JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Future Challenges and Convalescences in Machine Learning-based Cybersecurity

Dr. Lalit Kishore[1]
AP, MCA Deptt, SRM-IST, NCR Campus, Modinagar, Ghaziabad
Mr. Naved[2]
M.Tech (CS) Scholar, R.D Engineering College, Murad Nagar, Ghaziabad

Abstract: The threats of malware and cyber-attacks are rapidly increasing, and network protection and cyber security are the major issues in now a day for online communities. But the machine learning opens the new doors of huge opportunities in cybersecurity, because traditional approaches of network protection need human intervention to find and remove the vulnerability. Implementation of machine learning in cyber security made the stronger malware detection process, more active, accessible, and efficient than previous methods. The cybersecurity field facing several challenges in machine learning techniques which are need to be improved effectively. This paper includes various machine learning approaches which proven effective implementation to remove and detect most of the cyber-attacks. These approaches and models are most effective to develop reliable and secure systems.
Keywords: cybersecurity; machine learning; cyber-attacks; network protection.

1. Introduction

Due to the rapid and substantial rise in the significance of information technology over the past few decades, a multitude of security incidents have experienced a remarkable surge. These incidents encompass a range of unauthorized activities, including but not limited to illegal accessibility, service denial, worm infiltrations, data leakage, and re-analysis or hacking scams. The frequency of these incidents has escalated exponentially throughout the previous decade.

To illustrate this trend, in the year 2010, the security field documented a tally of under 50 million distinct executable malware files. By the time 2012 arrived, this documented figure had surged twofold to an approximate 100 million. Notably, as per statistics provided by AV-TEST, the security sector encountered a staggering 900 million executable instances in the year 2019, with this number continuing to expand.

The repercussions of e-crime and cyber war are profound, resulting in substantial monetary repercussions for both organizations and individuals. Therefore, the cost of an average data breach in the United States is projected at $3.9 million, while the global estimate reaches 8.19 million. Moreover, the worldwide economy bears the weight of a $400 billion annual expense due to cybercrime.

Projections from the security community indicate that this number is poised to experience an almost fourfold increase over the next five years, setting new records. Consequently, businesses are compelled to formulate and execute a comprehensive cybersecurity strategy to mitigate further financial setbacks. Recent socio-economic analyses underscore the imperative for governments and individuals to have secure access to data, applications, and tools, as this directly impacts national security.

Cybersecurity is exercise of protecting computer and networks from unwanted use, reading, release, disturbance, alteration, or damage [15…]. It is a broad term that encompasses a several security issues, such as:

- Physical safety: This includes measures to protect computer systems and networks from physical actions, including unwanted use to data centers or server rooms.
- Network security: This includes measures to to secure data from unwanted use, such as firewalls and intrusion detection systems.

- Information security: This includes measures to secure data from unwanted use, access, leakage, or alteration, such as encryption and access control lists.
- Application security: This includes measures to protect applications from security vulnerabilities, such as code reviews and penetration testing.

Conventional cybersecurity solutions include elements like utility software of antivirus and firewall, and systems to detect intrusion, integrated into network and computer security setups. The ongoing evolution of data science, particularly machine learning—an integral facet of "artificial intelligence"—holds significant potential in uncovering concealed patterns within data. This transformative role is revolutionary and new scientific model of data science, notably influencing the landscape of cybersecurity[2,3…]. Addressed in [4..], the progression of technologies associated with launching cyber threats has empowered attackers, rendering them more proficient, which in turn has led to a surge in interconnected technologies.

In 2015, both the fields of Cyber Security and Machine Learning exhibited popularity values below 30. However, these values are projected to surpass 70 by the year 2023, showcasing a significant surge in popularity, more than doubling within this time span. This study is primarily centered around the intersection of machine learning and cybersecurity. This intersection is rooted in their shared focus on decision making systems, safety, and different methods to process data, all aimed at real-life application.

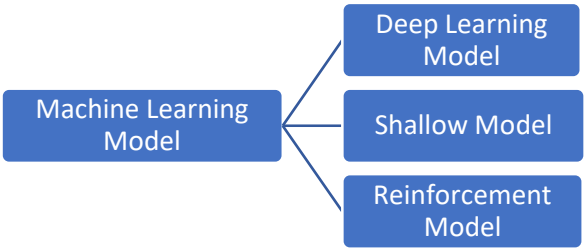
The primary focus of this research revolves around the utilization of machine learning algorithms on security data to assess cyber risks and optimize cybersecurity procedures. This endeavor holds relevance not only for academics but also for industrial researchers who are keen on exploring and crafting data-driven intelligent models for cybersecurity using machine learning methodologies.

Machine learning represents better option from previous methods to solve problems, such as user verification, accessibility monitoring, antivirus, and cryptographic models. It is not sure the these traditional approaches adequately cater to the dynamic cybersecurity requirements of the present era[16–18]. A significant issue arises when it comes to manually addressing these solutions in scenarios necessitating ad hoc data management [7…]. As the landscape of cybersecurity continues to witness a growth of various incidents, conventional methods are proving inadequate in managing the associated risks. This shortfall has led to the emergence of novel and intricate attacks that propagate rapidly through networks. Consequently, researchers are resorting to different data analysis and models for extracting the information to construct Cyber Security frameworks, which are discussed in next section. These models hinge on effectively identifying security insights and staying abreast of the latest security trends, which can be of more pertinence.

The research underscores the imperative of crafting adaptable and efficient security systems capable of responding to and mitigating attacks while dynamically updating security protocols to counteract them intelligently and promptly. Achieving this demands the analysis of a substantial volume of pertinent cybersecurity data gathered from varied sources like network and system resources. Furthermore, these techniques ought to be applied in a manner which improve less human efforts and made them fully automatic.

2. ML Techniques used for Cyber security

Mostly, Machine learning (ML) referred as an aspect of "artificial intelligence," intricately intertwined with statics and the concept of data mining. Its primary focus revolves around empowering systems to assimilate insights from old data[67,68]. Consequently, these models often encompass an assemblage of regulations, procedures, and intricate functions or expressions. Above attributes may be harnessed to unearth stimulating forms within data, identify sequences, or forecast behaviors[10…]. This positions ML as a potentially valuable asset within the realm of cybersecurity. In Figure 4, there is an overview of the frequently employed machine learning concepts.

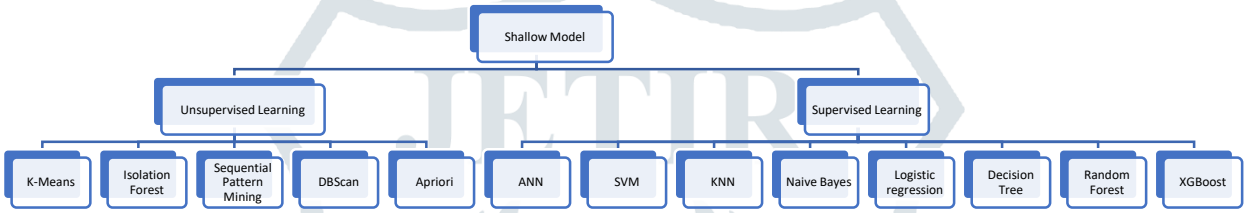


2.1 Shallow Model

The realm of machine learning algorithms, specifically shallow models, can be categorized into two main types: supervised learning and unsupervised learning. In supervised learning scenarios, models generally lack a dependent variable and primarily utilize inherent patterns of data. Various algorithms, discussed in literatures [70-71], can be employed for this purpose.

In supervised learning, models are typically equipped with class labels to authenticate predictions. For instance, Naïve Bayes employs a probability distribution to determine the class label for each data. Based on the training dataset, few decision trees constructed. When it comes to prediction, this tree structure can effectively sort unknown records. Concept of random Forest [13…] adopts a same strategy, but in place of building a single tree, it constructs a set of trees and employs a selection mechanism for record classification. Owing to the collaborative decision-making process, random forests often achieve heightened accuracy of classification.

A SVM (support vector machine) [14…] operates a decision line derived by dataset provided, akin to binary classification. Additionally, SVMs possess the capability to transform data through the application of the kernel trick, enabling them to classify non-linear datasets proficiently.



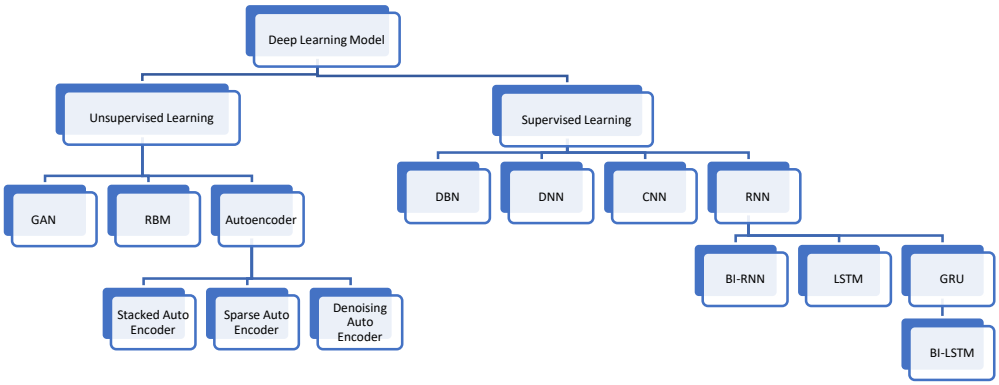
2.2 Deep Learning Model

Deep learning models offer a distinctive approach to classifying and clustering algorithms, diverging significantly from traditional machine learning models. These models are often referred to as "black box models" because they lack a fixed algorithm for prediction. Instead, they scrutinize data, discern patterns, and leverage these patterns for predictive purposes.

Deep learning models employ artificial neural networks constructed from numerous perceptrons. During the initial phases of model training, these perceptions establish connections in a randomized manner. As they analyze the data and undergo training over a specified period, these perceptions acquire values, commonly known as weights, that are better suited for classifying the provided dataset.

There exist various iterations of deep learning models tailored to specific tasks. Convolutional neural networks, for instance, are employed in the classification of image data and have even found application in categorizing cybersecurity datasets by converting the data into an image-like format. On the other hand, Recurrent Neural Networks (RNNs) are apt for classifying data with a temporal dimension. Enhanced versions of RNNs, including LSTM (long short-term memory) and Bi-LSTM, have further improved their performance.

Deep learning's unsupervised learning encompasses both autoencoders and generative adversarial networks. Autoencoders primarily employ dimensionality reduction, transforming information into a compressed representation before subsequent processing. This technique facilitates meaningful information compression, consequently enhancing predictive accuracy.

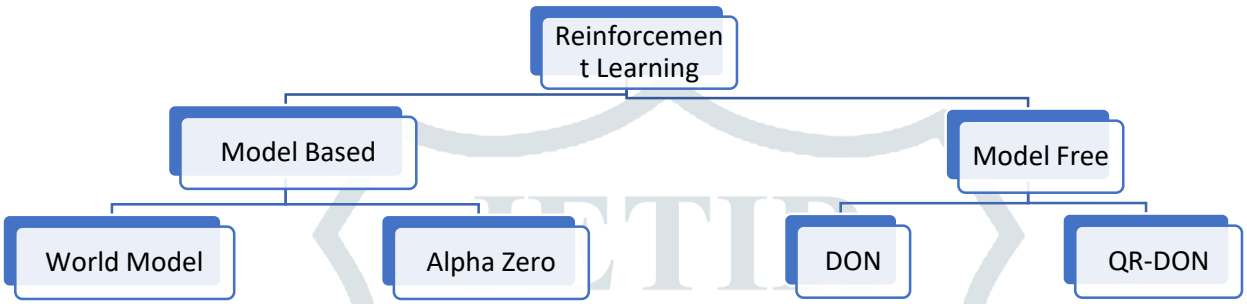


2.3 Reinforcement learning

Reinforcement learning offers a distinct approach to model training, enabling the differentiation between long-term and short-term goals. In this paradigm, agents engage with their environment, receiving rewards or penalties based on their actions. These variable rewards guide the model's improvement over time. One prominent example of this is Deep Q Networks (DQN) [15…], where deep learning is used to establish the mapping between states and actions, reducing the need for a large Q-learning table (TQL).

A derivative of DQN, known as QR-DQN [16…], employs quantile regression to model potential distributions instead of providing a mean distribution. This distinction can be likened to the difference between decision trees and random forests, as discussed earlier. In this section, we delve into various methodologies applicable to machine learning and their relevance to cybersecurity.

Conversely, traditional machine learning models are often referred to as "shallow models" in the context of intrusion detection systems (IDS). Many of these techniques have undergone extensive research and are well-established. They primarily concentrate on functions beyond intrusion detection, encompassing activities such as tagging, efficient attack detection, and the optimal management of available and processed data.



3. Challenges

Numerous research challenges and hindrances exist within the realm of applying machine learning to cybersecurity. These issues must be tackled to derive valuable insights from pertinent data, enabling informed and data-driven cybersecurity choices. Machine learning methods demand substantial computing power and extensive datasets for model training. While employing multiple GPUs can help, this approach is neither energy-efficient nor economical. Furthermore, it's important to note that machine learning techniques are not inherently tailored for cybercrime detection.

3.1 Current Challenges

Traditional machine learning techniques have historically not prioritized cybersecurity. There's a pressing need for robust and potent machine learning methodologies expressly crafted to handle security threats and adversarial inputs. It's vital to recognize that a single machine learning model cannot proficiently detect diverse security attacks. Instead, a tailored machine learning model should be designed for each specific cyberattack type.

Another formidable challenge lies in early-stage attack prevention. Machine learning techniques should possess the capability to swiftly identify real-time and zero-day attacks within a brief timeframe.

Machine learning models have demonstrated their utility in decision-making contexts, such as terrorism detection or medical diagnoses. However, in these instances, blind reliance on predictions could lead to catastrophic consequences. Hence, when employing machine learning techniques in situations of critical importance, such as self-driving cars, cybersecurity, or surgical robotics, it becomes imperative to prioritize high-level correctness guarantees over mere speed and accuracy [17-18…].

Trusted machine learning encompasses the secure application of machine learning techniques in the realm of cyberspace. The reliability of a classifier can be assessed through two avenues: (1) trusting the prediction, which involves evaluating whether users have confidence in a specific prediction model to guide a particular action, and (2) trusting the model itself, which pertains to whether users will have faith in the model when deployed as a tool in a rational manner.

In a study referenced as [19…], researchers delved into the dataset shift problem, wherein a model was trained and tested using dissimilar datasets. They also proposed strategies to mitigate dataset drift, such as eliminating leaked data or modifying the training dataset. These methods assist in determining the necessary steps to transform an untrusted model into a trusted one. It is worth noting that classical linear or shallow learning approaches tend to offer greater reliability, albeit at the cost of slower or less precise performance.

3.2 Future Challenges

Deep learning remains intricate and opaque, despite the ongoing advancements in theory. The advent of mobile phones and the Global Positioning System has opened up new possibilities for forensic science and epidemic control to ascertain the whereabouts of specific moving objects. However, maintaining the trustworthiness of a particular object's location is a formidable task due to potential errors or data distortions inherent in mobile devices. Chenyun [20…] introduced an approach to evaluate the similarity of location-related information gathered from multiple sources about a specific object. The reliability of position data derived from the trajectories of moving objects is inherently uncertain, primarily due to the objects' positional changes and network delays [21…]. In [22…], the authors proposed the use of a trust ontology approach to facilitate trustful interactions between service providers and consumers within an online web system.

Credibility also plays a vital role in natural language processing (NLP), especially in text classification for critical missions where message interpretation holds significant importance. Incorporating credibility considerations into text analysis, both in practical and semantic contexts, is essential to achieve the most accurate credibility detection results [23…]. Others have put forth a metric model for assessing software trustworthiness [24…]. Machine learning techniques find applications in the energy sector, where energy-efficient strategies have been suggested to curtail power consumption in data centers and businesses [25…]. This involves dynamically shutting down idle machines to reduce overall energy usage. Ensuring the reliability of the prediction model for deciding which machines to power down is of paramount importance.

In the realm of cybersecurity, the sensitivity of alarm detection is a critical concern, as it can lead to a higher false alarm rate, commonly referred to as alarm fatigue. A heightened false alarm rate has adverse implications for security personnel and can result in missed critical alarms or delayed responses. Addressing this issue is a challenging research endeavor in the field of cybersecurity [26…], [27…].

4. Major Challenge and Issues

4.1. Availability of Dataset

In the field of cybersecurity, much like in machine learning, the availability of source datasets is of paramount importance. However, a significant challenge arises from the fact that many publicly accessible datasets tend to be outdated and may not offer sufficient insights into the uncharted behavior patterns of diverse cyberattacks. Even though current data can be transformed into knowledge through a series of processing steps, there remains a gap in our comprehension of the characteristics of recent attacks and their recurring patterns. Consequently, the application of additional processing or machine learning techniques may yield suboptimal accuracy in the final decision-making process.

A fundamental hurdle in leveraging machine learning approaches for cybersecurity lies in the scarcity of up-to-date, domain-specific datasets, especially for tasks such as attack prediction or intrusion detection. Much of the information pertaining to data and cyberattacks tends to be repetitive, and machine learning models tend to perform more effectively when trained on larger datasets, which is often not the case with currently available datasets. Conversely, publicly accessible datasets are typically subject to strict anonymization and are beset by various limitations, primarily the fact that they do not accurately represent real-world and recent cyberattacks. Given these challenges, it remains challenging to distinguish between simulated benchmark datasets and the latest, real-world data.

4.2. Standard Dataset

Cybersecurity datasets often exhibit several challenges, including imbalances, noise, incompleteness, irrelevance, and inconsistencies in the instances of security breaches. These dataset issues have a detrimental impact on the quality of the learning process and the performance of machine learning-based models, as highlighted in references [28,29…]. To establish a data-driven cybersecurity solution, it is imperative to resolve these data-related problems before applying machine learning techniques.

One crucial step in this process is the establishment of benchmark and standard datasets containing extensive data for both training and testing, ensuring a balanced representation of attack categories. Data for security systems is sourced from various channels, encompassing social media and conventional sources like web and database access.

An essential aspect is comprehending these cybersecurity data challenges and addressing them effectively through existing or novel algorithms to accomplish tasks such as malware and intrusion detection, among others.

Feature engineering techniques, as mentioned in reference [30…], play a pivotal role in resolving these issues. These techniques involve the analysis and removal of redundant features, thereby reducing data dimensionality and complexity. Handling data imbalance is another critical aspect, which can be approached through methods like hybrid models, as reported in [31…], or through the generation of synthetic data, as mentioned in [162,163]. Additionally, addressing concerns related to data leakage is essential.

Furthermore, the sheer volume and diversity of data sources collected from various origins pose a significant challenge for machine learning models in the field of cybersecurity. It's worth noting that many datasets representing recent attacks are not publicly available due to privacy and security concerns.

4.3 Standard Metrics

In the work by the authors in reference [32……..], they introduced several evaluation metrics for assessing the classifier's performance. Nevertheless, it's noteworthy that many researchers have employed distinct parameters to appraise classification models, often overlooking the complementary aspects, even when working with the same dataset. There is a pressing necessity to establish a consensus on a standardized set of metrics for comparing models, which would pave the way for more effective enhancements in this field.

4.4. Hybrid Learning

Signature-based intrusion detection methods stand out as the most prevalent and firmly established approaches within the cybersecurity landscape [34,35…]. Nevertheless, these algorithms may falter in detecting novel attacks or incidents due to missing features, significant feature constraints, or limited profiling capabilities. To address these limitations, anomaly-based techniques, or hybrid approaches that merge both anomaly-based and signature-based detection methods, can be employed effectively.

For a more focused understanding within specific problem domains like intrusion detection, malware analysis, or phishing detection, harnessing a hybrid learning approach that combines diverse machine learning techniques proves invaluable. By amalgamating deep learning, statistical analysis, and traditional machine learning methodologies, one can make informed decisions when devising cybersecurity solutions.

4.5 Detection and Time Complexity of Techniques

The existing literature has paid limited attention to real-time attack environments, which is a noteworthy gap. When addressing such environments, it becomes imperative to evaluate both the detection rate of attacks and the algorithm's time complexity. Given that cybercriminals continually devise new attack strategies to exploit network vulnerabilities, the efficacy of attack detection holds great significance. In cases where the system produces false positives, security analysts are compelled to invest valuable time investigating non-malicious activities, which can undermine their confidence in the system if such occurrences become frequent. It's also crucial to take into account the computational complexity of various machine learning models, as demonstrated in Table 6. Furthermore, future research might explore enhancing detection speed and reducing computational costs by leveraging advanced hardware in a distributed approach.

4.6. Feature Engineering

The efficacy and performance of machine learning-based security models have come under scrutiny due to the immense volume of network traffic data and the multitude of smaller operational intricacies. To address the high dimensionality of this data, several techniques, such as principal component analysis (PCA), singular value decomposition (SVD), and linear discriminant analysis (LDA), have been employed. Establishing contextual connections between suspicious activities and low-level information within datasets can prove beneficial. These contextual data can be subjected to processing through an ontology or taxonomy to facilitate further investigation. Consequently, another challenging aspect in the domain of machine learning for cybersecurity pertains to the efficient selection of optimal features or the extraction of significant characteristics. This process should take into account both machine-readable features and contextual attributes to devise effective cybersecurity solutions.

4.7. Leakage

Data leakage, often referred to as "leakage," occurs when the training dataset includes relevant data that is not readily available or significantly diverges when models are utilized for predictions [37…]. This typically leads to overly optimistic predictions during the model development phase, followed by disappointing results when the prediction model is applied and tested on new data. In a notable research work [38…], this problem is termed "leaks from the

future," recognized as one of the "top 10 data mining mistakes," and is addressed through the recommendation of employing exploratory data analysis (EDA) to identify and eliminate potential sources of leakage.

EDA serves the purpose of enhancing dataset integrity, ultimately improving the accuracy of machine learning models when making predictions on unfamiliar data. Recent research [39…] emphasizes the detection and exploitation of leaks as a crucial factor in achieving success in data mining competitions, highlighting that it can also be a determinant of failure in data mining applications. Additionally, another study [40…] discusses the inclusion of indicative features that predict the target variable, often introduced at a later stage in the data collection process.

To mitigate the risk of leakage, researchers have proposed a two-stage approach [41…], which involves tagging each observation with a legitimacy marker during data collection and subsequently ensuring a clear separation between the learning and prediction phases. This approach yielded significant benefits, achieving a maximum accuracy of 91.2% with Naive Bayes, 87.5% using k-NN, and 94.2% with a centroid-based approach across different categories. In cases where machine learning scientists lack control over the data collection process, EDA remains a valuable technique for detecting and addressing potential leaks [42…], holding promise for future research endeavors.

4.8. Homomorphic Encryption

Homomorphic encryption represents a significant milestone in the field of cryptography, offering the capability to allow an untrusted third party to process data without revealing any sensitive information, thereby granting access to confidential data securely. In this encryption paradigm, neither the end-user nor an unauthorized remote server gains access to the decryption key, ensuring that the data remains within the designated domain. Its versatility extends across various domains, encompassing applications in cloud computing, financial transactions, and defense against potential threats from quantum computing technologies[43…].

Homomorphic encryption can be applied in two distinct ways: partial and full encryption. Fully Homomorphic Encryption (FHE) plays a pivotal role in enabling machine learning processes without compromising data privacy. Machine learning algorithms, whether deep learning or shallow, heavily rely on domain-specific data, often challenging to share publicly. FHE introduces a novel approach to delegate the sharing of sensitive data without exposing the actual meaningful data. However, it is essential to note that FHE's primary limitation lies in its restriction to integer-based operations. Therefore, ongoing research is focused on developing matrix-based schemes for FHE, with recent work demonstrating the efficacy of employing lowest degree polynomial approximation functions, such as Chebyshev, in conjunction with continuous functions like the sigmoid function. This innovation has paved the way for a new encryption method over FHE, particularly suitable for homogeneous networks[44-45…].

Federated learning, in tandem with FHE, has revolutionized the learning processes, particularly in scenarios involving extensive image data with sample expansions. This synergy has expanded the scope of FHE applications across various domains, including the highly confidential realm of medical and health information. Access to health data through FHE has opened up numerous possibilities for leveraging machine learning in the context of medical images and data from the Internet of Medical Things (IoMT). Notably, the integration of FHE with chaotic mapping, though successful in ensuring data transmission, raised concerns about computational privacy. Subsequent advancements in 2021 combined FHE with secret sharing and edge computing, enabling distributed mathematical operations without data leakage[46-47…].

Furthermore, the emergence of CryptoRNN, a recurrent neural network, has introduced a novel approach focusing on blockchain technology's privacy protection. Within cloud environments, the integration of FHE has become increasingly prevalent due to its adaptability in accessing domain data and harnessing substantial computing power. Machine learning as a service platform (MLaaS) further augments the utility of FHE in safeguarding confidential data by offering a wide array of machine learning algorithms[48-49…].

Exploring the application of homomorphic encryption in wireless sensor networks (WSNs), researchers have assessed its performance using the NS-2 network simulation tool. In such environments, where conditions remain consistent for each experimental agent, FHE outperformed alternative decryption methods like DAA, which decrypt data hop-by-hop, achieving a time complexity of O(n). Overall, the incorporation of homomorphic encryption serves to enhance global data flow, expand practical machine learning applications, and bolster cybersecurity efforts on a larger scale [51-52…].

4.9. Quantum Computing

In the early stages of quantum computing development, it became apparent that these emerging systems held the potential to compromise the security provided by asymmetric encryption techniques [53…]. Asymmetric key encryption relies on the generation of public and private keys through the factorization of two exceedingly large prime numbers. While factoring small primes is feasible, decrypting keys of significant size could take thousands of years, ensuring the security of our data. However, Shore's algorithm [54…], which offers an alternative for factorization, is also notably slow. Quantum computing, leveraging its superposition principle, can rapidly derive factors in a fraction of the time it would take for classical binary computing systems. Consequently, widely-used encryption algorithms like RSA, DES, elliptic curve algorithms such as ECDSA, and digital signature algorithms are rendered insecure by the speed of quantum computing [55…]. For instance, researchers have pointed out that cracking a 56-bit DES encryption using Grover's algorithm on a quantum computer would require just 185 searches to identify the key [56…].

On the other hand, symmetric key algorithms like AES remain resistant to quantum computing attacks. Scientists are exploring various avenues, including both quantum and mathematical techniques, to overcome these limitations. One notable example is the BB84 protocol, a quantum key distribution method [57…]. Additionally, mathematical approaches like lattice-based cryptography are under investigation [58…]. These efforts aim to enhance encryption methods that can withstand quantum computing advancements.

While quantum computing poses a threat to asymmetric encryption, it also offers opportunities for accelerating machine learning when employed as subroutines [59…]. This potential enhancement can significantly reduce prediction times, especially for algorithms such as Support Vector Machines (SVM), which may require extensive time for implementing kernel transformations to derive hyperplanes. Quantum computing can also be integrated into deep learning configurations, albeit with some challenges due to the linear dynamics inherent in quantum neural networks [60…].

4.10 Concept of Adversarial Inputs to Models

The concept of adversarial inputs to machine learning models poses several challenges, as discussed by the authors in [61…]. In military applications, the need for swift decision-making is paramount. Adversaries can manipulate messages by introducing hostile text sequences, potentially altering the message's entire meaning and leading to disastrous consequences [62…]. Training machine learning models in an adversarial environment is a crucial strategy for enhancing their resilience to such hostile inputs.

One proposed defense mechanism in this regard is DeepCloak, designed to identify and eliminate unnecessary features within deep neural network (DNN) models. DeepCloak's function is to curtail an attacker's ability to generate adversarial samples, thereby bolstering the model's robustness [63…]. However, it's worth noting that the assumption that test data comes from the same distribution as the training data is often violated. For instance, differences in the cameras used to capture images during training and testing can adversely affect model performance.

The work of Tony et al. in [64-65…] and [66…] sheds light on various adversarial attacks that can deceive the learning process of machine learning models. Additionally, Ibitoye et al. in [67…] have proposed a novel model for identifying the risk of adversarial attacks in network security, accompanied by an evaluation of different adversarial attacks on machine learning models used in network security scenarios.

In the realm of cybersecurity, there is a pressing need for deep learning models that exhibit resilience to noise and adversarial examples, although achieving this remains a challenging endeavor.

4.11 Adversarial Attacks and Defences

On the contrary, when a cyber attacker manipulates the attack pattern to influence the data as it is being distributed in order to deceive the trained model, this type of attack is referred to as an evasion attack [68…]. There exists a spectrum of adversarial attacks, including but not limited to the Fast Gradient Signal Method (FGSM), Multistage Bit Coordinate Ascent (BCAk), Multistage Bit Gradient Ascent (BGAk), Generative Adversarial Networks (GAN), and the Carlini & Wagner Attack (C&W) [68…].

In response to these adversarial threats, various defensive strategies have been proposed in the research literature. These strategies aim to thwart adversarial attacks and maintain the integrity of machine learning models. Among these strategies are adversarial training [69…], defense distillation [70…], compression specificity [71…], and the Magnet approach [72…].

Adversarial training involves the incorporation of adversarial examples during the model training process. While this approach is relatively straightforward to implement, it necessitates some level of training behind the model. The benefit is that attacks encountered during testing are as valuable as those experienced during training.

On the other hand, defense distillation requires retraining the model but is highly efficient for most datasets. It involves the distillation of neural networks to train new models with improved resilience to adversarial attacks.

Feature compression has proven effective in combatting various adversarial attacks, especially in image databases such as ImageNet and MNIST. This method involves compressing the data using various compression techniques, typically pixel-based methods. If the predictions of the original and compressed samples exhibit significant disparities, the compressed sample is flagged as a counterexample. Notably, this approach does not necessitate retraining the model but instead employs an autoencoder to detect counter-patterns [71…].

4.12 Growing Attacks

As the field of cybersecurity continues to evolve, cyberattacks are also evolving at a rapid pace. Leveraging machine learning (ML) to counter these emerging threats poses two distinct challenges. Firstly, ML models are employed to detect activities that have not been previously encountered [73…]. Secondly, new cyberattacks frequently differ in their technical characteristics from older ones. Typically, ML models are trained using historical features within the dataset, but new attacks may exhibit different sets of features. Modern cyberattacks have the ability to evade classifiers, potentially leading to false alarms or reduced detection rates.

4.13 Confidentiality and Protection

As per concerns of security and user privacy, have increased with the collection of data from structured and unstructured sources. This leads to Big-data and its protection issues for safety [74…]. Safe data protection against hostile attacks and tampering by unauthorized users is essential. Normal users should also have permission to access the data.

5. Other Unique Challenges

5.1 Much higher requirements for accuracy.

if you're processing an image and the system mistakes a dog for a cat, it might be annoying, but it probably won't be life-or-death. If a machine learning system mistakes a fraudulent data packet for a legitimate one, leading to an attack on a hospital and its equipment, the impact of miscategorization can be severe.

Organizations see large volumes of data packets pass through firewalls every day. If Machine Learning model miscategorized only 0.1% data, then we can mistakenly block a huge amount of normal traffic that would have a serious impact on the business. Understandably, in the early days of machine learning, some organizations were concerned that the models would not be as accurate as human security researchers. It takes time and also requires a huge amount of data to actually train a machine learning model to the same level of accuracy as a truly skilled human. However, people do not scale and are among the rarest resources in IT today. We rely on ML to effectively augment cybersecurity solutions. ML can also help us detect unknown attacks that are hard for humans to detect because ML can create baseline behaviors and detect any abnormalities that deviate from them.

5.2 Access to large amounts of training data, especially labeled data.

Machine learning requires large amounts of data to make models and predictions more accurate. Malware sample acquisition is much more difficult than image processing and NLP data acquisition. There is not enough data on attacks and much data on security risks is sensitive and unavailable for privacy reasons.

5.3 Dynamic Nature

Basic truth. Unlike images, the ground truth in cybersecurity may not always be available or fixed. The cybersecurity environment is dynamic and constantly changing. No malware database can claim to cover all the malware in the world, and more malware is being generated every moment. What is the ground truth that we should compare ourselves to in order to judge our accuracy?

6. Conclusions

Cybersecurity stands as a global concern, prompting ongoing enhancements in security measures to detect and combat cyber threats. The conventional security systems utilized in the past have become inadequate, lacking the efficacy to identify concealed and polymorphic attacks. Machine learning techniques have emerged as a pivotal component in various cybersecurity applications. Our examination reveals a substantial surge in interest surrounding

the intersection of machine learning and cybersecurity within both academic and industrial domains, particularly in the past decade. This surge has resulted in a marked increase in publications. In this paper, we endeavor to bridge the gap between machine learning techniques and the myriad threats facing computer networks and mobile communications by conducting a comprehensive study that explores the interplay between these two domains.

Our study encompasses a literature review of machine learning techniques for intrusion detection, spam detection, and malware detection in computer networks and mobile devices over the preceding decade. We offer a concise overview of the application of machine learning models in the realm of cybersecurity, with a specific focus on developments within the last decade. Each type of cyber threat presents unique characteristics that challenge even the most advanced machine learning models when addressing such attacks. Consequently, prescribing a singular recommendation for all attacks based on a single model proves unfeasible. Multiple criteria, such as detection speed, time complexity, classification time for identifying novel and zero-day attacks, and model accuracy, should be weighed when selecting a particular model for cyber attack detection.

In our exploration, we elucidate the fundamentals of cybersecurity, including the categorization of cyber attacks on both mobile devices and computer networks. Recognizing the pivotal role of machine learning, we provide introductory explanations of the basics, types, and key techniques of machine learning to facilitate comprehension for newcomers to the field. To our knowledge, there exists a scarcity of literature that delves into the application of machine learning techniques within the domain of cybersecurity concerning mobile devices and computer networks.

We present a visual overview of cyber attacks and the spectrum of machine learning techniques available for countering these cybercrimes. Additionally, we evaluate select popular machine learning tools and propose evaluation criteria for assessing the performance of any classifier. Datasets hold paramount importance for training and testing machine learning models, and we furnish descriptions of the most frequently utilized security databases. It is important to note that no single, comprehensive database exists for each threat domain.

Machine learning techniques were not originally conceived with cybersecurity in mind, making them susceptible to fuzziness, which can lead to misleading inputs. Reliable machine learning represents a facet of applying machine learning techniques in cyberspace that offers a degree of assurance regarding model speed and accuracy. We also provide a concise summary of key challenges in the application of machine learning techniques in the context of cybersecurity and supply an extensive bibliography to guide further exploration in this field. These challenges warrant substantial attention and exploration in future research endeavors.

6. References

[1.] O'Connell, M.E. Cyber security without cyber war. J. Confl. Secure. Law 2012, 17,187–209. [CrossRef]
[2.] Tolle, K.M.; Tansley, D.S.W.; Hey, A.J. The fourth paradigm: Data-intensive scientific discovery [point of view]. Proc. IEEE 2011, 99, 1334–1337. [CrossRef]
[3.] Benioff, M. Data, data everywhere: A special report on managing information (pp. 21–55). The Economist, 27 February 2010.
[4.] Cost of Cyber Attacks vs. Cost of Cybersecurity in 2021|Sumo Logic. Available online: https://www.sumologic.com/blog/costof-cyber-attacks-vs-cost-of-cyber-security-in-2021/ (accessed on 10 May 2022).
[5.] Anwar, S.; Mohamad Zain, J.; Zolkipli, M.F.; Inayat, Z.; Khan, S.; Anthony, B.; Chang, V. From intrusion detection to an intrusion response system: fundamentals, requirements, and future directions. Algorithms 2017, 10, 39. [CrossRef]
[6.] Tapiador, J.E.; Orfila, A.; Ribagorda, A.; Ramos, B. Key-recovery attacks on KIDS, a keyed anomaly detection system. IEEE Trans.
[7.] Saxe, J.; Sanders, H. Malware Data Science: Attack Detection and Attribution; No Starch Press: San Francisco, CA, USA, 2018..
[8.] Han, J.; Kamber, M.; Pei, J. Data mining concepts and techniques third edition. Morgan Kaufmann Ser. Data Manag. Syst. 2011, 5, 83–124.
[9.] Witten, I.H.; Frank, E.; Hall, M.A.; Pal, C.J. Practical machine learning tools and techniques. Morgan Kaufmann 2005, 2, 578.
[10.] Dua, S.; Du, X. Data Mining and Machine Learning in Cybersecurity; CRC Press: Boca Raton, FL, USA, 2016.

[11.] Ester, M.; Kriegel, H.P.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the KDD-94, Oregon, Portland, 2–4 August 1996; Volume 96, pp. 226–231.

[12.] Inokuchi, A.; Washio, T.; Motoda, H. An apriori-based algorithm for mining frequent substructures from graph data. In Proceedings of the European Conference on Principles of Data Mining and Knowledge Discovery, Lyon, France, 13–16 September 2000; Springer: Berlin/Heidelberg, Germany, 2000; pp. 13–23.

[13.] Breiman, L. Random forests. Mach. Learn. 2001, 45, 5–32. [CrossRef]

[14.] Cortes, C.; Vapnik, V. Support-vector networks. Mach. Learn. 1995, 20, 273–297. [CrossRef]

[15.] Mnih, V.; Kavukcuoglu, K.; Silver, D.; Graves, A.; Antonoglou, I.; Wierstra, D.; Riedmiller, M. Playing atari with deep reinforcement learning. arXiv 2013, arXiv:1312.5602.

[16.] Dabney, W.; Rowland, M.; Bellemare, M.; Munos, R. Distributional reinforcement learning with quantile regression. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; Volume 32.

[17.] M. T. Ribeiro, S. Singh, and C. Guestrin, ‘‘‘Why should I trustyou?’ Explaining the predictions of any classifier,’’ in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Aug. 2016,pp. 1135–1144.

[18.] S. Ghosh, P. Lincoln, A. Tiwari, and X. Zhu, ‘‘Trusted machine learning:Model repair and data repair for probabilistic models,’’ in Proc. Workshops 31st AAAI Conf. Artif. Intell., 2017.

[19.] J. Quionero-Candela, M. Sugiyama, A. Schwaighofer, andN. D. Lawrence, Dataset Shift in Machine Learning. Cambridge, MA, USA: MIT Press, 2009.

[20.] C. Dai, H.-S. Lim, E. Bertino, and Y.-S. Moon, ‘‘Assessing the trust-worthiness of location data based on provenance,’’ in Proc. 17th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst. (GIS), 2009,pp. 276–285

[21.] G. Trajcevski, O. Wolfson, K. Hinrichs, and S. Chamberlain, ‘‘Managing Uncertainty in moving objects databases,’’ ACM Trans. Database Syst., vol. 29, no. 3, pp. 463–507, Sep. 2004.

[22.] M. Zhu and Z. Jin, ‘‘A trust measurement mechanism for service agents,’’in Proc. IEEE/WIC/ACM Int. Joint Conf. Web Intell. Intell. Agent Technol., Sep. 2009, pp. 375–382.

[23.] Q. Su, C.-R. Huang, and H. K.-Y. Chen, ‘‘Evidentiality for text trustwor-thiness detection,’’ in Proc. Workshop NLP Linguistics, Finding Common Ground, Assoc. Compute. Linguistics, 2010, pp. 10–17.

[24.] H. Tao and Y. Chen, ‘‘A metric model for trustworthiness of softwares,’’ inProc. IEEE/WIC/ACM Int. Joint Conf. Web Intell. Intell. Agent Technol., Sep. 2009, pp. 69–72.

[25.] J. L. Barrel, Ì. Goiri, R. Nou, F. Julià, J. Guitart, R. Gavaldà, and J. Torres,‘‘Towards energy-aware scheduling in data centers using machine learning,’’ in Proc. 1st Int. Conf. Energy-Efficient Comput. Netw. (E-Energy), 2010, pp. 215–224.

[26.] X. Wang, Y. Gao, J. Lin, H. Rangwala, and R. Mittu, ‘‘A machine learning approach to false alarm detection for critical arrhythmia alarms,’’ in Proc. IEEE 14th Int. Conf. Mach. Learn. Appl. (ICMLA), Dec. 2015, pp. 202–207.

[27.] L. M. Eerikainen, J. Vanschoren, M. J. Rooijakkers, R. Vullings, andR. M. Aarts, ‘‘Decreasing the false alarm rate of arrhythmias in intensivecare using a machine learning approach,’’ in Proc. Comput. Cardiol. Conf.(CinC), Sep. 2015, pp. 293–296.

[28.] Kaelbling, L.P.; Littman, M.L.; Moore, A.W. Reinforcement learning: A survey. J. Artif. Intell. Res. 1996, 4, 237–285. [CrossRef]

[29.] Sarker, I.H.; Colman, A.; Han, J. Recencyminer: Mining recency-based personalized behavior from contextual smartphone data. J. Big Data 2019, 6, 1–21. [CrossRef]

[30.] Ahsan, M.; Gomes, R.; Chowdhury, M.; Nygard, K.E. Enhancing Machine Learning Prediction in Cybersecurity Using Dynamic Feature Selector. J. Cybersecur. Priv. 2021, 1, 199–218. [CrossRef]

[31.] Li, J.; Qu, Y.; Chao, F.; Shum, H.P.; Ho, E.S.; Yang, L. Machine learning algorithms for network intrusion detection. In AI in Cybersecurity; Springer: Berlin/Heidelberg, Germany, 2019; pp. 151–179.

[32.] Massaoudi, M.; Refaat, S.S.; Abu-Rub, H. Intrusion Detection Method Based on SMOTE Transformation for Smart Grid Cybersecurity. In Proceedings of the 2022 3rd International Conference on Smart Grid and Renewable Energy (SGRE), Doha, Qatar, 20–22 March 2022; IEEE: Piscataway, NJ, USA, 2022, pp. 1–6.

[33.] Ahsan, M.; Gomes, R.; Denton, A. Smote implementation on phishing data to enhance cybersecurity. In Proceedings of the 2018 IEEE International Conference on Electro/Information Technology (EIT), Rochester, MI, USA, 3–5 May 2018; IEEE: Piscataway, NJ, USA, 2018, pp. 0531–0536.

[34.] Liao, H.J.; Lin, C.H.R.; Lin, Y.C.; Tung, K.Y. Intrusion detection system: A comprehensive review. J. Netw. Compute. Appl. 2013, 36, 16–24. [CrossRef]

[35.] Tsai, C.W.; Lai, C.F.; Chao, H.C.; Vasilakos, A.V. Big data analytics: A survey. J. Big Data 2015, 2, 1–32.

[36.] Sarker, I.H.; Abushark, Y.B.; Khan, A.I. Context Pca: Predicting context-aware smartphone apps usage based on machine learning techniques. Symmetry 2020, 12, 499. [CrossRef]

[37.] Kaufman, S.; Rosset, S.; Perlich, C.; Stitelman, O. Leakage in data mining: Formulation, detection, and avoidance. ACM Trans. Knowl. Discover. Data TKDD 2012, 6, 1–21. [CrossRef]

[38.] Nisbet, R.; Elder, J.; Miner, G.D. Handbook of Statistical Analysis and Data Mining Applications; Academic Press: Cambridge, MA, USA, 2009.

[39.] Rosset, S.; Perlich, C.; Swirszcz, G.; Melville, P.; Liu, Y. Medical data mining: Insights from winning two competitions. ´ Data Min. Knowl. Discover. 2010, 20, 439–468. [CrossRef]

[40.] Kohavi, R.; Brodley, C.E.; Frasca, B.; Mason, L.; Zheng, Z. KDD-Cup 2000 organizers' report: Peeling the onion. ACM Sigkdd Explor. Newsl. 2000, 2, 86–93. [CrossRef]

[41.] Gupta, I.; Mittal, S.; Tiwari, A.; Agarwal, P.; Singh, A.K. TIDF-DLPM: Term and Inverse Document Frequency based Data Leakage Prevention Model. arXiv 2022, arXiv:2203.05367.

[42.] Stuart, M. Understanding robust and exploratory data analysis. J. R. Stat. Soc. Ser. D1984, 33, 320–321. [CrossRef]

[43.] Kjamilji, A.; Sava¸s, E.; Levi, A. Efficient secure building blocks with application to privacy preserving machine learning algorithms. IEEE Access 2021, 9, 8324–8353. [CrossRef] [

[44.] Aono, Y.; Hayashi, T.; Wang, L.; Moriai, S. Privacy-preserving deep learning via additively homomorphic encryption. IEEE Trans. Inf. Forensics Secur. 2017, 13, 1333–1345.

[45.] Takabi, H.; Hesamifard, E.; Ghasemi, M. Privacy preserving multi-party machine learning with homomorphic encryption. In Proceedings of the 29th Annual Conference on Neural Information Processing Systems (NIPS), Barcelona, Spain, 5–10 December 2016.

[46.] Fang, H.; Qian, Q. Privacy preserving machine learning with homomorphic encryption and federated learning. Future Internet 2021, 13, 94. [CrossRef]

[47.] Yang, Y.; Xiao, X.; Cai, X.; Zhang, W. A secure and high visual-quality framework for medical images by contrast-enhancement reversible data hiding and homomorphic encryption. IEEE Access 2019, 7, 96900–96911. [CrossRef]

[48.] Salim, M.M.; Kim, I.; Doniyor, U.; Lee, C.; Park, J.H. Homomorphic Encryption Based Privacy-Preservation for IoMT. Appl. Sci. 2021, 11, 8757. [CrossRef]

[49.] Bakshi, M.; Last, M. Cryptornn-privacy-preserving recurrent neural networks using homomorphic encryption. In International Symposium on Cyber Security Cryptography and Machine Learning; Springer: Berlin/Heidelberg, Germany, 2020; pp. 245–253.

[50.] Guan, Z.; Bian, L.; Shang, T.; Liu, J. When machine learning meets security issues: A survey. In Proceedings of the 2018 impact of quantum computing on present cryptography. arXiv 2018, arXiv:1804.00200.

[51.] Li, X.; Chen, D.; Li, C.; Wang, L. Secure data aggregation with fully homomorphic encryption in large-scale wireless sensor networks. Sensors 2015, 15, 15952–15973. [CrossRef] [PubMed]

[52.] Latif, S.; Dola, F.F.; Afsar, M.; Esha, I.J.; Nandi, D. Investigation of Machine Learning Algorithms for Network Intrusion Detection. Int. J. Inf. Eng. Electron. Bus. 2022, 14, 1–22.

[53.] Mavroeidis, V.; Vishi, K.; Zych, M.D.; Jøsang, A. Rosset , S.; Perlich, C.; Swirszcz, G.; Melville, P.; Liu, Y. Medical data mining: Insights from winning two competitions. ´ Data Min. Knowl. Discover. 2010, 20, 439–468. [CrossRef]

[54.] Shor, P.W. Algorithms for quantum computation: Discrete logarithms and factoring. In Proceedings of the 35th Annual Symposium on Foundations of Computer Science, Santa Fe, NM, USA, 20–22 November 1994; IEEE: Piscataway, NJ, USA, 1994, pp. 124–134.

[55.] Bone, S.; Castro, M. A Brief History of Quantum Computing; Imperial College in London: London, UK, 1997. Available online: http://www.doc.ic.ac.uk/~{}nd/surprise_97/journal/vol4/spb3 (accessed on 10 May 2022).

[56.] Grover, L.K. A fast quantum mechanical algorithm for database search. In Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing, Philadelphia, PA, USA, 22–24 May 1996; pp. 212–219.

[57.] Cerf, N.J.; Levy, M.; Van Assche, G. Quantum distribution of Gaussian keys using squeezed states. Phys. Rev. A 2001, 63, 052311. [CrossRef]

[58.] Ding, J.; Yang, B.Y. Multivariate public key cryptography. In Post-Quantum Cryptography; Springer: Berlin/Heidelberg, Germany, 2009; pp. 193–241.

[59.] Hassija, V.; Chamola, V.; Goyal, A.; Kanhere, S.S.; Guizani, N. Forthcoming applications of quantum computing: Peeking into the future. IET Quantum Commun. 2020, 1, 35–41. [CrossRef]

[60.] Schuld, M.; Sinayskiy, I.; Petruccione, F. The quest for a quantum neural network. Quantum Inf. Process. 2014, 13, 2567–2586. [CrossRef]

[61.] S. Huang, E.-H. Liu, Z.-W. Hui, S.-Q. Tang, and S.-J. Zhang, ‘‘Challenges of testing machine learning applications,’’ Int. J. Performability Eng., vol. 14, no. 6, pp. 1–8, 2018.

[62.] J. Gao, J. Lanchantin, M. L. Soffa, and Y. Qi, ‘‘Black-box generation of adversarial text sequences to evade deep learning classifiers,’’ in Proc. IEEE Secur. Privacy Workshops (SPW), May 2018, pp. 50–56.

[63.] J. Gao, B. Wang, Z. Lin, W. Xu, and Y. Qi, ‘‘DeepCloak: Masking deep neural network models for robustness against adversarial samples,’’ 2017, arXiv:1702.06763. [Online]. Available: http://arxiv.org/abs/1702.06763

[64.] I. Goodfellow, P. McDaniel, and N. Papernot, ‘‘Making machine learning robust against adversarial inputs,’’ Commun. ACM, vol. 61, no. 7, pp. 56–66, Jun. 2018.

[65.] T. Thomas, A. P. Vijayaraghavan, and S. Emmanuel, ‘‘Adversarial machine learning in cybersecurity,’’ in Machine Learning Approaches in Cyber Security Analytics. Singapore: Springer, 2020, pp. 185–200.

[66.] P. Dasgupta and J. B. Collins, ‘‘A survey of game theoretic approaches for adversarial machine learning in cybersecurity tasks,’’ 2019, arXiv:1912.02258. [Online]. Available: http://arxiv.org/abs/1912.02258

[67.] O. Ibitoye, R. Abou-Khamis, A. Matrawy, and M. O. Shafiq,‘‘The threat of adversarial attacks on machine learning in network security—A survey,’’ 2019, arXiv:1911.02621. [Online]. Available:http://arxiv.org/abs/1911.02621

[68.] F. Zhang, P. P. K. Chan, B. Biggio, D. S. Yeung, and F. Roli, ‘‘Adversarial feature selection against evasion attacks,’’ IEEE Trans. Cybern., vol. 46, no. 3, pp. 766–777, Mar. 2016.

[69.] I. J. Goodfellow, J. Shlens, and C. Szegedy, ‘‘Explaining and harnessing adversarial examples,’’ 2014, arXiv:1412.6572. [Online]. Available: http://arxiv.org/abs/1412.6572

[70.] N. Papernot, P. McDaniel, X. Wu, S. Jha, and A. Swami, ‘‘Distillation asa defense to adversarial perturbations against deep neural networks,’’ in Proc. IEEE Symp. Secur. Privacy (SP), May 2016, pp. 582–597

[71.] W. Xu, D. Evans, and Y. Qi, ‘‘Feature squeezing: Detecting adversarial examples in deep neural networks,’’ 2017, arXiv:1704.01155. [Online]. Available: http://arxiv.org/abs/1704.01155

[72.] D. Meng and H. Chen, ‘‘MagNet: A two-pronged defense against adversarial examples,’’ in Proc. Conf. Comput. Commun. Secur. (ACM SIGSAC), Oct. 2017, pp. 135–147.

[73.] R. Sommer and V. Paxson, ‘‘Outside the closed world: On using machine learning for network intrusion detection,’’ in Proc. IEEE Symp. Secur. Privacy, May 2010, pp. 305–316.

[74.] A. A. Cárdenas, P. K. Manadhata, and S. P. Rajan, ‘‘Big data analytics for security,’’ IEEE Security Privacy, vol. 11, no. 6, pp. 74–76, Nov./Dec. 2013.