IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 6, June 2024

DOI: 10.17148/IJARCCE.2024.13644

A survey on Next-Gen Intrusion Detection System

Netravati Gangappa Gokavi[1], Dr. Pijush Barthakur[2]

Dept. of Master of Computer Application, KLS Gogte Institution of Technology, Affiliated to Visvesvaraya Technological University, Belagavi, India[1]

Dept. of Master of Computer Application, KLS Gogte Institution of Technology, Affiliated to Visvesvaraya Technological University, Belagavi, India[2]

Abstract: The Internet links thousands of millions of computers which are running various software and hardware platforms to provide communication as well as commercial services. But since computers are connected, bad people can abuse resources and launch cyberattacks. Creating adaptable security-focused methods is becoming increasingly difficult due to an increasing number of cyberattacks. One of the most crucial tools for spotting online threats is the intrusion detection system (IDS). Several methods from several fields have been applied in literature to create effective IDS. AI-based techniques are widely used in intrusion detection systems (IDS) development and provide several advantages over alternative methods. To investigate and comprehend the current state of these approaches to address intrusion detection issues, there is not, however, a thorough analysis of AI-based strategies. This work has studied several AI-based methodologies, with a focus on IDS generation.

Keywords- Intrusion Detection System (IDS), Machine Learning, AI Powered, Datasets, Computer Security

I. INTRODUCTION

The security of sensitive data and vital infrastructure has emerged as a major concern in a world that is becoming more interconnected and dependent on technology.

The development of distributed networks and various kinds of access environments is the result of 5G wireless communication technology. As a result, data from a variety of heterogeneous sources, including computers, sensors, and the Internet of Things (IoT), is now communicated through network systems [1]. Because of the evolving nature of cyberattacks, traditional methods of protecting networks and systems have faced significant challenges. A hacking attempt or a threat of a prospective intentional unauthorized attempt to steal data or resources, alter data, or make a system useless is known as an intrusion. A hacker attempting to enter or abuse your system or network is also known as an incursion.

Cybersecurity is a dynamic field that need innovative approaches to address the increasing complexity of cyberthreats. AI now plays a significant role in the battle against these dangers, offering both intriguing benefits and challenging challenges. The sheer size and speed at which cyber threats can materialize renders traditional manual monitoring techniques frequently inadequate. Organizations can utilize AI to scan and interpret massive amounts of data from several sources, including network logs, user behavior, and system events. This allows the organizations to discover abnormalities and potential dangers more quickly and efficiently.

When network security provisions use AI and machine learning to improve protection, this is known as artificial intelligence (AI) for cybersecurity. Due to the rise in remote work and the resulting rise in the usage of Internet-connected services, the attack surface for cyber threats has been growing quickly. The effectiveness of traditional signature-based techniques to counter these attacks has been steadily declining. Organizations are left open to assault because the process of identifying a threat, obtaining user feedback, and creating a signature to neutralize it moves too slowly.

An essential element of cybersecurity is an Intrusion Detection System (IDS), which maintains an eye on and analyses network or system behaviour to detect any possible risks to security [3]. Unusual activity on hosts and networks is detected using a set of methods and tools referred to as intrusion detection. All incoming and outgoing network traffic is analyzed by an intrusion detection system (IDS) to look for any odd patterns that might indicate an attempt to break or infiltrate a system or network through a network or system attack. Installing an intrusion detection system (IDS) in a home or on a network accomplishes comparable goals. Both use several methods to detect the presence of an assailant,

DOI: 10.17148/IJARCCE.2024.13644

burglar, or intruder before sending out an alert or warning. IDS examines the resources of an information system and alerts to any detected malicious activities within the system. In some cases, highly sophisticated IDSs go beyond detection and possess the capability to respond actively to attacks. The proactive measure taken by these advanced IDSs involves blocking access for malicious users or activities, thereby preventing them from compromising computer resources. There exist two primary categories of Intrusion Detection Systems (IDS): Network Intrusion Detection System (NIDS) and Host Intrusion Detection System (HIDS) [5]

Network Intrusion Detection System (NIDS): Located strategically at key junctures in the network, NIDS examines all network traffic to spot possibly dangerous activity. It makes an important improvement to the general security of organizational networks by being essential in identifying attacks that originate from internal hosts.

Host Intrusion Detection System (HIDS): HIDS functions differently from NIDS and requires installation on individual client PCs (hosts) within the network. HIDS, in contrast to NIDS, monitors a single host's activity in addition to analyzing its traffic. HIDS sounds an alarm if it detects unusual activity. One of the most important elements in improving each host's security posture inside the network is HIDS.

It is worth noting that both NIDS and HIDS are integral elements of a comprehensive intrusion detection strategy. While NIDS focuses on network-wide traffic analysis to identify potential threats, HIDS operates at the individual host level, providing a more granular perspective on the security status of each client computer. Combining both NIDS and HIDS offers a layered approach to intrusion detection, strengthening the overall resilience of the network against a diverse range of cyber threats. It uses a variety of detection techniques, including behaviour-based, anomaly-based, and signature-based techniques to identify known as well as unknown threats. Nevertheless, this method may lack the capability to identify emerging or unfamiliar threats, representing a notable constraint. Additionally, the continuous updating of the signature database is imperative to maintain the effectiveness of the Signature Detection (SD) system.

An alternate method to SD is called Anomaly-based Detection (AD), which analyses network data with the goal to spot any unusual patterns of activity that can point to an impending attack. AD utilizes statistical analysis, machine learning algorithms, and other methods to create a baseline of typical behaviour and identify anomalies. This approach works well for identifying unknown or unusual assaults and is flexible enough to change with the way networks behave. While AD is superior to SD in terms of benefits, it has a greater false positive rate and necessitates a training period to establish the baseline of normal behavior. To lessen the detrimental consequences of AD, a hybrid strategy can be used to handle high false positives and low false negatives [3].

It seeks to identify instances of malware activity, denial-of-service attacks, illegal access, and other security breaches quickly and accurately. Achieving adaptation to changing cyber threats and maintaining detection accuracy while minimizing false positives and negatives are challenges. An intrusion detection system (IDS) and a firewall are related to network security, but an IDS is different from a firewall because an IDS is a next-generation firewall, whereas a traditional network firewall employs a static set of rules to allow or prohibit network connections. Implicit intrusion prevention is ensured, provided that a suitable set of rules has been established. Firewalls essentially prevent intrusion by restricting access between networks and by not alerting users to internal attacks. Once an intrusion is suspected, an IDS raises the alarm and details the incident. An intrusion detection system also keeps an eye out for internal system threats.

Traditionally, this has been accomplished by looking at network traffic, spotting patterns and heuristics (sometimes referred to as signatures) of typical computer attacks and taking appropriate action to notify operators [3].

AI and machine learning can quickly analyse millions of data sets and find a wide range of cyber risks, from malware menaces to dubious behaviour that could lead to a phishing attempt, these technologies are increasingly becoming indispensable to information security.

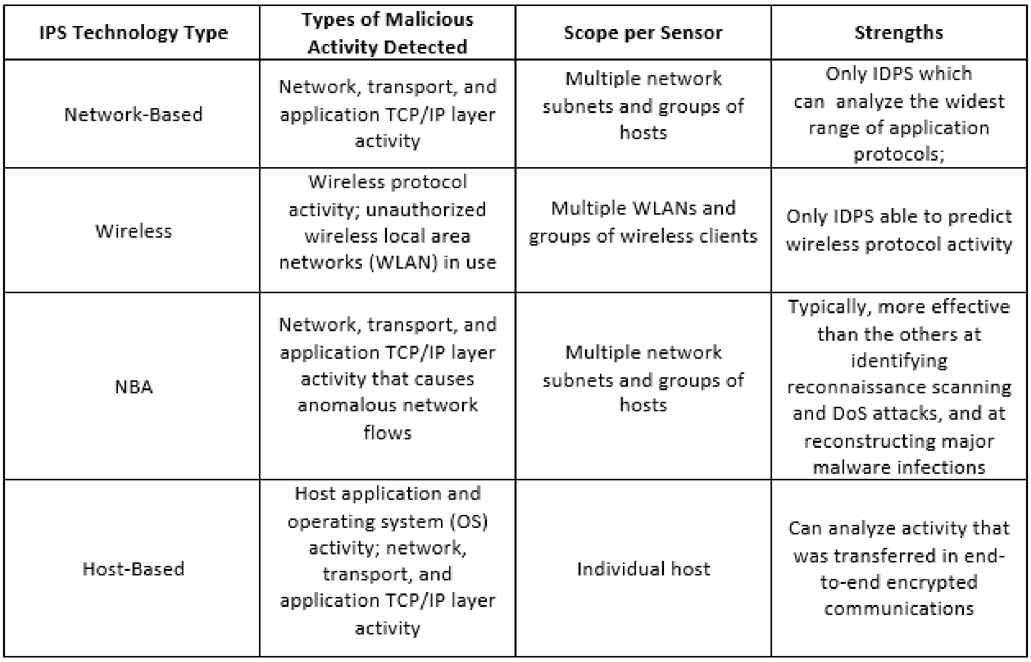
In cybersecurity, AI creates secure apps by default, removing user vulnerabilities. AI ensures accuracy in problem detection, accelerates inquiry, and automates response processes by eliminating unfavourable defaults. AI-driven solutions, such behavioural biometrics for user authentication, encourage the development of secure apps and a safe data ecosystem, which strengthens the infrastructure. Organisations can anticipate and stop cyberattacks before they happen thanks to AI's ability to recognise potentially harmful activity and threat actors. Systems may be continuously protected using AI-enabled automated monitoring, enabling businesses to take preventative action to preserve their digital assets before damage is done.

Cybersecurity relied mostly on signature-based detection methods prior to artificial intelligence (AI) to fend off threats. These systems carried out a database comparison between incoming network traffic and known threat signatures. When a match was detected, the system would sound an alert and safe data ecosystem, which strengthens the infrastructure. Organisations can anticipate and stop cyberattacks before they happen thanks to AI's ability to recognise potentially harmful activity and threat actors. Systems may be continuously protected using AI-enabled automated monitoring, enabling businesses to take preventative action to preserve their digital assets before damage is done.

Cybersecurity relied mostly on signature-based detection methods prior to artificial intelligence (AI) to fend off threats. These systems carried out a database comparison between incoming network traffic and known threat signatures. When a match

was detected, the system would sound an alert and take appropriate action to either restrict or eliminate the threat. To make matters worse, traditional cybersecurity operations relied heavily on manual analysis. Security analyst pays attention to security warnings and log data, searching for patterns or hints that can indicate potential security breaches. This tedious approach required a lot of time and primarily relied on the skills of individual security analysts to accurately spot risks.

The following figure illustrates the different types of Intrusion prevention systems (IPS) technologies.



IPS Technology Type	Types of Malicious Activity Detected	Scope per Sensor	Strengths
Network-Based	Network, transport, and application TCP/IP layer activity	Multiple network subnets and groups of hosts	Only IDPS which can analyze the widest range of application protocols;
Wireless	Wireless protocol activity; unauthorized wireless local area networks (WLAN) in use	Multiple WLANs and groups of wireless clients	Only IDPS able to predict wireless protocol activity
NBA	Network, transport, and application TCP/IP layer activity that causes anomalous network flows	Multiple network subnets and groups of hosts	Typically, more effective than the others at identifying reconnaissance scanning and DoS attacks, and at reconstructing major malware infections
Host-Based	Host application and operating system (OS) activity; network, transport, and application TCP/IP layer activity	Individual host	Can analyze activity that was transferred in end-to-end encrypted communications

Fig: 1.1 various kinds of IPS Technologies

In cybersecurity, AI creates secure apps by default, removing user vulnerabilities. AI ensures accuracy in problem detection, accelerates inquiry, and automates response processes by eliminating unfavourable defaults. AI-driven solutions, such behavioural biometrics for user authentication, encourage the development of secure apps and a safe data ecosystem, which strengthens the infrastructure. Organisations can anticipate and stop cyberattacks before they happen thanks to AI's ability to recognise potentially harmful activity and threat actors. Systems may be continuously protected using AI-enabled automated monitoring, enabling businesses to take preventative action to preserve their digital assets before damage is done [5].

Strong cybersecurity measures are crucial in an increasingly linked world where the digital landscape is growing at an unprecedented rate. Protection of sensitive data and digital assets has become a major concern for people, companies, and governments due to the rise of sophisticated cyber threats, which include ransomware, phishing, and malware. Artificial intelligence (AI), which offers a multitude of advantages that improve the effectiveness of cybersecurity efforts, has emerged as a game-changer as traditional cybersecurity tactics fail to keep up with the ever-evolving threat landscape. We examine the top ten advantages that artificial intelligence (AI) offers the field of cybersecurity in this post [9].

DOI: 10.17148/IJARCCE.2024.13644

Robust cybersecurity measures are now essential in an era of digital transformation where organisations largely rely on technology and data to drive their operations. Traditional security solutions are insufficient to secure sensitive data and vital systems due to the complexity and frequency of cyber threats. The combination of artificial intelligence (AI) and cybersecurity is changing how businesses prevent cyberattacks, identify security holes, and handle breaches.

II. LITERATURE REVIEW

Sowmya et al. [2] study exhibits a thorough review and classification of AI-based intrusion detection systems, analysing over 70 research papers on machine learning, deep learning, and ensemble methods. The analysis considers detection algorithms, performance metrics, and advantages and limitations. Key findings show deep learning models achieve over 99% accuracy on benchmark datasets, outperforming machine learning approaches, with convolutional neural networks being particularly effective. However, challenges remain in detecting novel attack types and performance issue.

An Artificial Intelligence powered network threat detection system (AI-NTDS) has been proposed by BO-XIANG WANG et al. [4]. It analyses attacker behaviour and categorizes threats into three severity categories using machine learning. Fifty-two features related to messages, hosts, and geography are extracted from a honeypot dataset. Experiments show the system achieves 99% accuracy in detecting threats, outperforming previous methods. The analysis identifies message length and file execution as the most important indicators of maliciousness.

Maurice Dawson et al. [3] explores using AI models like ChatGPT to enhance intrusion detection systems (IDS) for cybersecurity. It reviews limitations of traditional IDS methods and advantages of AI-based techniques, including improved adaptability, pattern recognition, and real-time threat detection. The research examines case studies showing successful implementations of AI-powered IDS across sectors like banking and proposes future directions like integrating ChatGPT to analyse network traffic and identify anomalies.

Cheolhee et al. [1] proposed a novel AI-based network intrusion detection system (NIDS) that leverages generative models to synthesize data and address class imbalance. The proposed autoencoder-based models outperformed previous machine learning approaches, achieving over 90% accuracy on benchmark datasets. Experiments demonstrated feasibility for real-world environments, motivating future work on ensemble systems and adversarial attack resistance.

Lin et al. [15] proposed an additional CNN-based IDS. Their response is divided into two parts. CNN training is done offline in the first. A maximum pooling layer, one or more convolutional layers, and a 9 × 9 input layer are used to reduce the model from its initial state to a 1 × 1 output layer. In the second step of their system, the online detection phase, traffic is intercepted using an open-source intrusion detection system called Suricata.

The result of the detection is then obtained by pre-processing the packets and applying the trained model to the network traffic. They used the CICIDS2017 dataset to evaluate their model. Testing was done utilizing both the raw traffic dataset and the feature dataset. Their model works better with raw traffic than with an extracted feature set, as evidenced by their respective accuracy scores of 96.55% and 99.56%.

Patrick Vanin et al [5] proposed that IDS is insufficient on its own for businesses looking to defend against intrusions. Intrusion Prevention Systems (IPS) are gradually replacing Intrusion Detection Systems (IDS). Similar to an IDS, an IPS has active components to thwart assaults before they have a chance to succeed. An IPS typically comprises of an IDS-rule-equipped firewall. IPSs are installed inline, as opposed to IDSs, which implies that they continuously scan the traffic as it goes through them. Because latency problems in a network can impair user experience, an intrusion prevention system (IPS) must be quick and have a large processing capacity.

Khraisat et al. [6] proposed that the sophistication of cyberattacks is rising, which makes it harder to identify breaches with accuracy. The confidence of security services, such as data confidentiality, integrity, and availability, could be damaged if the incursions are not stopped. To combat computer security risks, a variety of intrusion detection techniques have been put forth in the literature. These techniques can be broadly divided into two categories: Anomaly-based Intrusion Detection Systems (AIDS) and Signature-based Intrusion Detection Systems (SIDS). This survey study provides an overview of the datasets frequently used for evaluation, a taxonomy of modern IDS, and a thorough discussion of noteworthy recent publications. To increase computer system security, it also outlines attack evasion strategies and talks about upcoming research challenges to combat them.

Shruti Patil et al. [8] proposed that to stop and lessen threats, intrusion detection systems are widely used in the cyber security industry. Systems for detecting intrusions (IDS) aid in preventing threats and weak points from entering computer

DOI: 10.17148/IJARCCE.2024.13644

networks. There are several machine learning techniques that can be used to create intrusion detection systems that are efficient. In terms of learning, machine learning ensemble approaches have a strong track record. This research proposes a novel intrusion detection system using ensemble machine learning techniques. To enhance the precision of classification and remove false positives, characteristics from the CICIDS-2017 dataset were selected. In this research, an intrusion detection system based on SVM, random forests, and decision trees (IDS) machine learning techniques is proposed.

Diverse and heterogeneous data are exchanged via network systems in remote contexts as communication technology develops. In the meantime, worries about network security have grown as communication technology has advanced and the attack surface has grown. As a result, research on Network Intrusion Detection Systems (NIDS) has been aggressively undertaken to address possible threats. Recently, artificial intelligence (AI)-based anomaly detection systems have garnered attention among the many NIDS technologies, and several models have been put forth to enhance NIDS performance. Still, there is the issue of data imbalance, which prevents AI models from effectively learning criminal behaviour and from correctly identifying network risks. In this paper, we provide a novel intrusion detection method for networks based on artificial intelligence that can effectively address data imbalances.

Nour Moustafa et al. [13] proposed that the lack of a comprehensive network-based data set that can represent contemporary network traffic scenarios, a wide range of low-footprint intrusions, and depth-structured information about the network traffic is one of the main research problems in this subject. Ten years ago, benchmark data sets KDD98, KDDCUP99, and NSLKDD were created for the purpose of evaluating research efforts related to network intrusion detection systems. Nevertheless, a few recent studies have demonstrated that these data sets do not accurately represent network traffic and contemporary low-footprint attacks in the current network security environment. To address the issue of network benchmark data sets not being readily available, this research investigates the establishment of a UNSW-NB15 data set. The attack activities of the network traffic in this data set are a hybrid of the actual modern normal and the modern synthetic ones.

III. MACHINE LEARNING METHODS

ML algorithms are used in cybersecurity to generate numerous critical predictions that stop attacks by deleting data, hence preventing cybercrimes. The primary focus of this section is on the widely used machine learning algorithms, which serve as a security tool to anticipate assaults [2]. Machine learning (ML) has shown to be the most efficient security method for threat detection, and developing an anomaly detection system requires a grasp of the system's semantic features. Understanding the threat model is necessary for developing an ML-based security solution, and this may be done by identifying the behaviours of the attacker or the system environment [10]. Three essential elements are needed for the detection process to function: parameterization, training, and detection. Observable behaviours, such hosts and network connections, are converted into representations during parameterization. Using these representations, a classification model that can distinguish between normal and pathological behaviours is developed during the training phase. The model created during training is used in the detection step to forecast and locate new anomalies [2]. The main types of machine learning methods are: supervised, unsupervised and reinforcement machine learning.

A. Supervised Learning

Supervised machine learning creates a function in such a way that it maps an input to an output using labelled data. Labelled training data are used to develop the function [5]. The goal of the supervised learning technique is to develop a model that can predict whether a new event is abnormal or normal using this feedback. The system gets increasingly accurate as more feedback is received [3]. In supervised learning there are two types of models: Classification and Regression. In Ref. [2], [5-8] An algorithm is used in classification to precisely place test data into designated groups. It identifies entities in the dataset and tries to make recommendations on the definition or labelling of those items. The next section provides a more detailed description of several popular classification techniques, including random forest, k-nearest neighbour, decision trees, support vector machines (SVM), and linear classifiers.

To comprehend the link between dependent and independent variables, regression is utilised. It is frequently used to project things like sales income for a certain company. Popular regression algorithms include polynomial, logistical, and linear regression.

B. Unsupervised Learning

Unsupervised learning involves training a model on unlabelled data so that it can find patterns or anomalies without direct instruction on what the right output should be. Unsupervised learning issues fall into three categories: Dimensionality Reduction, Association and Clustering [5]. In Ref. [2], [5-8] some of the Unsupervised algorithm used for IDS are Kmeans, Fuzzy C-means, Gaussian Mixture and Apriori algorithm. Unsupervised learning is a type of machine learning where the models are trained on input data without explicit supervision or labeled responses. In other words, the algorithm learns from the data without any guidance or predefined labels. The main goal of unsupervised learning is to identify patterns, relationships, and structures within the data. It is commonly used for tasks such as clustering, dimensionality reduction, density estimation, and anomaly detection.

C. *Reinforcement Learning*

The interdisciplinary field of reinforcement learning (RL) combines optimal control and machine learning to study how an intelligent agent should behave in a dynamic environment to maximise the cumulative reward. Along with supervised learning and unsupervised learning, reinforcement learning is one of the three fundamental paradigms in machine learning. Reinforcement learning is distinct from supervised learning in that it does not need the presentation of labelled input/output pairings or the explicit correction of suboptimal behaviours. Rather, the emphasis lies in striking a balance between the exploration of unexplored area and the utilisation of existing information, with the goal of optimising the long-term reward—whose return on investment may be partial or delayed.

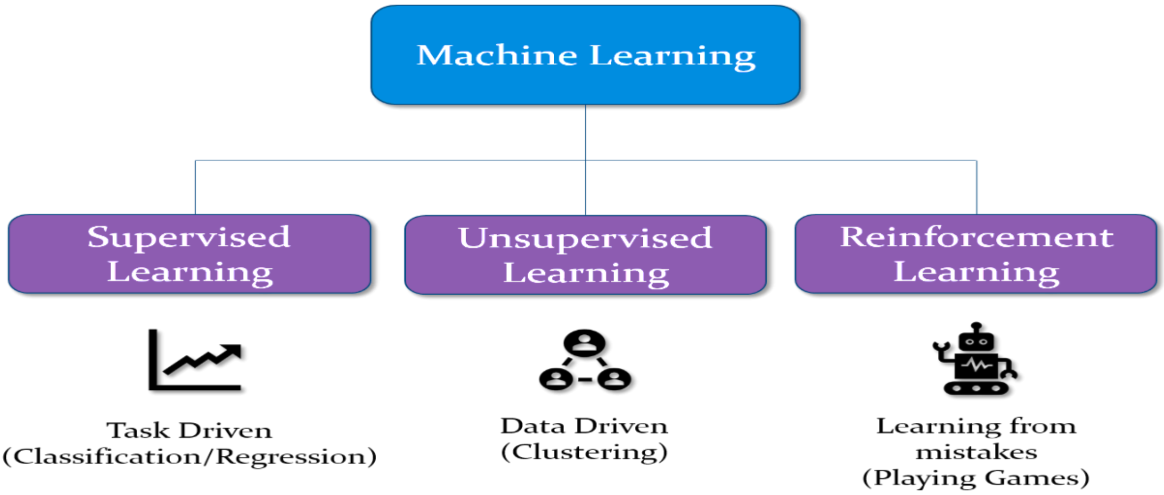


Fig 3. 1: Types of Machine Learning

IV. Data Set

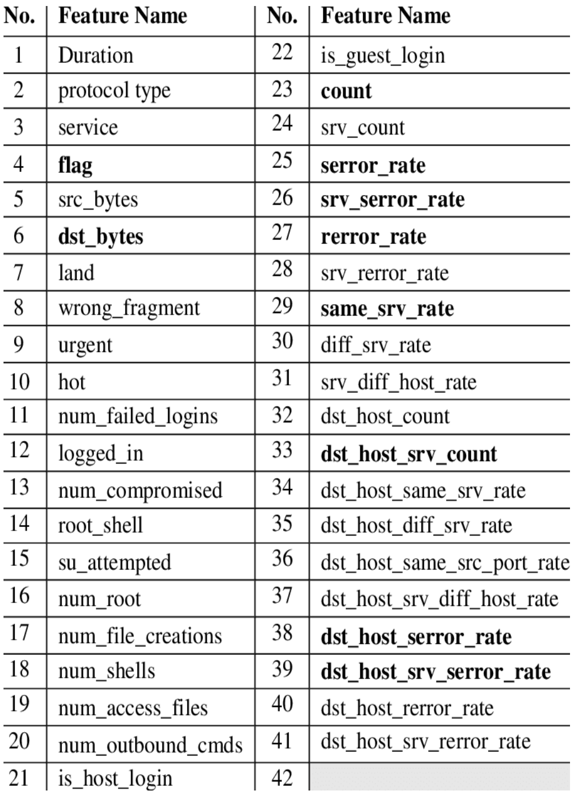
When creating and assessing intrusion detection systems, selecting the appropriate dataset is essential. The following are some well-known datasets that are frequently utilized in intrusion detection and cybersecurity:

A. *KDDCUP99*

most of the datasets are most frequently used to evaluate IDS is KDDcup99. The KDDcup99 has over 4,900,000 samples in it. Every sample is labelled as Attack or Normal and has 41 attributes. The attack samples are divided into four groups: Remote to Local (R2L), User to Root (U2R), Denial of Service (DoS), and Probe [5]. One of the main issues with this dataset is its imbalance; for instance, in key categories like DoS and Probe, there is a lot of similar samples, but not many in R2L and U2R. Some classes may not exist at all depending on whether portion of the dataset is used [11].

B. *NSL-KDD*

The KDDcup99 dataset's primary problem was resolved with the creation of this dataset. Tavallaee made the proposal in 2009 [12]. The KDDcup99's four assault categories are retained. The NSL-KDD suggests two data files: a training set and a testing set. The training set consists of 126,620 examples with 21 different types of attacks. The testing set has 22,850 examples and 37 different types of attacks [1], [7], [12]. Which is illustrated in figure 4.2



No.	Feature Name	No.	Feature Name
1	Duration	22	is_guest_login
2	protocol type	23	**count**
3	service	24	srv_count
4	**flag**	25	**serror_rate**
5	src_bytes	26	**srv_serror_rate**
6	**dst_bytes**	27	**rerror_rate**
7	land	28	srv_rerror_rate
8	wrong_fragment	29	**same_srv_rate**
9	urgent	30	diff_srv_rate
10	hot	31	srv_diff_host_rate
11	num_failed_logins	32	dst_host_count
12	logged_in	33	**dst_host_srv_count**
13	num_compromised	34	dst_host_same_srv_rate
14	root_shell	35	dst_host_diff_srv_rate
15	su_attempted	36	dst_host_same_src_port_rate
16	num_root	37	dst_host_srv_diff_host_rate
17	num_file_creations	38	**dst_host_serror_rate**
18	num_shells	39	**dst_host_srv_serror_rate**
19	num_access_files	40	dst_host_rerror_rate
20	num_outbound_cmds	41	dst_host_srv_rerror_rate
21	is_host_login	42	

Fig 4.2 NSL-KDD dataset

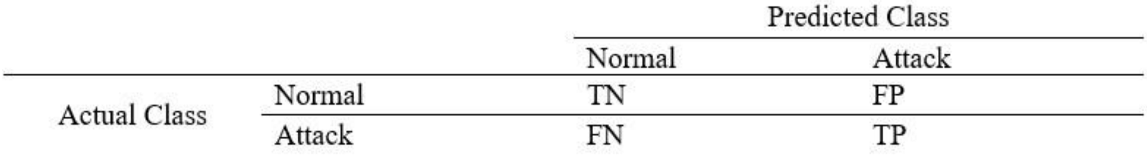
C. 4.3. UNSW-NB15

This data was created by the Australian Centre for Cyber Security. Its purpose was to produce traffic, a cross between regular activity and assault behaviours. There are nine different kinds of attacks in this dataset: worms, reconnaissance, generic, DoS, backdoors, fuzzers, and shellcode. UNSW suggests two files: a training set and a testing set. The original dataset's information about different types of traffic, such as attacks and ordinary traffic, can be found in these files. The original dataset consists of 2,540,044 records; the training set consists of up of 175,341 records; the testing set is built up of 82,332 records [13].

D. CIC-IDS 2017

The Canadian Institute of Cybersecurity has released the intrusion detection dataset CICIDS2017. It satisfies requirements for real-world assaults and includes current network attacks. The initial dataset included eight files including five days' worth of network traffic recordings. After merging the files, a sizable dataset with 83 features spanning 15 different attack type and over 3 million occurrences were created. Eight attack class labels, 2.8 million instances, 83 characteristics, and a lower-class imbalance ratio are present in the final dataset. Intrusion detection systems may benefit from using this enhanced dataset for training.

V. PERFORMANCE EVALUATION



		Predicted Class	
		Normal	Attack
Actual Class	Normal	TN	FP
	Attack	FN	TP

Fig 2. Confusion Matrix

DOI: 10.17148/IJARCCE.2024.13644

Performance metrics are used to assess IDSs, which employ machine learning techniques. A confusion matrix is a table that is used to assess how well a classification algorithm performs when applied to a collection of test data that has known true values. Metrics used to access the performance of IDS is given in Table1. The following terms are used to describe a confusion matrix: True Positive (TP): A precise identification of an assault sample as such has been made. True Negative (TN): Accurate identification of a normal sample as typical traffic. False Positive (FP): When an assault is unintentionally classified as a representative sample. False Negative (FN): An attempt has been made to identify a sample of assault as normal traffic.

A. Accuracy

$$\text{Accuracy} = \frac{TP+TN}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$$

Percentage of cases out of all predictions that were correctly classified.

B. Precision

Percentage of all positive instances which were accurately predicted out of all favourable circumstances.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

C. Recall

Percentage of all actual positive events that were accurately anticipated.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

D. F1-Score

Precision and Recall's harmonic mean, which combines their contributions for a single efficiency measure.

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

VI. CONCLUSION

According to the survey, utilizing the CIC-IDS 2017 dataset and the K-Nearest Neighbors (KNN) method to construct an intrusion detection system (IDS) would be a reasonable course of action. The main arguments in favor of this conclusion are: As noted in the literature review (Section 3.1), K-Nearest Neighbors (KNN) is a supervised machine learning method that has been widely employed for constructing intrusion detection systems.

The Canadian Institute of Cybersecurity released the CIC-IDS2017 dataset, which is a complete and current dataset that includes real-world traffic patterns and a variety of contemporary network attack methods (Section 4.4). Compared to other datasets like KDDCup99 and NSL-KDD, the CIC-IDS 2017 dataset has a lower-class imbalance ratio and is made to tackle the difficulties of real-world attacks. This can help the IDS model perform better.

The survey highlights how machine learning methods, such as supervised algorithms like KNN, have been effectively used to create intrusion detection systems, resulting in highly accurate and successful attack detection of a variety of kinds. Consequently, we can create an intrusion detection system that can potentially detect a variety of network attacks and adapt to new attack patterns by utilizing the KNN algorithm and the extensive CIC-IDS 2017 dataset. This will allow us to take advantage of the KNN algorithm's efficiency and simplicity

DOI: 10.17148/IJARCCE.2024.13644

REFERENCES

[1]. Cheolhee Park, Jonghoon Lee, Youngsoo Kim, Jong- Geun Park, Hyunjin Kim, and Dowon Hon, “An Enhanced AI-Based Network Intrusion Detection System Using Generative Adversarial Networks” in IEEE, 2023

[2]. Sowmya T.A , Mary Anita E.A.B, “A comprehensive review of AI based intrusion detection system” in Elsevier, 2023

[3]. Michal Markevych, Maurice Dawson, “A Review of Enhancing Intrusion Detection Systems for Cybersecurity using Artificial Intelligence (AI)” in Sciendo, 2023

[4]. BO-XIANG WANG, JIANN-LIANG CHEN (Senior Member, IEEE) AND CHIAO-LIN YU, “An AI-Powered Network Threat Detection System” in IEEE, 2022

[5]. Patrick Vanin, Thomas Newe, Lubna Luxmi Dhirani, Eoin O’Connell, Donna O’Shea, Brian Lee and Muzaffar Rao, “A Study of Network Intrusion Detection Systems Using Artificial Intelligence/Machine Learning” in MDPI, 2022

[6]. Ansam Khraisat, Iqbal Gondal, Peter Vamplew and Joarder Kamruzzaman, “Survey of intrusion detection systems: techniques, datasets and challenges” in Springer Open,2019

[7]. Elijah M. Maseno, Zenghui Wang and Hongyan Xing, “A Systematic Review on Hybrid Intrusion Detection System” in The Wiley Hindawi Publishing,2022

[8]. Shruti Patil, Vijayakumar Varadarajan, Siddiqui Mohd Mazhar, Abdulwodood Sahibzada, Nihal Ahmed, Onkar Sinha, Satish Kumar, Kailash Shaw and Ketan Kotecha, “Explainable Artificial Intelligence for Intrusion Detection System” in MDPI, 2022

[9]. Alex Shenfielda, David Dayb, Aladdin Ayeshb, “Intelligent intrusion detection systems using artificial neural networks” in Elsevier, 2018

[10]. Mesut Ugurlu, I. Dogru, “A survey on DL based intrusion detection system” in 2019 4th International Conference on Computer Science and Engineering (UBMK), 2019

[11]. L. Ashiku, C. Dagli, “Network Intrusion Detection System using Deep Learning” in Elsevier, 2021

[12]. Tavallaee.M, Bagheri.E, Lu Wei, Ali A. Ghorbani, “A detailed analysis of the KDD CUP 99 data set” in IEEE, 2009

[13]. Nour Moustafa, Jill Slay, “UNSW-NB15-A comprehensive data set for Network Intrusion Detection Systems in Military Communications and Information Systems” Conference (MilCIS), 2015

[14]. Samarjeet Borah, Ranjit Panigrahi, “A detailed analysis of CICIDS2017 dataset for designing Intrusion Detection Systems” in International Journal of Engineering & Technology, 2018

[15]. Chen.L, Kuang. X, Xu. A, Suo.S, Yang.Y, “A Novel Network Intrusion Detection System Based on CNN” in Eighth International Conference on Advanced Cloud and Big Data (CBD), Taiyuan, China, 2020