine learning's main goal is to automate human assistance by using pertinent data to create an algorithm. Machine learning (ML), a branch of artificial intelligence (AI), is concerned with creating machines that can learn from historical data, identify patterns, and draw logical conclusions with little to no assistance from humans. Cyber security is the idea of defending digital systems, such as computers, servers, mobile devices, networks, and the data they are connected to, against malicious attacks. The two key components of merging cyber security and ML are accounting for cyber security where machine learning is utilized and employing machine learning to facilitate cyber security. This union may help us in a number of ways, such as improving the security of machine learning models, boosting the efficacy of cyber security strategies, and facilitating the quick and easy identification of zero-day vulnerabilities with little to no human intervention. In this review paper, we integrate machine learning and cyber security to discuss two different concepts. We also discuss the advantages, challenges, and issues associated with integrating ML with cyber security. Furthermore, we investigate many attacks and provide an extensive examination of diverse strategies in two distinct groups. Lastly, we conclude with some recommendations for additional study. Keywords: Machine learning, algorithms, detection, and problems with cyber security.

# Introduction

The majority of the gadgets we use in the contemporary computer age are now a part of an Internet-connected ecosystem known as the Internet of Things (IoT). These devices exchange and send their data via the Internet, an insecure (open) communication channel. These are typically private bits of information (social security numbers, banking information, insurance information,

healthcare information, and other finance-related information). Hackers, or online attackers, are always searching for new ways to trick people. Some of the assaults they can perform include man-in-the-middle, replay credential guessing, impersonation, malware insertion, session key computation, and data manipulation.

Consequently, a number of researchers periodically suggest various security precautions to decrease these risks. The following categories comprise security protocols, often known as cyber security protocols: key management protocols, intrusion detection protocols, access control protocols, authentication protocols, and security strategies utilizing block chains. Below is a summary of these protocols.

* **Authentication protocols:** Finding out if a person or a device is who they say they are is the process of authentication. It can be carried out with credentials or elements that are specific to the user or device (e.g., smartcard, biometrics, password, and username). Device to device, user to user, and user to device are the three categories of authentication. Depending on the available factors, user authentication protocols can also be divided into three groups: one-factor, two-factor, and three-factor authentication protocols.
* **Protocols for access control:** Access control is the process of preventing unauthorized access to someone or something (s). Users or devices can safely communicate with one another after following all the stages of a user/device access control protocol. Access control protocols include those for block chain enabled security, intrusion detection, and key management. A synopsis of these protocols is given below.
* **Authentication protocols:** Finding out if a person or a device is who they say they are is the process of authentication. It can be carried out with credentials or elements that are directly associated with the person or device (e.g., smartcard, biometrics, passwords, and usernames). There are three types of authentication: device to device, user to user,

and user to user. Based on the available factors, user authentication protocols can also be divided into three groups: one-factor, two- factor, and three-factor authentication protocols.

* **Access control protocols:** The method of restricting unauthorized access to a person or a device (s). Users and devices can safely access one another after following every step of a user/device access control protocol. User access control and device access control are the two categories of access control protocols. Unauthorized devices can be controlled using the device access control protocol, and unauthorized users can be controlled via the user access control protocol. There are two types of access control methods: certificate- based and certificate-less. Authorization is the process by which an authority (a server) determines if an entity (a client) is authorized to use a resource. It is usually done in conjunction with authentication so that the server can identify the client submitting the access request. It determines who is allowed.
* **Protocols for detecting intrusions:** Anything or anyone acting maliciously is considered an incursion. This might be a malicious computer script or a system under the control of hackers that targets the Internet. Hackers typically try to infect online devices with malware in an effort to reduce their functionality or jeopardize their security (systems). For the purpose of detecting and mitigating intrusions, we need a certain class of protocols that are categorized as "intrusion detection protocols". Numerous techniques are available for carrying out the intrusion detection, such as hybrid intrusion detection, anomaly-based intrusion detection, and signature-based and anomaly-based intrusion detection. These days, machine learning and deep learning-based malware detection methods are growing in popularity.

# Related Work

This research aims to improve internet security by proposing an Intrusion Detection System (IDS) based on Convolutional Neural Networks. To identify network intrusions, the proposed IDS paradigm divides all network packet traffic into benign and malignant categories. The proposed model was trained and validated using the CICIDS2017 dataset from the Canadian Institute for Cyber security. The model's

overall accuracy, attack detection rate, false alarm rate, and training overhead have all been assessed. The effectiveness of the suggested model has been evaluated with nine other well-known classifiers.

Assist vector machines, ANNs, CNNs, Random Forests, and significant learning estimations have all been given reasonable assessments based on the most current CICIDS2017 dataset. Out of SVM, ANN, RF, and CNN, significant learning estimation performed the worst. Ultimately, we will leverage this dataset to do port scope attacks akin to other types of attacks by merging artificial intelligence and large-scale learning computations with Hadoop and shimmer advancements. By forecasting the four algorithms—SVM, ANN, RF, and CNN—this study ascertains which algorithm has the highest accuracy rates for predicting the best outcomes to assess whether or not a cyber-attack happened.

IEEE Member This study employs anomaly-based machine learning models to examine network intrusion attempts, providing a higher level of protection than to the conventional misuse-based methods. Two models were developed and implemented using a data set obtained from an actual, institutional production setting: an ensemble learning model and a convolutional neural network model. The UNSW-NB15 benchmarking data set was utilized to demonstrate the validity and reliability of the models. Limiting the type of attack to probing attacks, the study's scope was kept minimal. The results indicated high accuracy rates and that the CNN model was somewhat more accurate.

# UNITING CYBER SECURITY AND MACHINE LEARNING

## Cyber security and machine learning

In the cyberspace, connected systems can be subjected to a variety of attacks, including replay, man-in-the- middle (MiTM), impersonation, password guessing, credentials leakage, unauthorized data update, malware injection, flooding, denial of service (DoS), and distributed denial of service (DDoS). Therefore, we must some kind of security standard in order to identify and prevent these assaults. The machine learning models (machine learning ML algorithms) may learn about different cyber-attacks in the offline and online modes using the pre-processed dataset that is provided. In real time or in online mode, the machine learning algorithms detect any hint of infiltration (like a cyber- attack).

The scenario of "machine learning in cyber security" is depicted in Figure 1.

In this instance, a machine with an Internet connection (a laptop, desktop, smartphone, or Internet of Things device) can be utilized to do a range of online tasks, including financial transactions, online access to medical records, social security numbers, etc. Hackers search these systems for vulnerabilities all the time, and when they do, they attack. For the identification and mitigation of cyber-attacks, a variety of ML techniques, such as supervised learning, unsupervised learning, reinforcement learning, and deep learning, can be applied depending on the environment. The communication environment and resources available to a system determine which learning method— supervised learning, unsupervised learning, reinforcement learning, or deep learning—is most appropriate for it. Due of cloud servers' superior.

## Machine learning and cyber security

A scenario for "cyber security in machine learning," or machine learning (ML) security, is shown in Figure 2. ML models are used for a variety of event analysis and forecasting purposes. Nevertheless, certain assaults might have a detrimental effect on the performance of ML models, such as dataset poisoning and model poisoning disruption attacks [6]. Machine learning (ML) models may forecast the related phenomena erroneously as a result of these attacks. A "dataset poisoning attack" occurs when an attacker inserts adversarial examples (updated values) into the dataset, causing the machine learning model to provide inaccurate predictions. The "model poisoning attack" aims to corrupt the models by altering their settings and interfering with their internal workings.

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.

The attacker conducts a "privacy breach attack" in an attempt to expose personal data while also attempting to collect the model's relevant knowledge. An assault using membership inference is considered a privacy breach. In a "runtime disruption attack," the attacker

also targets the model's execution process, which affects the prediction results' accuracy and subverts the machine learning workflow. Therefore, specific cyber security approaches (such as encryption techniques, signature generation and verification procedures, and hashing processes) are required in order to fight against these assaults. The anticipated outcomes are accurate and the ML models and associated datasets are protected by these cyber security protocols.