ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 3, March 2023

A Comprehensive Study on Detection of Cyber-Attack using ML Techniques & Future Scope

[1] Reeta Mishra, [2] Dr. Neelu Chadhuary

[1] Research Scholar, Manav Rachna University, Faridabad, State Private University (under UGC)
[2] Associate Professor, Manav Rachna University, Faridabad, State Private University (under UGC)
Corresponding Author Email: [1] ree76shok@gmail.com, [2] Neelu@mru.edu.in

Abstract*— In cyber security, machine learning is becoming increasingly significant. The main goal of implementing machine learning in cyber security is to improve malware detection over conventional methods, which rely on human interaction, by making it more actionable, scalable, and efficient. Machine learning problems in the cyber security field call for effective methodical and theoretical management. Deep learning, support vector machines, and Bayesian classification, among other machine learning and statistical techniques, have all shown promise in resisting cyber-attacks. To create intelligent security systems, it is essential to identify hidden trends and insights in network data and to build a corresponding data-driven machine-learning model to stop these attacks. It has been highlighted how machine learning techniques have been used to reduce cyber security dangers that are now present. It has also been discussed how these cutting-edge models have drawbacks and how types of attack patterns have changed over the past ten years. Our objective is to evaluate how well these machine learning methods defend against the growing hazard of malware that affects the world in our online community.

***Index Terms*—Cyber Security; Attacks type, Machine Learning; Classification; Supervised & unsupervised learning.**

I. INTRODUCTION

The global world is currently living in a data-driven era Sarker et. al. (2020), said in which everything is digitally recorded and connected to a data source. The Internet of Things (IoT) data, cyber security data, smart city data, business data, smartphone data, social media data, health data, COVID-19 data, and many other types of data, for instance, are abundant in today's electronic environment. The amount of structured, semi-structured, and unstructured data is always growing. Building a variety of intelligent applications in the pertinent fields can be done using the insights that can be extracted from these data. For instance, the relevant mobile data can be utilized to create tailored context-aware smart mobile applications, or the appropriate cyber security data Sarker et. al. (2020) can be used to create a data-driven automated and intelligent cyber security system. Evaluate intrusion detection systems in light of cutting-edge innovations including cloud computing, edge computing, network virtualization, autonomous tractors, drones, the internet of things, industrial agriculture, and smart grids. Shaukat et. al.(2020) offer a thorough classification of intrusion detection systems in each developing technology based on the machine learning technique we utilized.Therefore, it is vitally necessary to develop data management tools and techniques that can quickly and intelligently extract insights or usable knowledge from the data, which will serve as the foundation for real-world applications.

In the context of data analysis and computing, according to Sarker et. al. (2021) artificial intelligence (AI), and in particular machine learning (ML), have expanded quickly in recent years. These technologies often enable applications to perform intelligently. A branch of artificial intelligence known as machine learning and deep learning has emerged as one of the most desirable employment choices. McIntosh et. al.(2018.2019), ML is normally referred to as the most popular and up-to-date technology in the detection of malicious code or threats since it gives systems the capacity to learn from behavior analysis , the experience of the datasets, and improve the performance without explicitly being coded. Sarker et. al. (2021) , explained that Utilizing cutting-edge smart technology, such as machine learning algorithms and exploratory data processing. As a subset of a larger family of machine learning techniques, deep learning was first developed from the artificial neural network that may be used to effectively evaluate data . Machine learning algorithms are therefore essential for effectively analyzing the behavior of the data set and creating related real-world applications.

Cyber security can be defined as V. Ambalavanan et. al. (2020) Information confidentiality, availability, and integrity must all be guaranteed via a solid, secure computer system. When an unauthorized person, program, or illegal entry into a computer or network with the intent to cause harm or interfere with regular operations compromises the integrity and security of the system, for sure the computer system is at risk. To overcome such situations various safeguards or measures are already available in the market. There are safeguards against cyber attacks at the software, network, host, and data levels. T. Thomas et. al. (2020)There are numerous solutions that operate in isolation to thwart assaults and spot security breaches, including firewalls, antivirus software, intrusion detection systems (IDSs), and intrusion

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 3, March 2023

protection systems (IPSs). Nevertheless, a lot of people continue to have an advantage because they only need to discover one weakness in the systems that require security. The attack surface grows along with the number of systems linked to the internet, raising the chance of an assault. Additionally, attackers are growing more skilled, creating malware that circumvents security mechanisms and zero-day exploits that allow them to remain for extended periods of time undetected.

Contributions: In this review paper, we build upon the existing literature of applications of ML models in cyber security and provide a comprehensive review of ML techniques in cybersecurity. The following are the contributions to this study:

(1) The best of our knowledge, we have made the attempt to provide a comparison of commonly used ML models in cyber security.
(2) Unlike other review papers, we have reviewed applications, datasets, and on use of supervised learning algorithms of ML models to common cyber threats which are malware detection, spam detection, and malware detection.
(3) We have comprehensively compared each classifier's performance based on frequently used datasets.
(4) We have listed the critical challenges of using machine learning techniques in the cyber security domain.

This review paper is organized as follows: Section 2 describes an overview of cyber security and threats, commonly used security datasets. Section 3, basics of machine learning, and classification with applications. Section 3 provides a comprehensive comparison of frequently used ML mechanisms based on different cyber threats and datasets. Section 4 concludes this study and points out the importance of various ML models in cybersecurity.

II. BACKGROUND

A cyber-attack is a hostile, unauthorized third-party system or network access. Its goal is to steal or delete sensitive data from a network connection, information system, and personal device. The perpetrator of this cyber attack is referred to as a hacker.Cyber security is a grouping of laws, methods, tools, and procedures that cooperate to defend the availability, confidentiality, and integrity of computer systems, networks, software, and data from intrusion. Today, according to Sarker IH et. al. (2021) many organizations are constantly facing sophisticated and aggressive cyber threats in today's ever-changing threat landscape. More organized, well-funded, and skilled than ever before are cyber attackers or criminals. Zero-day exploits are assaults that have never been seen before but are frequently alterations of well-known methods. colonization of attack techniques, which enables quick distribution without requiring knowledge of how to create exploits, exacerbates the issue.A computer system and network's configuration and implementation have internal and inherent flaws that lead to vulnerabilities that make it vulnerable to threats and cyberattacks.

2.1 Cyber-Attack types and their impact:

Today, technology rules the globe. Since the industrial revolution, numerous contemporary creations that have altered lifestyles have been created. The newest development in technology is the usage of computers. Computers have developed from complex, large-scale machines to simple, responsive tools that anybody can use. In addition, gadgets made communication easier when used with the Internet. In modern society, the role of the Internet and computers is widely acknowledged. M. Uma et. al. (2013) Wired networks and wires that transport messages to and from the Internet have been used to establish an artificial cyber warfare communication network. As more experience was transferred to it, this field slowly grew.Threats like this are known as cyber breaches. Such assaults are used to spread misinformation, tactically disable services, access personal information, keep track of people, steal records, and cause financial harm. M. Kashif et. al. (2018) these dangers have evolved over time in terms of style, sophistication, and longevity.Vulnerabilities in creating a computer network system include incorrect configuration, inadequate procedures, and unskilled or inexperienced staff. These flaws enhance the likelihood of threats and attacks coming from both inside and outside of a network. People from a wide range of professions are increasingly depending on cyber networks. An agent that alters the operations and behavior of a network or a computer via a specific penetration technique is referred to as a threat.The ultimate goal of cyber security is to Protect data, networks, and programs from cyber threats.In given fig.1, try to elaborate on the broad classification of cyber security.In this various types of attacks are classified on the bases of specific parameters or modes.

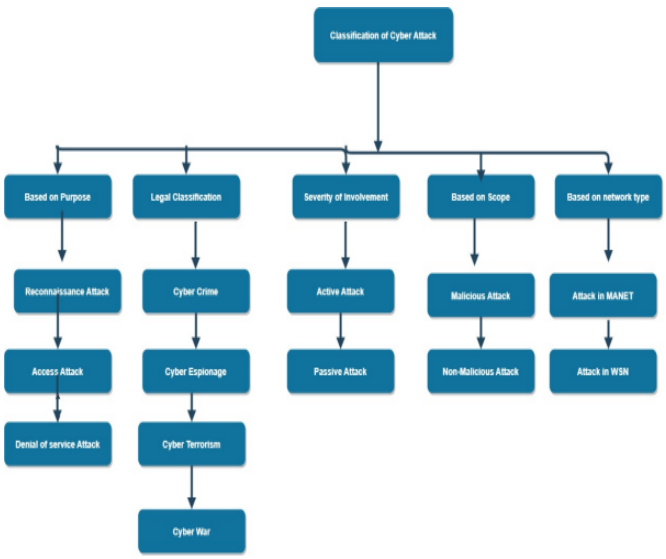

Fig. 1: Broad classification of cyber attack ,Sujayraj et. al. (2019)

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 3, March 2023

The ability to access a computer when the attacker does not have a login or account is given by the unauthorized attacker. Anyone without control authorization has the ability to compromise the data or develop software that makes use of a targeted or exploited coding flaw. In order to get illegal access to user accounts, personal databases, and other sensitive information, established bugs can influence security systems, FTP (File Transfer Protocol) services, and web servers. Sujayraj et. al. (2019),access was subject to a variety of attacks, including secret code attacks, use of the confidence port, channel redirection, socio-technical attacks, attacks based on the environment and severity, and many others.

Table. 1: Attacks name with example

S.N	Classification Type	Name of Attacks	Example
1	**Based on purpose**	Reconnaissance Attack	Packet Sniffers Port Scanning Ping Sweeps DNS Queries
		Access Attack	Port trust Utilization Port redirection Dictionary Attack Man-In-The-Middle Attack Social Engineering Attacks Phishing
		Denial of Service	Smurf SYN Flood DNS Attack DDOS
2	**Legal Classification**	Cyber Crime	Identity Theft Credit card Fraud
		Cyber Espionage	Tracking Cookies RAT Controllable
		Cyber Terrorism	Crashing the powder grids via a network Poisoning of the water supply.
		Cyber War	Russia's war on Estonia (2007) Russia's war on Georgia (2008)
3	**Severity of Involvement**	Active Attack	Masquerade Reply Modification of message
		Passive Attack	Traffic Analysis Release of message contents
4	**Based on Scope**	Malicious Attack	Sasser Attack
		Non-Malicious Attack	Registry Corruption Accident Erasing of hard disk
5	**Based on Network Type**	Attacks in MANET	Byzantine Attack Black hole Attack Flood Rushing Attack Wormhole Attack
		Attack in WSN	Application Layer Attack Transport Layer Attack Network Layer Attack Multi-Layer Attack

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 3, March 2023

Now, in tabular form, we discuss the most common type of attack that occurred in the last few years, particularly after 2019. Some of the common types of cyber-attacks along its description which occurred beyond COVID-19 onwards:

Table. 2: Types of cyber attacks occurred beyond 2019

S.No	Type of Attacks
1	**Malware**
2	**Phishing-**
3	**Man-in-the-middle attack (MITM)**
4	**Distributed Denial-of-Service (DDoS) attack**
5	**SQL injection**
6	**Zero-day exploit**
7	**DNS Tunnelling**
8	**Business Email Compromise (BEC)**
9	**Cryptojacking**
10	**Drive-by Attack**
11	**Cross-site scripting (XSS) attacks**
12	**Password Attack**
13	**Eavesdropping attacks**
14	**AI-Powered Attacks**
15	**IoT-Based Attacks**

2.2.Protection Mechanism from attacks

Traditional, well-known security measures such as anti-virus software, firewalls, access control, data encryption, and cryptography systems are examples of that may not be as effective as needed in the cyber business today. We highlighted some of the protection mechanisms based on AI and ML that need to be in place to reduce the risk of such attacks. Protection mechanisms [Jain AK et. al. (2022)] such as (a) Multi-pronged approach ,(b) Anti-phishing add-ons,(c) Rotating passwords regularly,d) Using a next-generation firewall or Intrusion Prevention System (IPS),(e) Don't ignore updates,(f) Using a VPN (f) virtual private network, (g) Properly sanitized inputs (h) Use Next-Generation Antivirus (NGAV) solutions, (i) Use specialized tools, such as TunnelGuard, Zscaler, and DNSFilter, (j) Monitor the CPU usage of all network devices, including any cloud-based infrastructure.,(k) Install an ad-blocker, or use a privacy/security-focused web browser. (l) Use Multi-Factor Authentication (MFA),(m) Use an intrusion prevention solution to monitor your network for suspicious traffic and reject any packets with spoofed addresses.(n) Good password hygiene, robust access controls, network monitoring, and all of the other solutions.(o) Change the default router settings, use a strong and unique password, and disconnect IoT devices.(p)AI-enabled threat detection systems can predict new attacks and notify admins of any information breach,(q) Agile technique to construct greater secure software program in each aspect.(r)Use of cloud applications such as Google or Microsoft. (s) Follow New strict measures General Data Protection Regulation (GDPR).

III. MACHINE LEARNING IMPORTANCE

AI-based systems perform better than conventional ones,according to Shaukat et. al. (2020) when identifying and retaliating to attacks. Regarding error rate, efficiency, and response to a cyberattack, the AI-based system performs much better than conventional threat-detecting techniques. AI-based systems have a lower mistake rate than conventional systems when it comes to both detecting and reacting to an attack. Mirlekar et. al. (2022) in terms of error rate, the accuracy of attack prediction, and the number of false positives, AI-based solutions also shorten the time needed for the analysis of network flaws, their correction, and the patching of malware-infected networks.

Machine learning is the best choice in the current landscape because of a development, analysis, and implementation process that results in the establishment of a systematic procedure and is related to the field of datasets. The use of datasets or Big Data gives devices the ability to solve challenging problems.This presents a chance to examine and draw attention to any connections among two or more specific occurrences, as well as to forecast the various effects of those correlations. Because models may independently adapt when they are exposed to fresh data, the iterative nature of ML is intriguing. They gain knowledge from prior calculations to provide reliable, repeatable judgments and outcomes. Today, in many spheres of life, including finance, banking, and healthcare, the use of machine learning techniques and AI approaches is expanding quickly.Such as in the fields of education (AlDaajeh et. al. and Khan et. al. (2022)), medicine, and healthcare (AlZubi et. al. and Adil et. al. (2021)), as well as manufacturing (Corallo et. al.and Kayan et. al. (2022)) and, especially, cyber security.Still, many updates are required for better and more accurate outcomes.To address these security difficulties, academics are currently concentrating on the urgent need for new automated security methods. Utilizing automated machine learning approaches to identify new and previously unidentified cyber threats is one of the greatest and most successful practices currently being used.

The aim of this article is to assess the most significant machine learning techniques applied to cyber security and to pinpoint the expanding pattern of such application. We've provided a brief review of machine learning techniques and also how they can be applied to identify and classify risks on computer networks and smart devices, such as intrusion detection, virus detection, spam detection, and traffic analysis.

Actually , How effective and efficient a machine learning solution is developed generally defines the source and quality of the data, as well as the effectiveness of the learning algorithms. The aim of various learning algorithms varies,

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 3, March 2023

and even the results of several learning algorithms in a similar category can differ depending on the qualities of the data. As a result, it's critical to comprehend the fundamentals underlying different machine learning algorithms and how they can be used in a variety of real-world contexts, including IoT systems, cyber security services, business and recommendation systems, smart cities, healthcare, COVID-19, context-aware systems, sustainable agriculture, and many more.

3.1. Machine Learning Development Techniques

Machine learning models come in a variety of forms. In this paper, we try to carry out a task, each model offers precise instructions. Machine learning can quickly identify patterns in millions of data sets. While the basics of machine learning are well understood, practically, little is known about the various types of machine learning models. Experts employ this technique for data analysis to automate the development of analytical models. Mohammed et. al. (2016), machine learning systems are constantly changing and learning from data, finding patterns, and offering insightful information with little involvement from humans. The foundation of machine learning applications and programs that businesses employ for their technology initiatives is made up of algorithms and models.

Main models can be categorized in general under four different methods. The various methods utilized in machine learning development will be highlighted in this paper, along with the best examples of machine learning models and the algorithms that allow the execution of applications for gaining insights from data. For the early detection and forecasting of various assaults, ML approaches are essential in many areas of cyber security such as spam classification, malware detection. fraud detection, phishing, deep websites, IDS, and many more.Now-a-days ,the majority of ML/DL techniques are hybrids, which offer higher results in detection. Xin et. al. (2018) by using hybrid detection, recognized incursions are found more often while unknown attacks receive fewer false positives.

In the area of data science, artificial intelligence, and machine learning, researchers use various widely used datasets for different purposes. Such as, cyber security datasets such as NSL-KDD [Tavallaee et. al. (2009)], ISCX'12 [Canadian institute of cybersecurity, university of new brun swick, (2019)], Bot-IoT Koroniotis et. al. (2019), CICDDoS2019 , UNSW-NB15 [Moustafa et. al. (2015)], etc., smart phone datasets such as phone call logs [Santi et. al. (2011), Sarker et. al. (2016)], SMS Log, mobile application usages logs [Srinivasan et. al. (2014),Zhu et. al. (2014)], mobile phone notification logs [11], etc., IoT data [Balducci et. al. (2018), Lade et. al. (2017), Khadse et. al. (2018)], agriculture and e-commerce data [Tsagkias et. al. (2021), Zikang et. al. (2020)], health data such as heart disease [Safdar et. al. (2018)] , diabetes mellitus [Perveen et. al. (2018), Sarker et. al. (2021)], COVID-19 [Harmon et. al. (2020), Mohamadou et. al. (2020)], ADFA data [Xin et. al. (2018)]etc.CSE-CIC IDS-2018, UNSW-NB15, ISCX-2012, and CIDDS-0018[Kilincer et. al. (2021)], and many more in various application domains. The dataset can be in any of the following categories, and their characteristics may change depending on the application used in the actual world. Various kinds of machine learning approaches can be used, depending on their learning capabilities, to analyze such data in a specific problem domain and to extract insights or relevant information from the data for developing real-world intelligent applications. These techniques are discussed in the following. Four distinct types of learning algorithms: supervised, unsupervised, semi-supervised, and reinforcement learning[Mohamme et. al. (2016)].

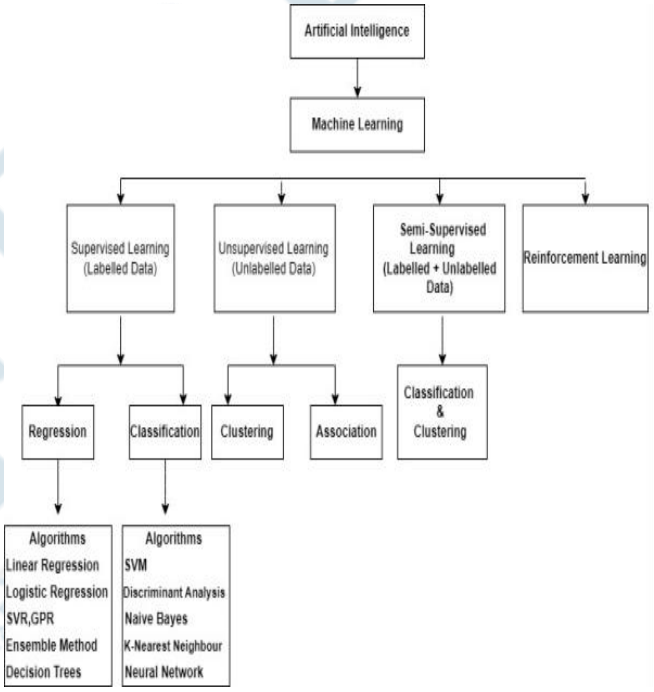

Fig. 2: Various types of machine learning methods.

In the domain of machine learning algorithms, approaches like classification, regression, data clustering, feature engineering, and dimensionality reduction are available to effectively develop data-driven systems. As part of a larger family of machine learning techniques, the artificial neural network (ANN) can be used to effectively examine the data is where deep learning began to take shape[Mohamme et. al. (2016)]. So it can be challenging to choose a learning algorithm that is appropriate for the target application in a given domain. Regarding, this we explain the techniques in detail to reduce the selection challenges. First, **supervised learning** use labeled datasets, and computers are trained for this type of learning, i.e., a task-driven approach [Sarker et. al. (2020)]. Supervised learning used any one way among the two categories: **Classification and Regression.** A classification algorithm is used to tackle the issues when the output may be in binary and/or categorical response. Such as Yes or No, Available or Unavailable, Pink or Blue, etc. are

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 3, March 2023

all possible response options. It is used all around the world to detect spam. and the side, the r**egression** approach is used to resolve issues where the output and input variables have a linear connection. Forecasts for the weather and market conditions are made using regression. In the next second technique, **unsupervised learning,** machines are trained using unlabeled, unclassified datasets, which are then deployed to conclude unstructured data, i.e., a data-driven process [Han et. al. (2011)]. They then make output predictions without oversight or human involvement. This technique is frequently used to group or classify unsorted data according to their characteristics, resemblances, and differences.Unsupervised learning can be split into two types: **The clustering** method helps the computers to group the data based on characteristics, similarities, and differences. Additionally, machines verify object classification and identify intrinsic groups in complex data. This is often used, especially across geographies, to understand client groups and purchasing behavior. **The association** technique is used when massive datasets are offered as input, and machines uncover meaningful interactions and connections between variables. Some researcher arise such questions ,How are the data related to one another?

What steps are involved in mapping variables?

How are these connections profitable?

These are the essential factors to keep in mind when using this learning method. This method is well-liked for checking for plagiarism in research work and mining online activity. Third type, **Semi-Supervised Learning,** this method was developed with consideration for both the advantages and disadvantages of supervised and unsupervised learning. The machines are trained using both labeled and unlabeled datasets during the training phase [Sarker et. al. (2020)][Han et. al. (2011)]. In contrast, the majority of input datasets are not categorized in the actual world. The benefit of using all accessible data rather than just information makes this strategy extremely cost-effective. This learning involves a Hybrid type of learning. Two other crucial techniques are : **Self-directed education,** which uses supervised learning algorithms to solve an unsupervised learning problem, the problem is re-framed as a supervised problem. **Multi-instances Learning,** Although it is an issue of supervised learning, the individual examples are not labeled. Instead, data groups or clusters are labeled. Fourth, **Reinforcement Learning,** directly mimics how people learn from data in the daily lives. Labeled data is not a concept in reinforcement learning. Machines can only learn from their environment or experiences. Learning is a trial-and-error process that is feedback-based. The AI investigates the data, takes note of traits, gains knowledge from past mistakes, and enhances its overall performance. When the output is accurate, the AI agent is rewarded. and disciplined if the outcomes are unfavorable. The use of reinforcement learning in multi-agent systems and game theory is very common.There are two different ways to categorize reinforcement learning: **Learning Through Positive Reinforcement and Learning through Negative Reinforcement.**

"The estimated value of the global market in 2027 is $117.19 billion. Machine learning is being embraced at a rapid rate due to its enormous potential to impact enterprises all around the world."

3.2. Between supervised and unsupervised techniques [Dhanaraj et. al. (2020)]:

In tabular form,we try to highlight the comparison between supervised and unsupervised learning in ML. Dhanaraj et. al.(2020),Machine learning algorithms are applied to the data set.In order to uncover features, produce usable output, recognize patterns or make judgement based on the data , draw inferences from real-time streaming data, make their findings available to analysts, also embed their findings directly into business processes.

Table. 3: Difference between ML techniques

S.N	Key	Supervised Techniques	Unsupervised Techniques
1	**Input**	It uses known,trained and labeled data. (Pre categorized data)	It uses unknown and unlabeled data.
2	**Based on**	Task driven	Data driven
3	**Goals**	Goal is to predict outcomes for new data.	Goal is to gather insights from large volumes of new data.
4	**Data set Type**	Using training datasets	Uses just input datasets
5	**Evaluation**	Quantitative or direct	Qualitative or indirect
6	**Complexity**	A simple method for machine learning, typically calculated through the use of programs like R or Python.	It is complex because they need a large data set to produce intended outcomes.
7	**Drawbacks**	This learning models can be time-consuming to train, and the labels for input and output variables require expertise.	This learning methods may have inaccurate results ,so human intervention to better the output variables.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 3, March 2023

8	**Algorithm**	Decision tree,Logistic regression, SVM,Linear regression,Naive Bayes and neural network.	K-means clustering,hierarchical ,Apriori algorithm
9	**Mechanism**	It has a feedback based mechanism	Follows no feedback mechanism
10	**Techniques used**	Prediction	Analysis
11	**Classes**	Known number of classes	Unknown number of classes.
12	**Learning**	Machines Learn explicitly	Machines understand the data (identifies patterns/structure)
13	**Applications**	Spam detection, sentiment analysis, weather forecasting and pricing predictions, among other things.	Recommendation engines, customer personas images , medical imaging.and Anomaly detection,detection of anomaly
14	**Techniques used to resolves problems**	Artificial neural network ->Regression & classification Convolutional neural network-> computer vision Recurrent neural network->time series analysis	Self-organizing maps->Feature extraction Deep BM & Auto Encoders -> Recommendation system

IV. MACHINE LEARNING MECHANISM ON PERFORMANCE BASED

Details study on use of machine learning , Supervised Learning algorithm and mechanism used to detect and minimized the attack in various domains [Wang et. al. (2022)] to [Dhanaraj et. al. (2020)].

Table. 4 : Different Supervised mechanism already applied in malware detection and globally used to reduction it effect in cyber attacks

S.No	Author's Name	Objectives	Data set Source	Mechanism	Result / Accuracy %
1	Vinod Jain et al (2020)	To detect the frauds accomplished using credit cards	data of 284808 credit cards	Decision Tree, Random Forest and XGBoost	99.96%.
2	Osama et al (2020)	To detect anomalies and also identify the variables that caused these anomalies (data for two series of products).	Manufacturing Industry	KNN, ABOD	95.00%.
3	Gethzi et al (2020)	Anomaly detection is a key challenge to ensure the security in WSN.	Intel Berkeley Research Lab (IBRL) from 54 mica sensors	Online Locally Weighted Projection Regression (OLWPR) for anomaly detection , Regression methods, Principal Component Analysis	86.00%
4	Ruttala Sailusha et al (2020)	To detect the fraudulent activities the credit card fraud detection system was introduced.	European credit card company	random forest algorithm and the Adaboost algorithm	99.00%
5	Ritwik Giri et al. , (2020)	A new self-supervised classification framework for anomaly detection in audio signals.	DCASE	Ensembles Method (MoblieNetV2, ResNet-50 and Group-Made models)	improvement of over 12.8% (average AUC metrics), 95.5%

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 3, March 2023

6	T. Ergen and et al.(2017) Agarap et al.(2017)	Introducing a linear support vector machine (SVM) as the replacement for Softmax in the final output layer of a GRU model.	2013 network traffic data	GRU-SVM model	81.54% 84.14%
7	Liu et al., (2020)	Enhance IoT security through the experimentation of multiple machine learning methods on the IoT network intrusion datasets.	IoT Network Intrusion Datasets	LR, SVM, RF, KNN, XGBoost	Accuracy when using KNN algorithms was 99%
8	Al-Akhras et al., (2020) Alrashdi et al., (2019)	Examine various ML algorithms' effectiveness to detect Attacks and anomaly in IoT Networks.	UNSW-NB15	RF, KNN, Naïve bayes	Perform best with RF 100% accuracy and KNN 99% accuracy. AD-IoT 99.34%
9	Khattak et al (2021)	To reduce the amount of sparseness in the data collected from stock market using machine learning	KSE dataset	KNN Classifier	The proposed system's model (KNN classifier) gives better results of low error as compared to previous .
10	Ioannou, et al. (2020)	For anomaly detection techniques to detect abnormal behaviors within the network.	IoTtest bed data	SVM	81.00%
11	Thamaraiselvie t al. (2020)	serviceability issues with ML for detecting anomalies in IoT networks	IoT-23	SVM, RF, Naiive Bayes,Decision tree	RF algorithm achieved the best results with an accuracy of 99.5%
12	Susilo and Sari (2020)	Improving IoT Security by using machine learning techniques	BoT-IoT	Random Forest, CNN, and MLP	Random forests and CNN have recorded the highest results in terms of accuracy
13	Rani, D.; Kaushal (2020)	To secure data, device and IoT network,seeking for secure and accurate Intrusion Detection System (IDS).	KDD-99	c-SVM	99.90%
14	Wan, Y.; Xu (2020)	To introduce IoTArgos, a multi-layer security monitoring system, which collects, analyzes, and featured data communications of heterogeneous IoT devices .	NSL-KDD and KDDCUP99	Random Forest	98.7%%

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 3, March 2023

15	Krishnan, S.; Neyaz (2021)	Perform various supervised feature selection methods and employ three classifiers on IoT network data. The classifiers predict with high accuracy when the network traffic against the IoT device was malicious or benign.	Network traffic data	NN, LR, RF, NB, and KNN	Three logistic regression techniques (SVC, Random Forest, and XGBoost) performed with high accuracy.
16	Morfino, V. et al (2020)	To identifying cyber-attacks (namely SYN-DOS attacks) to IoT systems are compared both in terms of application performances, and in training/application times.	IoTID20	RF, SVC, and XGBoost	Spark algorithms >99%
17	Khonde, S et al (2019)	An intelligent IDS system is presented which classifies the normal traffic in a network with abnormal or attacked ones	SYNDOS2M	RF, DT, LR, SVM, and GBT	increase accuracy of detection by 10%, reduces false positive rate to 0.05
18	HuiLia ,Qi Chen et al (2021)	To implement a new method to study the extraction and classification of online dating services (ODS)'s comments.	4,300 comments of negative/positive emotions published on dating websites	Machine learning and lexicon-based method	Achieve higher accuracy than any type of sentiment analysis.
19	Y. Y. Aung et al (2020)	To analysis network traffic action	KDD cup99	hybrid K means and KNN	99.99% rate of detection,time 0.18
20	C. Liang et al (2019)	K Apache Spark for anomaly detection	ISCX 2012	K Means, decision Tree, Random Forest [8]	RF acc 99.5%, DT acc 93.5%
21	S. Sharma et al (2020)	Analysis E-web application	CSIC HTTP 2010	J48, DT, NB One R	DT J48 good result detection rate of 94.5%
22	V. Ravindranath et al. (2020)	Hacked reath Network attack	WOA	Cloud computing	WOA accuracy 80%
23	Kotpalliwar,Wajgi et al. (2015)	To calculates parameter values related to intrusion-detector performance	Mixed,'10% KDD Cup 99	SVM	89.85% and 99.9%
24	Saxena et. al. (2014)	To analysis IDS	41 mixed attribute	PSO-SVM	89.4%
25	Pervez et. al. (2014)	intrusion detection	NSL-KDD	Support Vector Machine Classifier	99%
26	Yan ,Liu (2018) Kokila et al.(2015)	To create a transductive method and introduces the simulated annealing method to degenerate the optimization model. Focus on DDoS attacks on the SDN controller	DARPA 1998	Support Vector Machine Classifier	80.1% 95.1%

27	Rao et. Al.(2018)	To experiment with various attack types and different k values	12,597 sample	k-Nearest Neighbor(IKPDS)	99.6%
28	Sharifi et al. (2015) B. Ingre et al.(2017) Shapoorifard et al.(2017)	IDS detection To improve the classification performance of KNN in IDS	NSL-KDD	k-Nearest ,correlation feature selection Neighbor(IKPDS)	90% 90.3%
29	Relan et. al. (2015)	To analysis feature selection	KDD Cup 99 and NSL-KDD	C4.5 (with pruning), C4.5 decision tree	98.45%
30	Azad et. al. (2015)	To solves the problem of small separation in the decision tree, improves the accuracy of classification, and reduces the false positive rate.	Random Tree, Naïve Bayes and Reptree	C4.5 decision tree and the genetic algorithm	99.8%

V. CONCLUSION

In the present cyber environment, especially for cyber security, machine learning techniques have emerged as its most fundamental component. Both the defending and the attacker sides are using machine learning techniques. In order to bypass and avoid the security system and firewall, the attacker's side uses machine learning techniques. These methods aid security professionals in defending against unauthorized access and illegal penetration of security systems. In order to identify cyber security concerns, machine learning techniques are compared and contrasted in this study. The three main hazards to the internet that we have taken into account are virus, spam, and intrusions.We have contrasted machine learning models: Deep belief network, Decision tree, Naive Bayes, Random forest, and Support vector machine. Each cyber threat has a unique set of sub-domains. The sub-domains of intrusion detection are anomaly-based, signature-based, and hybrid-based. The sub-domains for malware detection either are static detection, dynamic detection, or hybrid detection. In order to categories spam such photos, videos, emails, SMS, or calls, sub-domains for spam are the medium through which the models are implemented.Furthermore, in order to handle hostile inputs, robust machine learning models are required. To build resilient models against hostile inputs, there should be a focus on training the model in adversarial conditions. We have examined machine learning models for detecting cyber threats based on various datasets, but we recommend a novice in this field to explore the comprehensive bibliography provided in this review paper. Future research will examine more ML and DL methods used to counter many other cyber security threats. We will assess the ML models in other cyber security domains, including IoT, smart cities, API-based approaches, cellular networks, and smart grids.

REFERENCES

[1] McIntosh, T.; Jang-Jaccard, J.; Watters, P.; Susnjak, T. The inadequacy of entropy-based ransomware detection.In Proceedings of the International Conference on Neural Information Processing, Sydney, Australia, 12–15 December 2019; Springer: Berlin/Heidelberg, Germany,2019; pp. 181–189.

[2] McIntosh, T.R., Jang-Jaccard, J., Watters, P.A.: Large scale Behavioural analysis of ransomware attacks. In: Cheng, L.,Leung, A.C.S., Ozawa, S. (eds.) ICONIP 2018. LNCS, vol. 11306,pp. 217–229. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-04224-0_19

[3] Sarker IH, Hoque MM, MdK Uddin, Tawfeeq A. Mobile data science and intelligent apps: concepts, ai-based modeling and research directions. Mob Netw Appl, pages 1–19, 2020.

[4] Canadian institute of cybersecurity, university of new brunswick, iscxdataset, http://www.unb.ca/cic/datasets/index.html/ (Accessed on 20 October 2019).

[5] Cic-ddos2019[online].available: https://www.unb.ca/cic/datasets/ddos-2019.html/ (Accessed on 28 March 2020).

[6] Sarker IH. A machine learning based robust prediction model for real-life mobile phone data. Internet Things. 2019;5:180–93.

[7] Sarker IH. Ai-driven cybersecurity: an overview, security intelligence modeling and research directions. SN Comput Sci. 2021.

[8] Sarker IH, Kayes ASM, Badsha S, Alqahtani H, Watters P, Ng A. Cybersecurity data science: an overview from machine learning perspective. J Big Data. 2020;7(1):1–29.

[9] Koroniotis N, Moustafa N, Sitnikova E, Turnbull B. Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: bot-iot dataset. Fut Gen Comput Syst. 2019;100:779–96.

[10] Mehrotra A, Hendley R, Musolesi M. Prefminer: mining user's preferences for intelligent mobile notifcation management. In:Proceedings of the International Joint

Conference on Pervasive and Ubiquitous Computing, Heidelberg, Germany, 12–16 September, 2016; pp. 1223–1234. ACM, New York, USA.

[11] Moustafa N, Slay J. Unsw-nb15: a comprehensive data set for network intrusion detection systems (unsw-nb15 network data set). In: 2015 military communications and information systems conference (MilCIS), 2015;pages 1–6. IEEE .

[12] Tavallaee M, Bagheri E, Lu W, Ghorbani AA. A detailed analysis of the kdd cup 99 data set. In. IEEE symposium on computational intelligence for security and defense applications. IEEE. 2009;2009:1–6.

[13] Santi P, Ram D, Rob C, Nathan E. Behavior-based adaptive call predictor. ACM Trans Auton Adapt Syst. 2011;6(3):21:1–21:28.

[14] Sarker IH, Colman A, Kabir MA, Han J. Phone call log as a context source to modeling individual user behavior. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (Ubicomp): Adjunct, Germany, pages 630–634. ACM, 2016.

[15] Srinivasan V, Moghaddam S, Mukherji A. Mobileminer: mining your frequent patterns on your phone. In: Proceedings of the International Joint Conference on Pervasive and Ubiquitous Computing, Seattle, WA, USA, 13-17 September, pp. 389–400. ACM, New York, USA. 2014.

[16] Zhu H, Chen E, Xiong H, Kuifei Y, Cao H, Tian J. Mining mobile user preferences for personalized context-aware recommendation. ACM Trans Intell Syst Technol (TIST). 2014;5(4):58.

[17] Balducci F, Impedovo D, Pirlo G. Machine learning applications on agricultural datasets for smart farm enhancement. Machines. 2018;6(3):38.

[18] Sarker, I. H. "Machine Learning: Algorithms, Real-World Applications and Research Directions. SN ComputSci2,160."(2021) https://link.springer.com/article/10.1007/s42979-021-00592-x

[19] Safdar S, Zafar S, Zafar N, Khan NF. Machine learning based decision support systems (dss) for heart disease diagnosis: a review. Artif Intell Rev. 2018;50(4):597–623.

[20] Perveen S, Shahbaz M, Keshavjee K, Guergachi A. Metabolic syndrome and development of diabetes mellitus: predictive modeling based on machine learning techniques. IEEE Access.2018;7:1365–75.

[21] Khadse V, Mahalle PN, Biraris SV. An empirical comparison of supervised machine learning algorithms for internet of things data. In: 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), IEEE.2018; 1–6.

[22] Mohammed M, Khan MB, Bashier Mohammed BE. Machine learning: algorithms and applications. CRC Press; 2016.

[23] Emeritus organization: https://emeritus.org/blog/types-of-machine-learning/

[24] Lade P, Ghosh R, Srinivasan S. Manufacturing analytics and industrial internet of things. IEEE Intell Syst. 2017;32(3):74 - 9.

[25] Mohamadou Y, Halidou A, Kapen PT. A review of mathematical modeling, artifcial intelligence and datasets used in the study, prediction and management of covid-19. Appl Intell ,2020;50(11):3913–25.

[26] Tsagkias M. Tracy HK, Surya K, Vanessa M, de Rijke M. Challenges and research opportunities in ecommerce search and recommendations. In: ACM SIGIR Forum. volume 54. NY, USA: ACM New York; 2021. p. 1–23.

[27] Zikang H, Yong Y, Guofeng Y, Xinyu Z. Sentiment analysis of agricultural product ecommerce review data based on deep learning. In: 2020 International Conference on Internet of Things and Intelligent Applications (ITIA), IEEE, 2020; pages 1–7

[28] Han J, Pei J, Kamber M. Data mining: concepts and techniques.Amsterdam: Elsevier; 2011.

[29] Harmon SA, Sanford TH, Sheng X, Turkbey EB, Roth H, Ziyue X, Yang D, Myronenko A, Anderson V, Amalou A, et al. Artifcial intelligence for the detection of covid-19 pneumonia on chest ct using multinational datasets. Nat Commun. 2020;11(1):1–7.

[30] Sarker IH. Deep cybersecurity: a comprehensive overview from neural network and deep learning perspective. SN Comput Sci. 2021

[31] Zheng T, Xie W, Xu L, He X, Zhang Y, You M, Yang G, Chen Y. A machine learning-based framework to identify type 2 diabetes through electronic health records. Int J Med Inform. 2017;97:120–7.

[32] Md Sahrom Abu, Siti Rahayu Selamat,, Aswami Ariffin,, Robiah Yusof4,"Cyber Threat Intelligence – Issue and Challenges",Indonesian Journal of Electrical Engineering and Computer Science,Vol. 10, No. 1, April 2018, pp. 371~379ISSN:2502-4752,DOI: 10.11591/ijeecs.v10.i1.pp371-379.

[33] Sujayraj, S. “Classification of Cyber Attacks and its Associated Laws.” Journal of emerging technologies and innovative research 6 (2019): 289-298-289-298.

[34] Uma, M., and Ganapathi Padmavathi. "A Survey on Various Cyber Attacks and their Classification." *Int. J. Netw. Secur.* 15, no. 5 (2013): 390-396.

[35] M. Kashif, S. A. Malik, M. T. Abdullah, M. Umair, and P. W. Khan, “A systematic review of cyber security and classification of attacks in networks,” International Journal of Advanced Computer Science and Applications, vol. 9, no. 6, pp. 201–207, 2018, doi: 10.14569/IJACSA.2018.090629.

[36] Shaukat, Kamran, Suhuai Luo, Vijay Varadharajan, Ibrahim A. Hameed, Shan Chen, Dongxi Liu and Jiaming Li. “Performance Comparison and Current Challenges of Using Machine Learning Techniques in Cybersecurity.” Energies 13 (2020): 2509.

[37] Mirlekar, Swati and Komal Prasad Kanojia. “Role of Intrusion Detection System in Network Security and Types of Cyber Attacks-A Review.” International Journal of Innovations in Engineering and Science (2022): n. page.

[38] Shaukat, Kamran, Suhuai Luo, Vijay Varadharajan, Ibrahim A. Hameed, Shan Chen, Dongxi Liu and Jiaming Li. “Performance Comparison and Current Challenges of Using Machine Learning Techniques in Cybersecurity.” Energies 13 (2020): 2509.

[39] Mirlekar, Swati and Komal Prasad Kanojia. “Role of Intrusion Detection System in Network Security and Types of Cyber Attacks-A Review.” International Journal of Innovations in Engineering and Science (2022)

[40] Li, Chenggang, Tao Lin, and Zhenci Xu. "Impact of Hydropower on Air Pollution and Economic Growth in China." *Energies* 14, no. 10 (2021): 2812.

[41] Jain, Ankit Kumar, and B. B. Gupta. "A survey of phishing attack techniques, defence mechanisms and open research

challenges." *Enterprise Information Systems* 16, no. 4 (2022): 527-565.

[42] Kilincer, Ilhan Firat, Fatih Ertam, and Abdulkadir Sengur. "Machine learning methods for cyber security intrusion detection: Datasets and comparative study." Computer Networks 188 (2021): 107840.

[43] Vinod Jain,Mayank Agrawal,Anuj Kumar,"Performance Analysis of Machine Learning Algorithms in Credit Cards Fraud Detection",2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)Amity University, Noida, India. June 4-5, 2020

[44] O. Abdelrahman and P. Keikhosrokiani, "Assembly Line Anomaly Detection and Root Cause Analysis Using Machine Learning," in IEEE Access, vol. 8, pp. 189661-189672, 2020, doi: 10.1109/ACCESS.2020.3029826.

[45] I. Gethzi Ahila Poornima, B. Paramasivan,"Anomaly detection in wireless sensor network using machine learning algorithm", ComputerCommunications, Volume151,2020, Pages 331-337, https://doi.org/10.1016/j.comcom.2020.01.005.

[46] R. Sailusha, V. Gnaneswar, R. Ramesh and G. R. Rao, "Credit Card Fraud Detection Using Machine Learning," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), 2020, pp. 1264-1270, doi: 10.1109/ICICCS48265.2020.9121114.

[47] Ritwik Giri, Srikanth V. Tenneti , Fangzhou Cheng, Karim Helwani, Umut Isik, Arvindh Krishnaswamy,"SELF-SUPERVISED CLASSIFICATION FOR DETECTING ANOMALOUS SOUNDS" , https://assets.amazon.science/8f/33/04709ab4460da4af7f80528ab61c/self-supervised-classification-for-detecting-anomalous-sounds.pdf

[48] D. -H. Shin, R. C. Park and K. Chung, "Decision Boundary-Based Anomaly Detection Model Using Improved AnoGAN From ECG Data," in IEEE Access, vol. 8, pp. 108664-108674, 2020, doi: 10.1109/ACCESS.2020.3000638.

[49] Bagui S, X. Wang, S. Bagui, "Machine Learning Based Intrusion Detection for IoT Botnet," International Journal of Machine Learning and Computing. 2021; 11.

[50] Mary DRTaSAS. "Attack and Anomaly Detection in IoT Networks using Machine Learning," Int. J. Comput. Sci. Mob.Comput. 2020;9:95–103.

[51] Wang T, M. Z. A. Bhuiyan, G. Wang, L. Qi, J. Wu, T. Hayajneh, "Preserving balance between privacy and data integrity in edgeassisted Internet of Things," IEEE Internet of Things Journal. 2019;7:2679-2689.

[52] Liu Z, Thapa N, A. Shaver, K. Roy, X. Yuan, S. Khorsandroo. "Anomaly Detection on IoT Network Intrusion using Machine Learning," in 2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD). 2020;1-5.

[53] Al-Akhras M, M. Alawairdhi, A. Alkoudari, and S. Atawneh, "Using machine learning to build a classification model for IoT networks to detect attack signatures."; 2020.

[54] Alrashdi I, Alqazzaz A, Aloufi E, Alharthi R, Zohdy M, Ming H. "Ad-iot: Anomaly detection of iot cyberattacks in smart city using machine learning," in 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC). 2019;0305-0310.

[55] Khattak, A. et al. (2022). An Efficient Supervised Machine Learning Technique for Forecasting Stock Market Trends. In: Guarda, T., Anwar, S., Leon, M., Mota Pinto, F.J. (eds) Information and Knowledge in Internet of Things. EAI/Springer Innovations in Communication and Computing. Springer, Cham. https://doi.org/10.1007/978-3-030-75123-4_7

[56] Sarkar, T., Mukherjee, A. & Chatterjee, K. Supervised Learning Aided Multiple Feature Analysis for Freshness Class Detection of Indian Gooseberry (Phyllanthus emblica). J. Inst. Eng. India Ser. A 103, 247–261 (2022). https://doi.org/10.1007/s40030-021-00585-2

[57] Ioannou, C.; Vassiliou, V. Experimentation with local intrusion detection in IoT networks using supervised learning. In Proceedings of the 16th International Conference on Distributed Computing in Sensor Systems (DCOSS), Marina del Rey, CA,USA, 25–27 May 2020; pp. 423–428.

[58] Rani, D.; Kaushal, N.C. Supervised machine learning based network intrusion detection system for internet of things. In Proceedings of the 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT),Kharagpur, India, 1–3 July 2020; pp. 1–7.

[59] Wan, Y.; Xu, K.; Xue, G.; Wang, F. Iotargos: A multi-layer security monitoring system for internet-of-things in smart homes. In Proceedings of the IEEE INFOCOM 2020-IEEE Conference on Computer Communications, Toronto, ON, Canada, 6–9 July 2020;pp. 874–883.

[60] Krishnan, S.; Neyaz, A.; Liu, Q. IoT network attack detection using supervised machine learning. Int. J. Artif. Intell. Expert Syst.2021, 10, 18–32.

[61] Morfino, V.; Rampone, S. Towards near-real-time intrusion detection for IoT devices using supervised learning and APACHE Spark. Electronics 2020, 9, 444. [CrossRef]

[62] Khonde, S.; Ulagamuthalvi, V. Ensemble-based semi-supervised learning approach for a distributed intrusion detection system.J. Cyber Secur. Technol. 2019, 3, 163–188. [CrossRef]

[63] Sujayraj, S. “Classification of Cyber Attacks and its Associated Laws.” Journal of emerging technologies and innovative research 6 (2019): 289-298-289-298.

[64] C. Liang, B. Shanmugam, S. Azam, M. Jonkman, F.De Boer, and G. Narayansamy, “Intrusion Detection System for Internet of Things based on a Machine Learning approach,” Proc. - Int. Conf. Vis. Towar. Emerg. Trends Commun. Networking,ViTECoN 2019, pp. 1–6, 2019

[65] S. Sharma, P. Zavarsky, & S. Butakov, “Machine Learning based Intrusion Detection Web-Based Attacks,” Proc. - 2020 IEEE 6th Intl Conference

[66] V. Ravindranath, S. Ramasamy, R. Somula, K. S. Sahoo, and A. H. Gandomi, “Swarm Intelligence Based Feature Selection for Intrusion and Detection System in Cloud Infrastructure,” 2020 IEEE Congr. Evol. Comput. CEC 2020 - Conf. Proc., pp. 1–6,2020.

[67] Shaukat, Kamran, Suhuai Luo, Shan Chen, and Dongxi Liu. "Cyber threat detection using machine learning techniques: A performance evaluation perspective." In 2020 International Conference on Cyber Warfare and Security (ICCWS), pp. 1-6. IEEE, 2020.

[68] M. V. Kotpalliwar and R. Wajgi, ‘‘Classification of attacks using support vector machine (SVM) on KDDCUP’99 IDS database,’’ in Proc. Int. Conf.Commun. Syst. Netw. Technol., 2015, pp. 987–990

[69] R. T. Kokila, S. T. Selvi, and K. Govindarajan, ‘‘DDoS detection and analysis in SDN-based environment using support vector machine classifier,’’ in Proc. 6th Int. Conf. Adv. Comput., 2015, pp. 205–210.
[70] B. B. Rao and K. Swathi, ‘‘Fast kNN classifiers for network intrusion detection system,’’ Indian J. Sci. Technol., vol. 10, no. 14, pp. 1–10,2017
[71] A. M. Sharifi, S. A. Kasmani, and A. Pourebrahimi, ‘‘Intrusion detection based on joint of K-means and KNN,’’ J. Converg. Inf. Technol., vol. 10,no. 5, pp. 42–51, 2015.
[72] H. Shapoorifard and P. Shamsinejad, ‘‘Intrusion detection using a novel hybrid method incorporating an improved KNN,’’ Int. J. Comput. Appl.,vol. 173, no. 1, pp. 5–9, 2017.
[73] N. G. Relan and D. R. Patil, ‘‘Implementation of network intrusion detection system using variant of decision tree algorithm,’’ in Proc. Int. Conf.Nascent Technol. Eng. Field, 2015, pp. 1–5.
[74] C. Azad and V. K. Jha, ‘‘Genetic algorithm to solve the problem of small disjunct in the decision tree based intrusion detection system,’’ Int.J. Comput. Netw. Inf. Secur., vol. 7, no. 8, pp. 56–71, 2015.
[75] K. Alrawashdeh and C. Purdy, ‘‘Toward an online anomaly intrusion detection system based on deep learning,’’ in Proc. IEEE Int. Conf. Mach.Learn. Appl., Dec. 2017, pp. 195–200.
[76] Xin, Yang, Lingshuang Kong, Zhi Liu, Yuling Chen, Yanmiao Li, Hongliang Zhu, Mingcheng Gao, Haixia Hou, and Chunhua Wang. "Machine learning and deep learning methods for cybersecurity." Ieee access 6 (2018): 35365-35381.
[77] Shaukat, Kamran, Suhuai Luo, Vijay Varadharajan, Ibrahim A. Hameed, and Min Xu. "A survey on machine learning techniques for cyber security in the last decade." IEEE Access 8 (2020): 222310-222354.
[78] V. Ambalavanan, "Cyber threats detection and mitigation using machine learning" in Handbook of Research on Machine and Deep Learning Applications for Cyber Security, Hershey, PA, USA:IGI Global, pp. 132-149, 2020.
[79] T. Thomas, A. P. Vijayaraghavan and S. Emmanuel, "Machine learning and cybersecurity" in Machine Learning Approaches in Cyber Security Analytics, Singapore:Springer, pp. 37-47, 2020, [online] Available: https://link.springer.com/chapter/10.1007/978-981-15-1706-8_3.
[80] S. Emerson, R. Kennedy, L. O’Shea and J. O’Brien, "Trends and applications of machine learning in quantitative finance", Proc. 8th Int. Conf. Econ. Finance Res. (ICEFR), pp. 1-9, 2019.
[81] A. F. Agarap. (2017). ‘‘A neural network architecture combining gated recurrent unit (GRU) and support vector machine (SVM) for intrusion detection in network traffic data.’’ [Online]. Available: https://arxiv.org/abs/1709.03082
[82] T. Ergen and S. S. Kozat, ‘‘Efficient online learning algorithms based on LSTM neural networks,’’ IEEE Trans. Neural Netw. Learn. Syst., vol. 4,no. 2, pp. 1–12, 2017.
[83] Manufacturin-Corallo, Angelo, Mariangela Lazoi, Marianna Lezzi, and Angela Luperto. "Cybersecurity awareness in the context of the Industrial Internet of Things: A systematic literature review." Computers in Industry 137 (2022): 103614.
[84] Kayan, Hakan, Matthew Nunes, Omer Rana, Pete Burnap, and Charith Perera. "Cybersecurity of industrial cyber-physical systems: a review." ACM Computing Surveys (CSUR) 54, no. 11s (2022): 1-35.
[85] AlZubi, Ahmad Ali, Mohammed Al-Maitah, and Abdulaziz Alarifi. "Cyber-attack detection in healthcare using cyber-physical system and machine learning techniques." Soft Computing 25, no. 18 (2021): 12319-12332.
[86] Adil, Muhammad, and Muhammad Khurram Khan. "Emerging iot applications in sustainable smart cities for covid-19: Network security and data preservation challenges with future directions." Sustainable Cities and Society 75 (2021): 103311.
[87] AlDaajeh, Saleh, Heba Saleous, Saed Alrabaee, Ezedin Barka, Frank Breitinger, and Kim-Kwang Raymond Choo. "The Role of National Cybersecurity Strategies on the Improvement of Cybersecurity Education." Computers & Security (2022): 102754.
[88] Gamal, Merna, Hala Abbas, and Rowayda Sadek. "Hybrid approach for improving intrusion detection based on deep learning and machine learning techniques." In *Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020)*, pp. 225-236. Springer International Publishing, 2020.
[89] Khan, Manzoor Ahmed, Adel Merabet, Shamma Alkaabi, and Hesham El Sayed. "Game-based learning platform to enhance cybersecurity education." Education and Information Technologies (2022): 1-25.
[90] Dhanaraj, R.K., Rajkumar, K., Hariharan, U. (2020). Enterprise IoT Modeling: Supervised, Unsupervised, and Reinforcement Learning. In: Haldorai, A., Ramu, A., Khan, S. (eds) Business Intelligence for Enterprise Internet of Things. EAI/Springer Innovations in Communication and Computing. Springer,Cham. Doi.org/10.1007/978-3-030-44407-5_3

ABBREVIATIONS AND ACRONYMS

S.No.	Abbreviation	Description
1	KNN	K-nearest neighbors
2	CNN	Convolutional Neural Network
3	SVM	Support Vector Machine
4	LR,RF	Logistic regression,Random Forest
5	GRU	Gated Recurrent Unit
6	GBT	Gradient boosting
7	SVC	Support Vector Classifier
8	XGBoost	Extreme Gradient Boosting
9	J48,NB	C4.5 algorithm,Naive Bayes
10	ABOD	Angle Based Outlier Detector