IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

COMPARATIVE ANALYSIS ON RESEARCH CHALLENGES, APPLICATIONS AND METHODOLOGIES FOR CYBER SECURITY THROUGH AI BASED ALGORITHMS

Ms. Sruthi Mol P[1], Dr. Sathish Kumar N[2]

Assistant Professor[1], Professor[1]

Department of Information Technology[1], Department of Electronics and Communication Engineering[1]

KGiSL Institute of Technology[1], Sri Ramkrishna Engineering College[2]

Coimbatore, Tamil Nadu, India

Abstract: Cyber analysts find it more and more challenging to efficiently monitor the volume, velocity, and diversity of data produced as the number of Internet-connected equipment grows. Cyber security tactics that rely on signatures are unlikely to deliver the performance needed to find new attack vectors. The development of advanced attack techniques that can evade detection by present security systems is also made possible by technical advancements. We require cutting-edge tools and technologies to identify, investigate, and take prompt action on new assaults and threats as the cyber threat scenario gets worse. To identify various sorts of attacks, a cyber-security system must be constructed. The use of various intelligence algorithms in cyber security enabled the detection and analysis of attacks on computer networks. Artificial intelligence, machine learning, and deep learning algorithms are used in cyber security to take the best feature representation possible out of a large data set. This has been used in a number of cyber security scenarios, including the analysis, prediction, and detection of attacks. The purpose of this work is to analyze cyber security attack datasets using clever methods. Additionally, it offers a thorough comparing of algorithm effectiveness and field application to explain the advantages of network protection optimization methods. We analyze the key traits of reflective deep learning approaches used in cyber security application domains, we bring the recent developments in deep learning, and we offer an insight of resources required like a systematic framework and appropriate datasets before analytically and comparably assessing state-of-the-art services from the research. We point out the shortcomings of the examined works and present a picture of the current problems in the field, offering helpful advice and best practices for academics and developers tackling relevant issues. Finally, we identify current research directions and pain issues that need to be addressed.

Index Terms **- Cyber Attacks, Intelligent Algorithms, Cognitive Science, Machine Learning, Deep Learning**

I. INTRODUCTION

Cyberattacks have increased dramatically as a result of the quick evolution of computing networks. Every sector of our organization, including management, frugal living, and vital infrastructures, depends heavily on computing networks and data processing solutions. Then, it is clear that they are vulnerable to cyberattacks. An attack launched from one or more calculations against other calculations or networks is referred to as a cyberattack. Cyberattacks often try to either disable the goal calculation, take the tasks offline, or get access to the goal calculation's dossier.Indeed, maintaining cybersecurity has grown to be one of the most difficult challenges in the field of computer science, and it is anticipated that cyberattacks will continue to get more sophisticated and numerous.

In order to protect computing resources, networks, programs, and data against unwanted access, modification, or annihilation, cybersecurity defense measures must be developed. New cybersecurity dangers are rapidly emerging as a result of the information and communication technologies' significant advancements. The sophistication and speed of cybercriminals' attacks are growing as they adopt new methods. Therefore, there is a need for more adaptable, versatile, and strong cyber defense systems that can recognize a wide range of threats in real-time. Artificial intelligence (AI) approaches have become more widely used in recent years and continue to play a significant role in the detection and avoidance of cyber threats.

Although the idea of AI was first put forth in the 1950s, it has recently grown at a rapid rate and is now having an impact on all facets of society and occupations. AI is useful in many industries, including gaming, NLP, medical, production, schooling, and others. The cybersecurity industry, where AI has been used for both attacking and defending in cyberspace, is also being impacted by this trend. Cyber threats can use AI to increase the sophistication and scale of their offensive operations. On the defense front, AI is used to improve defense tactics, making defense systems stronger, more adaptable, and more effective. This entails being responsive to environmental factors to lessen the effects that may occur.

The cyber world is experiencing an increase in cyberattacks. To reduce or prevent the number of cyberattacks, enhanced security measures should be implemented. There are many different types of attacks, including DDoS attacks, Man in the Middle, information espionage, PROBE, User-To-Root, and Remote-To-Local attacks. Crooks or squatters use these assaults to get illegal access to any private network, websites, data, or maybe our personal machines. Therefore, to protect the sensitive information, information, and financial data, outside or inside hackers utilize cutting-edge approaches or find ways to irritate or breach any defense systems. Intelligent intrusion weaponry should strive to manage or block a variety of inventive attacks that hackers have designed or coded.

Networks, devices, processes, and data are protected by networks, processes, and practices known as cybersecurity from intrusions, damage, and unauthorized access. "Cybersecurity relates to the set of actions and procedures, both technological and non-technical, geared to secure the 'real geographical location' of

cyberspace as well as equipment, software, and the content they contain conveyed, from all potential dangers," says Myriam Dunn Cavelty [1]. One of the most crucial challenges in cyberspace nowadays is cybersecurity [2, 3].

Traditional cybersecurity techniques rely on static control of security equipment and operate in reaction to an attack. Security systems, for instance, keep an eye on nodes in the event of network intrusion attacks in accordance with a predetermined set of criteria. These procedures hold off until they receive word that an attack has happened. However, the conventional strategy is no longer effective given the rise in cyberattacks.The recent Equifax hack in 2017 is only one illustration of how inadequate standard cybersecurity techniques are, putting sensitive data at serious risk by exposing information for up to 143 million consumers.

Researchers have released a number of surveys in the fields of cybersecurity and AI. Some of them, however, were just concerned with using machine learning techniques for online issues such those in [4–7]. Other studies [8,9] only looked at deep learning techniques. Additionally, there is a dearth of literature addressing the sinister application of AI. A survey on ML and DL techniques for cyber security was conducted by Apruzzese et al. [10]. However, their research only looked at attacks involving malware analysis, spam identification, and network intrusion detection. The author of [11] explored how cybersecurity and AI interact. The paper evaluated certain ML and DL strategies to defend against cyberattacks in more detail.

We will apply these concepts to cyber security using machine learning technologies to strengthen the defenses built into the intrusion detection system. The information must first be fed into the machine-learning model. The dataset collection trains the model, resulting in a trained model. The next step is to employ and implement the machine-learning procedure after feeding the dataset instance.The improvement of the protective mechanisms in this intrusion detection system is mostly due to machine learning algorithm. The two categories of ML algorithms are supervised learning and unsupervised learning. By the information (i.e., input) they choose, they can be distinguished from one another. Algorithms that are given a set of labelled training data with the objective of determining what sets the labels apart are said to be learning under supervision. Unsupervised learning describes techniques where algorithms are given unlabeled training data and left to deduce the classes on their own. Most of the time, labelled data is extremely rare, or even labelling the data itself is a laborious operation, and we may not be able to tell if labels are indeed present.In order to provide the best solutions for cyber environments and strengthen cybersecurity capabilities against cyber-attacks, this research examines the necessity for the evolution of cybersecurity strategies. Additionally, it gives a brief explanation of a few AI subset technologies, including deep learning, expert systems, machine learning, and bio-inspired computations [12-14].

The following is a list of this study's main contributions:

- We briefly describe AI and examine its effects in the cyber realm in order to illustrate the impact of AI techniques on cybersecurity.
- Applications of AI for Cybersecurity: We survey the use of AI for cybersecurity, which includes a wide spectrum of cyber-attack kinds.
- We examine different potential dangers and attacks that may occur through the usage of AI systems in our discussion of the potential security threats from hostile uses of AI technologies.
- Hurdles and future directions: We talk about the open research directions for AI in cybersecurity as well as potential research challenges.

The reminder of the research article is as follows: section 2 gives the overview of current research works in AI for Cyber Defense. Section 3 provides the impactfulness of AI, ML and DL for cyber security. Section 4 provides the working nature and their possibilities of various techniques involved through ML and DL for cyber defense. Section 5 gives the overall nature of the surveyed literature and research directions. Section 6 concludes the work.

II. RELATED WORKS

In the field of cybersecurity research, AI models are a significant player. However, it can be difficult to guarantee that AI models are used correctly in a cybersecurity setting. A thorough overview of the uses of ML in cybersecurity is given by Shaukat et al. [15]. Network security, computer security, mobile security, etc. are all examples of cybersecurity applications for AI.

A computer system and network's setup and execution have internal and inherent flaws that lead to weaknesses that make it vulnerable to threats and cyberattacks. Vulnerabilities in creating a computer network system include improper setup, inadequate procedures, and unskilled or unskilled staff. These flaws enhance the likelihood of threats and attacks coming from both inside and beyond a network. People from a wide range of professions are increasingly depending on cyber systems. An entity that alters the operations and behavior of a computer or network using a certain penetration technique is referred to as a threat [16]. The goal of cyber security is to safeguard data, networks, and software against online attacks [17].

There has been a competition between cybercriminals and guards since the first computer virus appeared in 1970 [18]. Fighting these cyber security threats and keeping up with their increasing speed is becoming more and more difficult. To address these security difficulties, academics are currently concentrating on the urgent need for new automated security methods. Utilizing automated machine learning approaches to find new and undiscovered cyberthreats is one of the greatest and most effective practices [19].

Machine learning algorithms have been researched and examined in several articles. The authors of [20] compared the effectiveness of the C 4.5 method and the support vector machines (SVM). The two algorithms' accuracy was examined in relation to four different computer network assaults. The results show that C 4.5 performs better than SVM. This discovery is consistent with the outcome we found using the J48 tree, a Weka version of the C 4.5 algorithm.The authors of [21] looked into a number of classifier algorithms

for attacks such as denial of service (DoS), remote to local (R2L), user to root (U2R), and surveillance (PROBE). The authors made an effort to choose the most effective algorithms for every type of assault. The findings indicate that the best algorithms for PROBE, U2R, and R2L assaults, respectively, are Naive Bayes, Bayes Net, and One-R. The majority of the algorithms reportedly performed admirably for the DoS attack category.

The lack of qualified employees with experience in these specialized cybercrime detection technologies can be addressed by ML approaches. Additionally, aggressive strategies are required to identify and respond to the emerging generation of assaults (automated and evolutionary). One way to respond swiftly to such attacks is through the use of machine learning (ML), which can learn from past attacks and promptly counter new ones. Nearly all facets of cyber security are currently using machine learning and deep learning models to identify and combat cyberattacks [22].To mention a few, SEIM Solutions [23], intrusion prevention system (IPS) [24], unified threat management (UTM) [25], firewalls, and antivirus software are currently being deployed as traditional cyber security systems. These conventional solutions rely on static control of devices in accordance with specified rules for network security and lack automation (use of AI techniques). In terms of mistake rate, performance, and defending against the cyberattack, the AI-based system outperforms conventional threat detecting methods [26].

AI-based systems perform attacks detection and response with a lower mistake rate than conventional systems. When detecting and responding to an attack, AI-based systems perform better than conventional ones in terms of error rate, accurate prediction of an assault, and the number of false positives. AI-based solutions also speed up the process of identifying network flaws, repairing them, and patching compromised networks [27]. Research found that more than 60% of assaults are discovered after they had already damaged the internet [28]. To address these security issues and dangers, new automated security techniques are currently required.Smartphones are targets of cybercriminals because to their fast expansion and availability of complex functions. Additionally, machine learning techniques are essential for enhancing the effectiveness of techniques for detecting and preventing attacks to mobile devices [29].

In the early stages of developing intrusion detection systems, ML/DM (Data mining)-based cyber analytics support was investigated. Anomaly-based techniques simulate the behavior of the system and the conventional network, making it easier to identify anomalies as deviations from the norm. Its advantage is that traditional activity profiles are built specifically for each system, application, or network, making it difficult for attackers to understand the kinds of actions that can be carried out covertly and without detection. They appear appealing because of their unique capacity to detect zero-day assaults.Anomaly detection and misuse are combined in hybrid approaches. They are used to lower the frequency of unidentified attacks and increase the detection rates of acknowledged intrusions. Once more, the creation of intelligent intrusion detection systems depends on the availability of a solid data set. An information set with a lot of high-quality data and one that simulates real time will only make it easier to train an intrusion detection system's associated check.

The roles that machine learning techniques play are on both the attacker and cyber security sides. Cyber attackers and criminals are employing ML approaches to identify the system's weaknesses and

sophisticated attack strategies to get past the defenses. In order to increase performance and early attack detection to lessen the impact and damage that occurs, ML models are playing a crucial role on the defense side [30, 31]. Combining machine learning approaches improves the accuracy of early and accurate cyberattack categorization [32].

This article's goal is to discuss the main machine learning methods used in cyber security and to highlight the growing popularity of these methods. The use of machine learning approaches to identify and characterize cyberattacks, such as intrusion detection, virus detection, and spam detection on both computer networks and mobile or smartphone devices, has been briefly described here.We are not aware of any survey that addresses the use of ML approaches in computer and mobile network cyber security. Our work also gathered together future challenges like trustworthiness and adversarial machine learning, as well as frequently used ML tools, security datasets, a graphic summary of key cyber security components, and ML techniques that are currently available to combat threats and attacks in cyberspace. Many recent surveys either offer applications in a particular field or fail to provide the fundamental information a new researcher would need to enter or comprehend that field. Additionally, the majority of survey articles primarily cover specific network dangers and assaults.

Machine learning techniques in particular have raised dangers to computer networks while also holding great promise for the identification and categorization of assaults and threats on mobile devices and networks. Our study examines cyberthreats to computer networks and mobile devices. When compared to previously published survey papers in the field, our survey is comprehensive and distinctive in that it offers the following features: basic insights into cyber security threats on both mobile devices and computer networks; characterizations of frequently used safety datasets; summaries of state-of-the-art ML techniques to manage these risks; indications of prevalent ML tools; characterizations of performance indicators to assess the effectiveness of ML techniques; and a list of current research initiatives.

To combat these cyberattacks, we have developed a graphic description of the key cyber security components and machine learning tools now in use.

III. IMPACT AI, ML & DL IN CYBERSECURITY

3.1 What is AI?

In the past ten years, the field of artificial intelligence (AI) has gained popularity and become a common notion. John McCarthy first introduced the term AI in 1956. He defined AI as an approach that formalizes fundamental facts about events and their effects using mathematical reasoning. Artificial intelligence, or AI, is intelligence presented by a machine. It makes it possible for programmers to write their programmes quickly.AI simulates human thought by using intricate mathematical algorithms. AI technologies are capable of comprehending, learning from, and acting on data obtained from events and effects.

There are two ways that AI can be defined. It is a science that aims to understand the nature of intelligence and create intelligent machines by applying knowledge, logic, self-learning, and tenacity to make machines intelligent. Simply said, intelligent machines are made by humans. This intelligence has the

same abilities as human cognition, including the ability to think, learn, make decisions, and work. On the other hand, scientists define artificial intelligence (AI) as a branch of study that investigates and creates solutions for complexity issues that can never be handled without the use of intelligence.For instance, with vast volumes of data, scientists can create an AI system for real-time analysis and decision-making. Artificial intelligence (AI) has recently led to advancements in a wide range of sectors of science and technology, including computerized robots, image recognition, natural language processing, expert systems, and others.

People's daily lives and jobs have been significantly impacted by the internet's and computing technology's quick development. Unfortunately, it also gave rise to a number of new, difficult cybersecurity problems: First off, manual analysis is impractical due to the explosion of data. Second, threats are expanding quickly, which also results in the emergence of new, transient species and highly adaptive threats. Third, the threats now use a variety of strategies for proliferation, infection, and evasion, making it difficult to identify and forecast them. Additionally, the cost of averting hazards should be taken into account. A lot of time, money, and effort must be invested to create and put into practice an algorithm. Additionally, it is difficult and expensive to hire or train specialists in the subject.

Stuart Russell and Peter Norvig stated that “AI aims not simply to understand but also to produce intelligent beings”. In their definition of AI, which they divided into two broad categories as depicted in fig 1:

- Both mental process and reasoning-which can be divided into the categories of human and rational thinking-measure success in terms of thinking.
- Behavior, which can be divided into acting humanely and acting logically, assesses success based on the ideal performance and action.

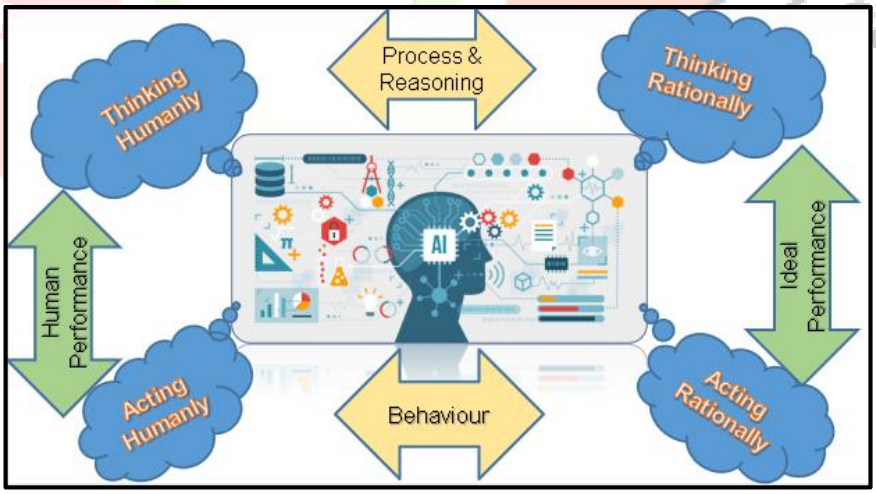

Fig. 1 Ways of Defining AI

The diagnosis and look-ahead components of decision-making theory are two. AI pays insufficient attention to this element and disregards multi-attribute human reasoning since look-ahead judgements are unpredictable. A new kind of artificial intelligence that responds like human intellect is what AI aims to create. Machine learning must be accurate in order for this to be accomplished, hence learning methods must be used to train the machines. Algorithms are used in AI techniques. However, even if methods are not greatly improved, AI can exploit large data and extremely powerful computing to learn by sheer force.

AI works in 3 ways:

- Assisted intelligence is intelligence that helps individuals accomplish things better already.
- People with enhanced intelligence are able to perform tasks they previously could not.
- Autonomous intelligence, which are characteristics of autonomous machines.

Given these three categories, it is possible to draw the conclusion that AI aims to solve some of the most challenging issues, and cybersecurity fits into this category because cyberattacks have developed into highly sophisticated, potentially disastrous issues that have become complicated in the cyberspace.Additionally, a variety of threats are constantly developing and spreading. Therefore, AI-based approaches are anticipated to handle these cybersecurity concerns.

3.2 Machine Learning & Deep Learning

Machine learning (ML) is a subfield of artificial intelligence that uses data to enable systems to learn and advance without explicit programming. Mathematical methods that enable the process of information extraction, pattern recognition, and inference from data are closely related to machine learning (ML). Although there are many distinct kinds of machine learning algorithms, they can be broadly divided into three groups: reinforcement learning, unsupervised learning, and supervised learning. Decision trees (DT), support vector machines (SVM), Bayesian algorithms, k-nearest neighbor (KNN), random forests (RF), association rule (AR) algorithms, ensemble learning (EL), k-means clustering, and principal component analysis are the common machine learning (ML) algorithms used in the computer security domain (PCA).

A subfield of computer science known as artificial intelligence (AI) creates methods, theories, and software. Early attempts to adopt a streamlined model that was modelled after how neurons trigger other neurons in a biological system, such as an organic brain, led to the development of Artificial Neural Networks (ANNs). AI's machine learning (ML) division. Models are created by ML algorithms based on training data, enabling the models to make predictions (or choices) about incoming data without being explicitly told how to do so. ML has uses in a variety of contexts [33].

Regarding approach, machine learning can be divided into three main groups: semi-supervised machine learning, unsupervised machine learning, and supervised machine learning. In supervised machine learning, the intended labels or classes for the data are already known, and these labels or classes are used to train for calculations, such as classification and regression. The targeted value is unknown in unsupervised machine learning. Unsupervised learning is primarily concerned with identifying patterns in the data. It functions by identifying patterns in the data, such as clustering. Semi-supervised machine learning (ML) refers to a method where some of the data is labelled or requires the assistance of human specialists during the data collection phase.Semi-supervised machine learning (ML) refers to a method where some of the data is labelled or requires the assistance of human specialists during the data collection phase. The human expert during the labelling procedure will undoubtedly assist in solving the issue and enhancing the model's accuracy. Another branch of machine learning is called reinforcement learning (RL). Because there is feedback to the algorithms against any incorrect prediction, RL is sometimes referred to as learning with a critic. The algorithm hasn't been informed of how to remedy it, though.

A kind of machine learning called deep learning (DL) uses data to educate computers to perform tasks that, at the time, can only be performed by humans. The processing of impulses by neurons in the human brain serves as its inspiration. The fundamental idea behind deep learning is that when we build larger neural networks and give them more training data, their performance keeps improving. The main benefit of DL over traditional ML is that it performs better on large datasets. supervised learning, unsupervised learning, and reinforcement learning are all features of DL approaches, just like ML methods. The advantage of DL is the use of unsupervised learning to automatically select feature.Feed forward neural networks (FNN), convolutional neural networks (CNNs), recurrent neural networks (RNN), deep belief networks (DBNs), stacked autoencoders (SAE), generative adversarial networks (GANs), restricted Boltzmann machines (RBMs), and ensemble of DL networks are typical DL algorithms frequently used in the cybersecurity domain (EDLNs).

3.3 Bio-inspired Computation

It is a group of clever algorithms and techniques that utilize biological traits and behaviors to address a variety of challenging issues. Because of how they learn, traditional AI and bio-inspired AI are not the same. Machines can build traditional AI, which produces intelligence. These programmes, which also produce other programmes with intelligence, produce this intelligence. However, simple rules and creatures that closely adhere to those laws form the foundation of bio-inspired computing. These organisms gradually evolve in response to some circumstances. The approaches listed below are the ones most frequently utilized in the cybersecurity field among bio-inspired computations: genetic algorithms, evolutionary algorithms, ant colony optimization, particle swarm optimization, and artificial immune systems [12].

IV. RESEARCH METHODOLOGIES OF AI, ML & DL IN CYBER DEFENSE

Thanks to improvements in computing technology, our society is changing quickly. This has a big impact on people's daily lives and jobs. Machines that can think, learn, make decisions, and solve problems similarly to humans have been made possible by some of these technologies. AI, as an illustration, adopts intelligence and has the capacity to analyze massive volumes of data while performing real-time analysis and decision-making. The use of AI techniques is advantageous in many domains of science and technology. It goes without saying that there is a ton of personal information on the Internet, which leads to several cybersecurity problems.

First, the amount of the data makes manual analysis all but impossible. Second, there may be dangers based on AI or rising threats. Additionally, the expense of preventing threats rises due to the high cost of hiring specialists. The development and application of algorithms to identify those dangers likewise involves a significant amount of time, money, and effort. Utilizing AI-based techniques is one remedy for those problems.

AI is capable of quickly, correctly, and efficiently analyzing massive amounts of data. An AI-based system can predict future assaults that will be similar to those that have already occurred by using threat history, even if the patterns of those attacks vary. Because of these factors, AI can be applied in cyberspace

[12]: AI can handle vast data, find new and significant changes in attacks, and continuously improve its security system's response to threats.

AI does, however, have certain drawbacks. For example, an AI-based system requires a sizeable amount of data, and processing this volume of data demands a lot of time and resources. Frequent false alarms are also a problem for end users, and delaying any necessary responses reduces efficiency. In addition, attackers can harm the AI-based system by introducing hostile inputs, data tampering, and model theft.

Recently, scientists presented a number of ways that made use of AI techniques to identify domains produced by domain generation algorithms, detect network intrusions, and categorize malware, phishing, and spam campaigns (DGAs).

In this area, we divide the literature into four primary categories: identifying malware, detecting network intrusions, recognizing phishing and SPAM, and other, which includes identifying DGAs and thwarting APT. The main areas of applying AI for cybersecurity are shown in Figure 2.

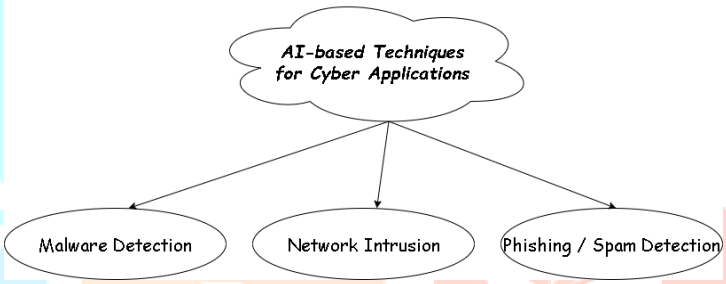

Fig. 2 AI based techniques for Cyber Applications

This section also describes common machine learning techniques. Figure 3 gives the overview of major research-oriented techniques involved and Figure 4-6 gives a virtual view of different methodologies. Table I provides a compact overview of ML models including the time complexity, pros, and cons.

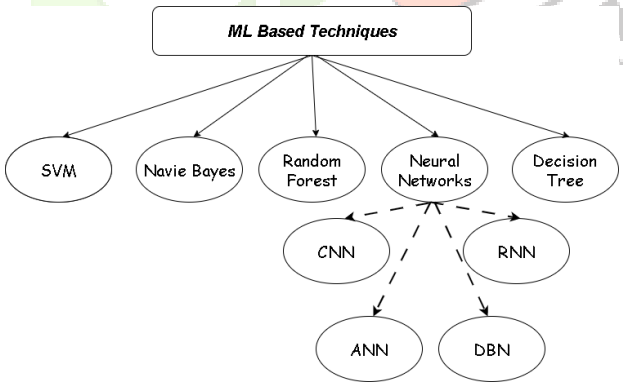

Fig. 3 ML Based Techniques

4.1 SVM – Support Vector Machines

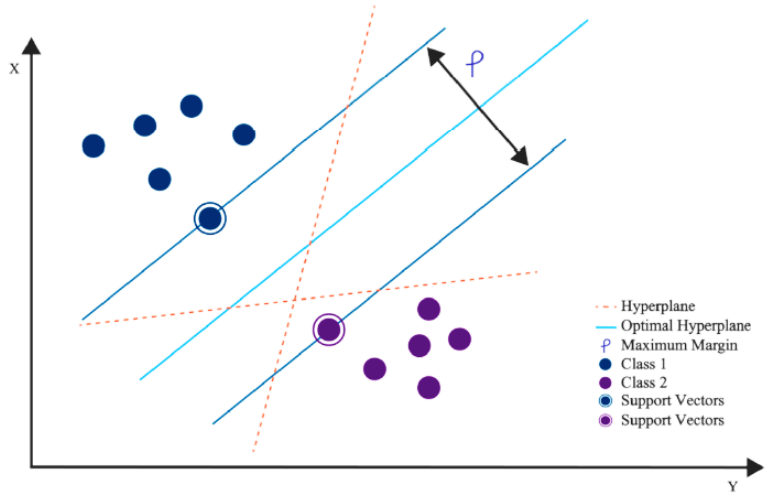

Fig. 4 Support Vector Machine Model

Support Vector Machine (SVM) is regarded as the most popular and effective machine learning (ML) technology for cyber security jobs, particularly for IDS. Based on the margin notation on either side of the hyperplane, SVM divides and classifies the two data classes. Fig. 4 shows the model of SVM.

4.2 Decision Tree

A supervised machine learning method based on a recursive tree-structure is called a decision tree (DT). As shown in Figure 5, DT is made up of three nodes: a root orintermediate node, a path, and a leaf node.

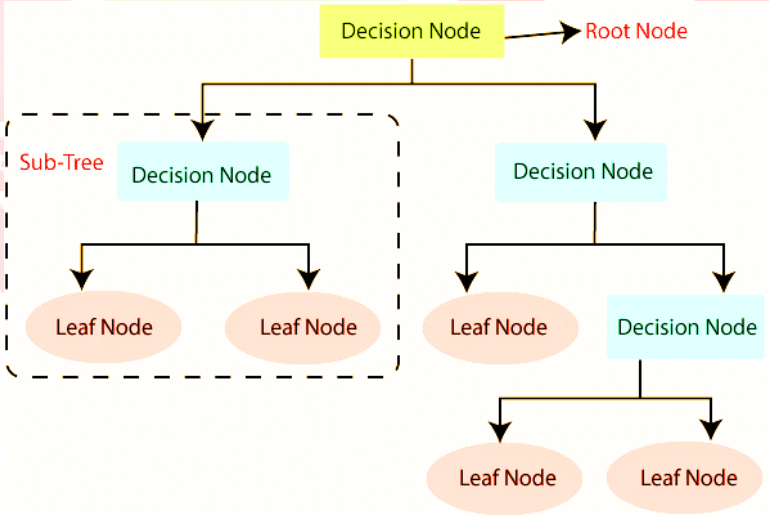

Fig. 5 Decision Tree Model

4.3 Naïve Bayes

The Naive Bayes (NB) class of classifiers decomposes the conditional probability of an issue under analysis using Bayes' theorem (or Bayes' Rule). However, in terms of cyber security, different attack types do not satisfy this criterion of independence.

Figure 6 shows the various representation of neural networks.

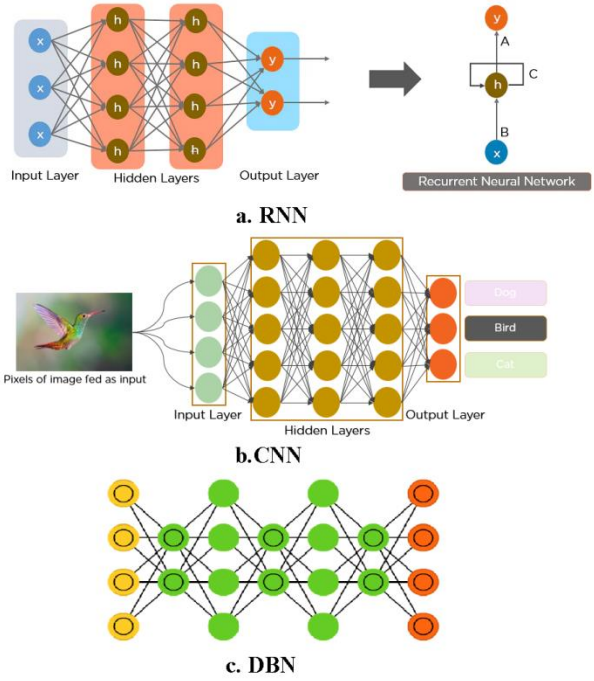

Fig. 6. Graphical Views of a. RNN, b. CNN and c.DBN

Table 1. Comparison of Various common ML Techniques

ML Technique	***Time Complexity***	***Pros***	***Cons***
SVM	$O(n^2)$[1]	Used for Classification & Regression	High computational Cost
Naïve Bayes	$O(mn)$[2]	Less Computational time due to probabilistic classifier	Massive data required for quality results
Random Forest	$O(Mm\ \log n)$[3]	Composed of many Decision Trees	High computational Cost
Decision Tree	$O(mn^2)$[5]	Works on if-then rules	Complex and expensive
Neural Networks	-	DBN - Higher performance RNN – sequential data CNN – a smaller number of neurons	DBN – resource consumption high RNN – difficult to train CNN – requires more convolution layers

The above algorithms or techniques utilizes the following evaluation criteria or metrics to prove their performance:

- True Positive (TP) – count of positive samples
- True Negative (TN) – count of attacks
- False Positive (FP) – count of attacks misclassified as positive
- False Negative (FN)– count of attacks misclassified as negative

- $Precision = \frac{TP}{TP+FP}$ (1)
- $Recall = \frac{TP}{TP+FN}$ (2)
- $Accuracy = \frac{TP+TN}{(TN+FN+TP+FP)}$ (3)
- $Error\ Rate = \frac{FP+FN}{(TN+FN+TP+FP)}$ (4)
- Miss Rate
- $False\ Positve\ Rate\ (FPR) = \frac{FP}{FP+TN}$ (5)
- $False\ Negative\ Rate\ (FNR) = \frac{FN}{FN+TP}$ (6)
- F1-Score
- Received Operating Characteristics
- Area Under Curve
- Root Mean Square Error (RMSE)

The following table 2 shows the performance of the above said algorithms under various criteria mentioned.

Table 2 Comparative Analysis of ML Models

Learning Model	**Sub-domain**	**Results**		
		Accuracy	**Precision**	**Recall**
SVM	Email – Spam	95.23%	94.52%	92%
NB		98.28%	98.47%	97.21%
RF		96.25%	96%	96%
SVM	Anomaly-Based Intrusion Detection	92.57%	90.14%	61.68%
NB		98%	85%	76%
RF		87%	90%	91%
SVM	Malware Detection	93%	86%	100%
NB		94.60%	94.60%	94.60%
RF		92.56%	87.63%	92.56%

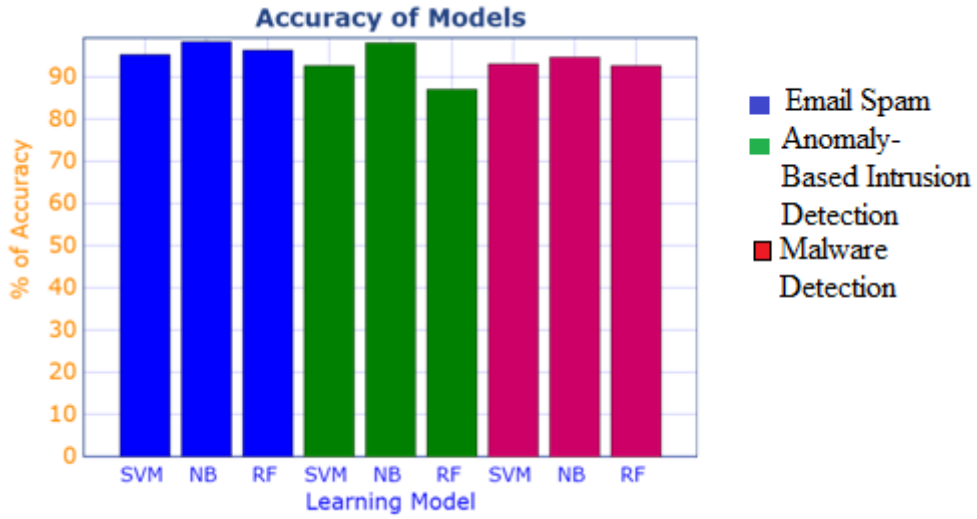

Fig. 7 . Comparative Analysis of ML Models for various sub-domains

Fig. 7 gives the accuracy of the various models under the sub-domains Email – Spam (blue), Anomaly-Based Intrusion Detection (Green), and Malware Detection (Red). This proves most models give more accurate detection of security issues. In each of the sub domains Naïve Bayes is providing high accuracy.

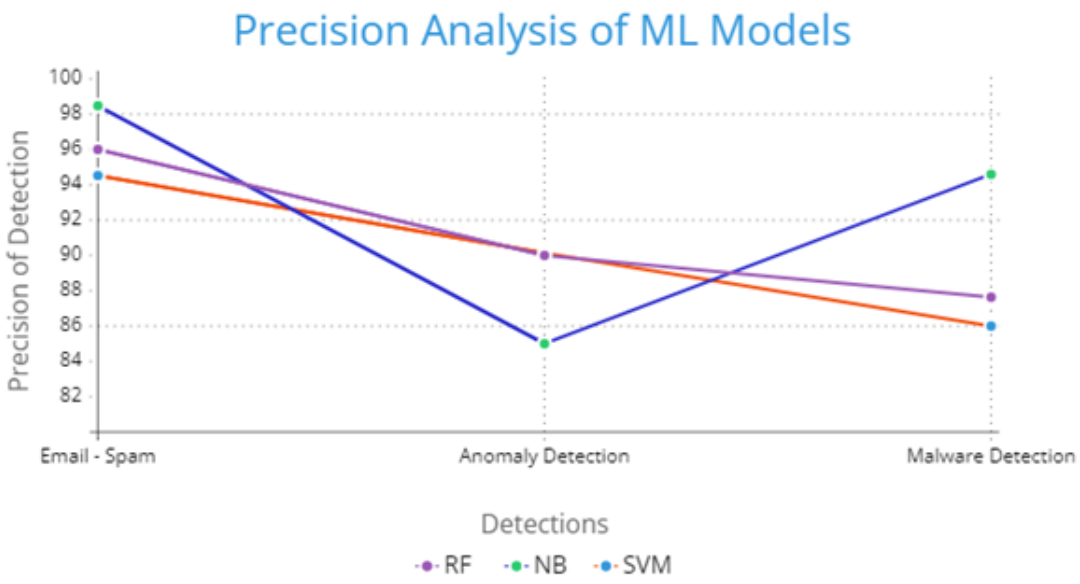

Fig.8. Comparative Analysis of Precision for ML Models

From Fig.8 it is understood that Naïve Bayes provides high precision value for email-spam detection and malware detection, SVM for Anomaly-Based Intrusion Detection.

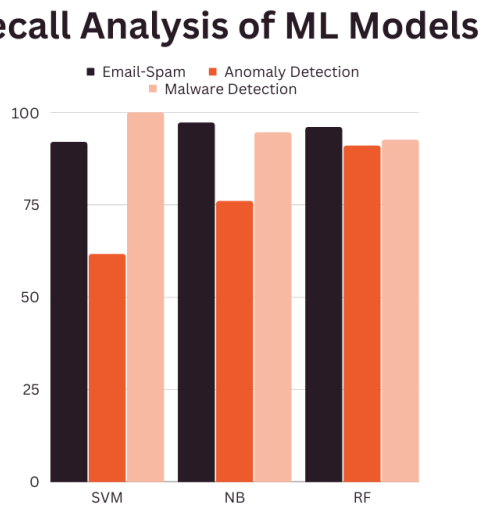

Fig.9. Comparative Analysis of Recall efficiency in ML Models

From the fig.9 it is understood that RF is more accurate in Anomaly Detection, NB in Email-spam and SVM in Malware detection.

As cyber-attacks and malware evolve over time, they pose increasing challenges for traditional cyber security measures, making it difficult to detect and mitigate new generation attacks. Artificial intelligence (AI) can provide a solution to such problems. AI models have the capability to learn from historical attack data and apply that knowledge to effectively respond to emerging attack types. Traditional security measures, including access control, antivirus software, cryptographic software, intrusion detection systems (IDS), intrusion prevention systems (IPS), sandboxes, and Security Information and Event Management (SIEM), have been employed to safeguard against attacks. However, despite the presence of these approaches, the number of attacks continues to rise. AI offers the potential to bolster cybersecurity defenses by leveraging its ability to analyze vast amounts of data and detect patterns that may go unnoticed by traditional methods, thereby enhancing threat detection and response capabilities.

In the current landscape, several threats and challenges persist, including the evolution of ransom ware, the expansion of AI, IoT threats, vulnerabilities in server less applications, and more. Cyber security is inherently driven by technology, and the introduction of new technologies not only affects cyber security but also gives rise to novel types of attacks and defense mechanisms. Looking ahead, emerging technologies such as quantum computing, cloud computing, predictive semantics, behavioral identity, and dynamic networks will undoubtedly impact the field of cyber security. These advancements will necessitate the adoption of new approaches to address security threats effectively. Machine learning (ML) and deep learning (DL) are expected to play increasingly prominent roles in cyberspace as these new security threats emerge. By leveraging ML and DL, cyber security professionals can enhance their ability to analyze vast amounts of data, detect patterns, and develop proactive defense mechanisms in response to evolving cyber threats.

Fig. 10 illustrates the comparison of different IDS approaches through a graph. Additionally, we discussed five methods for malware detection. Out of which one was unsupervised learning approach. The comparison of the entire Malware Detection model's performance can be seen in Figure. 11.

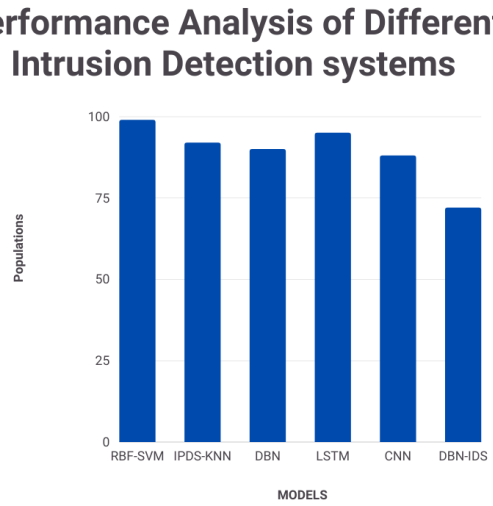

Fig. 10. IDS Approaches Performance Analysis

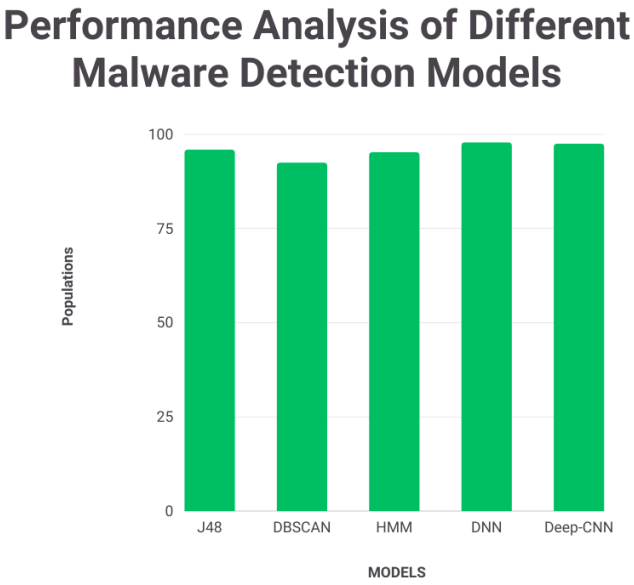

Fig. 11. Malware Detection Approaches Performance Analysis

Machine learning techniques have become a crucial aspect of the modern cyber world, especially in the realm of cyber security. These techniques are employed on both the attacker and defender sides. On the attacker side, machine learning is used to discover new methods to bypass security systems and firewalls, thus evading detection. Conversely, on the defender side, these techniques aid security professionals in safeguarding security systems from unauthorized access and illegal penetrations.

This paper presents a comparative analysis of machine learning techniques utilized for detecting cyber security threats, with a specific focus on three significant threats in cyberspace: intrusion detection, spam detection, and malware detection. The study assesses six machine learning models, namely, random forest, support vector machine, naïve Bayes, decision tree, artificial neural network, and deep belief network.

Further, the comparison is extended to different sub-domains of each cyber threat. For intrusion detection, the sub-domains considered are anomaly-based, signature-based, and hybrid-based detection methods. In the case of malware detection, the sub-domains involve static detection, dynamic detection, or hybrid-detection techniques. For spam detection, the sub-domains entail the mediums on which the models are applied, such as images, videos, emails, SMS, or calls.

By exploring various machine learning models and their performance across diverse sub-domains of cyber security threats, this research aims to provide insights into the effectiveness of these techniques in countering specific cyber threats effectively. This comprehensive analysis aids in understanding the strengths and limitations of different machine learning approaches and assists in developing more robust and efficient cyber security defense systems.

Fig.12 shows the performance comparison of six machine learning techniques to detect intrusion detection.

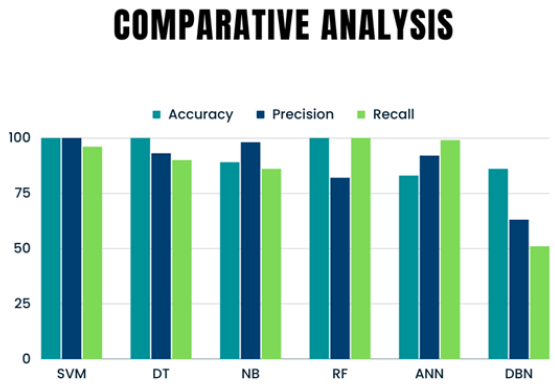

Fig.12. Comparative Analysis of Intrusion Detection using Machine Learning Techniques

4.4 5G Security for real-world applications

The arrival of 5G technology brings forth a host of advanced and rapid communication capabilities, surpassing anything witnessed thus far. Nevertheless, this groundbreaking system also presents heightened security challenges that must be addressed. While significant advancements have been achieved in the field, it is crucial to recognize that the imminent advent of 5G will render existing intrusion detection and security systems obsolete unless they are appropriately enhanced.

The analysis conducted in this paper explores various attack prediction methods, revealing three prominent approaches derived from the literature review. The initial examination reveals that the majority of prediction methods employed in the field of security utilize models or projects to represent the state of security situations and potential attacks. Predictions are made by employing models that exhibit clear divisions, encompassing discrete models and continuous models, both of which are instrumental in projecting attacks and forecasting network security. These two primary cases often intersect and complement each other in numerous instances. In the subsequent scenario, a noteworthy trend emerges, where a significant portion of approaches relies on machine learning (ML) and data mining techniques, thereby revolutionizing the current state of research in the domain of cyberspace prediction.

The data mining approach addresses vulnerabilities by constructing an artificial prediction model, whereas the machine learning (ML) approach refines the model based on general input. However, numerous challenges arise when evaluating predictions in the cyberspace realm. Existing datasets often suffer from outdated or unreliable information, having been created for different purposes. Additionally, evaluating predictions on live networks proves to be non-reproducible, as there is no standardized set of metrics to compare different methods. Research indicates that attack prediction lags behind intrusion detection and suggests several directions for improvement. Firstly, there is a shift towards processing network data and

transitioning from batch-based alerting to stream processing, driven by the advancement of big data analytics. Secondly, cooperative efforts in attack prediction, such as intrusion finder systems or distributed warning alert platforms, are being explored. In the future, machine learning and data mining will undoubtedly play crucial roles in attack prediction simultaneously. By using machine learning alone, a better understanding of attacks and their prediction can be achieved. Alternatively, utilizing both machine learning and data mining for attack prediction can enhance pattern matching capabilities. However, the fundamental challenge in attack prediction lies in the absence of a definitive method for successful and accurate forecasting of cyber-attacks. Although attack prediction has not yet been applied in practice, it remains an important research problem to be addressed.

V. CHALLENGES & RESEARCH DIRECTIONS

5.1 Challenges

AI techniques have been essential in cyber security applications and will keep moving in a path that is intriguing for research. However, there are various concerns that must be taken into account while using AI-based cyber security techniques. The accuracy of AI models is a major roadblock, to start. Particularly, erroneous alerts might waste processing time or cause an AI system to completely miss a hack. The fact that many of the current methodologies are model-free methods is another obstacle to acceptance.

These models need a lot of training data, which is difficult to come by in actual cyber security practice. Next, strategies for creating AI-based cyber security solutions need to take the opponent into account. Since adversarial attacks are a byproduct of a conflict between AI systems, they are difficult to identify, avoid, and defend against. AI can assist in defending the system from cyber threats but it can also enable harmful assaults, or attacks that use AI. AI can be used by malicious actors to create flexible, complex attacks that evade defenses and breach computer networks or systems.

Techniques for machine learning are frequently employed in the field of cyber security. There are a number of obstacles in this path, though. While building the models, ML approaches require a sizable number of high-performance resources and data. Using numerous GPUs is one option, but this is neither a cost-effective nor a power-efficient solution. Additionally, ML methods are not intended to identify cybercrimes. Traditional ML algorithms did not place much emphasis on cyber security. It is necessary to have strong and reliable ML methods that are especially made to counter security threats and deal with adversarial inputs. It should be noted that different security assaults cannot be effectively detected by a single ML model. It is necessary to have strong and reliable ML methods that are especially made to counter security threats and deal with adversarial inputs. It should be noted that different security assaults cannot be effectively detected by a single ML model. To address a particular kind of cyber attack, a corresponding ML model should be created. Another difficult task is the early attack prevention. These real-time and zero-day threats should be able to be quickly detected using ML approaches.

The creation of a cyber environment will result in social and technological problems. covering topics including artificial intelligence (AI), the internet of things (IoT), big data, smart environments, and quantum computing, and will examine the complicated concerns that may arise as a result of their adoption. The

hypothesis of cyber warfare, which has been postulated for about twenty years now, has occasionally come to pass. Cyber warfare is more difficult to describe. Cyber warfare differs depending on the nation and location.

The following are the present challenges that will have a significant impact on how future cyber security is designed and operates: passwords, IoT, quantum computing, artificial intelligence, and big data privacy and security. Therefore, we ought to pay more attention to the problems we now have when designing and running new cyber security systems. What then can people and businesses do to better prepare for the following challenges: The first step toward a better future for cyber security is a change in mentality. The most recent strategy for cyber security is the perimeter defense model, which keeps an organization secure solely from the outside using a coconut-like hard shell.

In the belief that the shell is thick enough, this hard-shell has been constructed around the entire organization. When an insider threat results as a side effect, this strategy fails. The avocado approach, then, is another more secure strategy in which companies are secured by multiple layers. These days, organizations have started using this model. The second is to think holistically. Attackers will always discover ways around designs, no matter how good they are. In order to be ready for defense, we must add a second layer of monitoring. The first layer is detection, the second is protection, and the third is response. Greater prevention results from faster response times.

The third is an artificial immune system and intelligence enhancement (AIS). If we are successful in building a powerful artificial immune system, we will be able to recognize attacks, react to them promptly, and do so at a lower cost of effect thanks to the real-time application of AI/ML. We can cope with very high impact cyber-attacks if we can put AIS at the network's edge for dealing with big volume/low impact attacks, along with a more effective centralized monitoring and response mechanism.

5.2 Open Research Directions

The integration of AI approaches and cyber security spans a wide range of promising and unexplored problems. The following are some research areas:

- First, it may still be a promising research area to combine several AI-based strategies in a protection solution. For instance, combining ML/DL techniques with bio-inspired computation yields promising results in detecting malware for network penetration. Due to the restricted use of bio-inspired algorithms in cyber security, the combination of these two methodologies represents a particularly promising study area.
- Second, research is required on how human intelligence and robots interact in cyber protection. In this human-machine architecture, the agents will carry out the task independently, with human oversight and intervention only when needed.
- Third, research has shown that threat actors could use an AI-based technique to go around or subvert AI models. Therefore, a future trend would be the defense plan against these kinds of attacks.
- Fourth, choosing the right characteristics is essential for training a model to generate successful results. Both statically and dynamically extracted features can be used to identify

malicious behavior. Most of the investigated publications used only a small number of characteristics or only statically or dynamically extracted features, which raises the possibility that some aspects that could have been crucial in identifying harmful nature were overlooked. Non-optimal characteristics were highlighted in many articles, and it is recommended that this be done in future research.

- Fifth, to prevent the discovery and analysis of their distributed software, malware developers employ a variety of anti-analysis tactics. Tools to obfuscate samples, compress/pack the binary/exe, and encrypt the file are all used to make malware harder to identify and analyze. Future development may involve reducing the effectiveness of all these anti-analysis technologies in order to completely eliminate the harm that malware poses.

VI. CONCLUSION

Information technology has dramatically advanced, creating new problems for cyber security. Cyber attacks' computational complexity necessitates the development of new strategies that are more reliable, scalable, and adaptable. In order to improve security measures to recognize and respond to assaults, cyber security has grown to be a concern on a global scale. The standard security systems that were previously in use are no longer adequate since they are ineffective at identifying new and polymorphic assaults. In a variety of applications, machine learning techniques are essential in cyber security systems. Our review has shown that there is a fast-expanding interest in machine learning and cyber security among academics, business, and government, leading to an increase in publications, notably in the previous ten years. The use of the AI-based approach to cyber security challenges is the main topic of this research. According to current research, network intrusion detection, malware analysis and categorization, phishing, and spam emails are the main goals for AI use in cyber security. The use of DL there gradually took over as the dominant trend. Researchers were also interested in the pairing of ML/DL with other intelligent techniques, such as bio-inspired methodologies. These combinations produce highly encouraging outcomes and maintain the need for more study.

The use of machine learning in cyber security was not their primary purpose. Evasion can quickly deceive the ML model by providing hostile inputs. Instead of focusing on the model's accuracy and speed, trustworthy machine learning involves the secure application of machine learning techniques for cyberspace. We have also provided a comprehensive literature in this field and briefly outlined some of the major difficulties associated with applying machine learning techniques to cyber security. Future study should pay attention to the aforementioned difficulties.

VII. REFERENCES

[1] Cavelty, Myriam Dunn, "The Routledge Handbook ofNew Security Studies," 154-162, 2018.

[2] Guan ZT, Li J, Wu LF, et al., "Achieving efficient and secure data acquisition for cloud-supported Internet of Things in smart grid,". IEEE Internet Things J, 4(6): 1934-1944,2017.

[3] Wu J, Dong MX, Ota K, et al., "Big data analysis-based secure cluster management for optimized control plane in software-defined networks,". IEEE Trans Netw ServManag, 15(1):27-38.

[4] Buczak, A.L.; Guven, E. A survey of data mining and machine learning methods for cyber security intrusiondetection. IEEE Commun. Surv. Tutor. 2015, 18, 1153–1176.

[5] Torres, J.M.; Comesaña, C.I.; García-Nieto, P.J. Machine learning techniques applied to cybersecurity. Int. J.Mach. Learn. Cybern. 2019, 1–14.

[6] Guan, Z.; Bian, L.; Shang, T.; Liu, J. When machine learning meets security issues: A survey. In Proceedings of the 2018 IEEE International Conference on Intelligence and Safety for Robotics (ISR), Shenyang, China,24–27 August 2018; pp. 158–165.

[7] Xin, Y.; Kong, L.; Liu, Z.; Chen, Y.; Li, Y.; Zhu, H.; Gao, M.; Hou, H.;Wang, C. Machine learning and deeplearning methods for cybersecurity. IEEE Access 2018, 6, 35365–35381.

[8] Berman, D.S.; Buczak, A.L.; Chavis, J.S.; Corbett, C.L. A survey of deep learning methods for cyber securityInformation 2019, 10, 122.

[9] Wickramasinghe, C.S.; Marino, D.L.; Amarasinghe, K.; Manic, M. Generalization of Deep Learning for Cyber-Physical System Security: A Survey. In Proceedings of the IECON 2018-44th Annual Conference ofthe IEEE Industrial Electronics Society,Washington, DC, USA, 21–23 October 2018; pp. 745–751.

[10] Apruzzese, G.; Colajanni, M.; Ferretti, L.; Guido, A.; Marchetti, M. On the effectiveness of machine and deep learning for cyber security. In Proceedings of the 2018 10th International Conference on Cyber Conflict(CyCon), Tallinn, Estonia, 30 May–1 June 2018; pp. 371–390.

[11] Li, J.H. Cyber security meets artificial intelligence: A survey. Front. Inf. Technol. Electron. Eng. 2018,19, 1462–1474.

[12] Thanh Cong Truong, Quoc Bao Diep and Ivan Zelinka. "Artificial Intelligence in the Cyber Domain: Offense and Defense", Symmetry 2020, 12, 410.

[13] K. Morovat and B. Panda, "A Survey of Artificial Intelligence in Cybersecurity," *2020 International Conference on Computational Science and Computational Intelligence (CSCI)*, 2020, pp. 109-115.

[14] R. Devakunchari, Sourabh, Prakhar Malik, "A Study of Cyber Security using Machine Learning Techniques", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075,Vol-8, Issue-7C2, May 2019.

[15] Shaukat K et al, "A survey on machine learning techniques for cyber security in the last decade". In: IEEE Access, vol 8, pp 222310–222354, 2020.

[16] F. Farahmand, S. B. Navathe, P. H. Enslow, and G. P. Sharp, "Managing vulnerabilities of information systems to security incidents," in *Proceedings of the 5th international conference on Electronic commerce*, 2003: ACM, pp. 348-354.

[17] D. Craigen, N. Diakun-Thibault, and R. Purse, "Defining cybersecurity," *Technology Innovation Management Review*, vol. 4, no. 10, 2014.

[18] P. Szor, *The Art of Computer Virus Research and Defense: ART COMP VIRUS RES DEFENSE _p1*. Pearson Education, 2005.

[19] I. Firdausi, A. Erwin, and A. S. Nugroho, "Analysis of machine learning techniques used in behavior-based malware detection," in *2010 second international conference on advances in computing, control, and telecommunication technologies*, 2010: IEEE, pp. 201-203.

[20] Solanki, M., &Dhamdhere, V,"Intrusion Detection System Using Means of Data Mining by using C 4.5 Algorithm", International Journal of Application or Innovation in Engineering & Management (IJAIEM), 4(5), 2015.

[21] Nguyen, H. A., & Choi, D.,"Application of Data Mining to Network Intrusion Detection: Classifier Selection Model", Proceedings of the 11th Asia-Pacific Symposium on Network Operations and Management: Challenges for Next Generation Network Operations and Service Management (APNOMS '08), (pp. 399-408). Beijing, China, 2008.

[22] R. Prasad and V. Rohokale, "Artificial Intelligence and Machine Learning in Cyber Security," in *Cyber Security: The Lifeline of Information and Communication Technology*: Springer, 2020, pp. 231-247.

[23] "Security Information and Event Management (SIEM)." https://www.esecurityplanet.com/products/top-siem-products.html (accessed Jan 09, 2023).

[24] "Top Intrusion Detection and Prevention Systems: Guide to IDPS." https://www.esecurityplanet.com/products/top-intrusion-detection-prevention-systems.html (accessed Jan 09, 2023).

[25] "Unified threat management." https://en.wikipedia.org/wiki/Unified_threat_management (accessed Jan 09, 2023).

[26] B. Geluvaraj, P. Satwik, and T. A. Kumar, "The future of cybersecurity: Major role of artificial intelligence, machine learning, and deep learning in cyberspace," in *International Conference on Computer Networks and Communication Technologies*, 2019: Springer, pp. 739-747.

[27] "How Artificial Intelligence is Transforming Cybersecurity." https://www.plugandplaytechcenter.com/resources/how-artificial-intelligence-transforming-cybersecurity/ (accessed Jan 09, 2023).

[28] A. Sharma, Z. Kalbarczyk, J. Barlow, and R. Iyer, "Analysis of security data from a large computing organization," in *2011 IEEE/IFIP 41st International Conference on Dependable Systems & Networks (DSN)*, 2011: IEEE, pp. 506-517.

[29] B. Arslan, S. Gunduz, and S. Sagiroglu, "A review on mobile threats and machine learning based detection approaches," in *2016 4th International Symposium on Digital Forensic and Security (ISDFS)*, 2016: IEEE, pp. 7-13.

[30] T. T. Nguyen and V. J. Reddi, "Deep Reinforcement Learning for Cyber Security," *arXiv preprint arXiv:1906.05799*, 2019.

[31] K. Geis, "Machine Learning: Cybersecurity that Can Meet the Demands of Today as Well as the Demands of Tomorrow," Utica College, 2019.

[32] M. Thangavel, A. S. TGR, P. Priyadharshini, and T. Saranya, "Review on Machine and Deep Learning Applications for Cyber Security," in *Handbook of Research on Machine and Deep Learning Applications for Cyber Security*: IGI Global, 2020, pp. 42-63.

[33] T. M. Alam*et al.*, "Corporate Bankruptcy Prediction: An Approach Towards Better Corporate World," *The Computer Journal*, 2020.