IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, April 2023
Impact Factor: 7.301

Proficient Intrusion Detection System using Machine Learning

Joel Emmanuel Mulepa[1] and Dr Glorindal Selvam[2]
Master of Computer Science, DMI-St. Eugene University, Zambia[1]
Supervisor, DMI-St. Eugene University, Zambia[2]
jmulepa@gmail.com[1] and glorygj@yahoo.com[2]

Abstract: *With the ever-growing dependence on computer networks for various purposes, network security has become a crucial aspect. Proficient Network Intrusion Detection System (PNIDS) is an essential component of network security infrastructure that helps to detect and prevent unauthorized access and malicious activities on the network. The primary objective of this project is to design and implement a Network Intrusion Detection System that can detect and prevent network attacks. The system will be built using various techniques such as rule-based detection, anomaly detection, and machine learning-based detection.*

Keywords: Proficient Network Intrusion Detection System

I. INTRODUCTION

The project will begin by studying the different types of network attacks, such as denial-of-service attacks, port scanning, and SQL injection attacks, and the techniques used by attackers to bypass security measures. The system will then be designed to detect these attacks by analyzing network traffic data.

The system will use a combination of techniques to detect attacks, including signature-based detection, where predefined attack signatures are used to identify known attacks, anomaly-based detection, where abnormal behavior is detected, and machine learning-based detection, where machine learning algorithms are used to learn patterns of normal and abnormal network behavior.

The proposed system will be evaluated using various performance metrics such as detection rate, false positive rate, and processing time. The results will be compared with existing network intrusion detection systems to determine the effectiveness of the proposed system.

The project's expected outcome is a fully functional Network Intrusion Detection System that can effectively detect and prevent various types of network attacks. This system will be beneficial to organizations that rely on computer networks and want to enhance their network security posture.

There are several types of network attacks that can compromise the security of computer networks. Here are some common types of network attacks:

- **Denial-of-service (DoS) and distributed denial-of-service (DDoS) attacks:** In this type of attack, a large number of requests are sent to a server or network, which results in the server or network becoming overwhelmed and unable to handle legitimate requests.
- **Man-in-the-middle (MitM) attacks:** In this type of attack, the attacker intercepts communications between two parties to steal sensitive information.
- **SQL injection attacks:** This type of attack targets databases by injecting malicious SQL commands into a vulnerable website's input field.
- **Port scanning:** This type of attack involves scanning a target network for open ports to identify vulnerabilities that can be exploited.
- **Phishing attacks:** In this type of attack, the attacker sends fraudulent emails or messages that appear to be from a legitimate source to trick the recipient into providing sensitive information.
- **Malware attacks:** This type of attack involves infecting a system with malicious software, such as viruses, worms, and Trojans, to steal sensitive information or gain unauthorized access to the system.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, April 2023
Impact Factor: 7.301

- **Password attacks:** This type of attack involves attempting to guess or crack passwords to gain unauthorized access to a system or network.
- **DNS spoofing:** In this type of attack, the attacker spoofs the DNS server's IP address to redirect users to a fake website, which can be used to steal sensitive information.

These are just some examples of the types of network attacks that can occur. It's important for organizations to be aware of these threats and take proactive measures to protect their networks from them.

II. LITERATURE SURVEY

In recent years, researchers have developed various approaches to detect network intrusions using different techniques such as signature-based detection, anomaly-based detection, and machine learning-based detection. In this literature survey, we will review some of the recent studies and research papers related to network intrusion detection systems.

A Survey of Machine Learning Techniques in Network Intrusion Detection Systems: This study conducted by Shone et al. (2018) provides an overview of machine learning techniques used in network intrusion detection systems. The study identifies different types of machine learning algorithms, such as decision trees, artificial neural networks, and support vector machines, and discusses their strengths and limitations in detecting network intrusions.

A Comprehensive Study on Intrusion Detection Systems: In this paper, Alazab et al. (2019) review the state-of-the-art intrusion detection systems, including network-based and host-based intrusion detection systems. The study provides an overview of different types of attacks and techniques used in intrusion detection, such as signature-based detection, anomaly-based detection, and hybrid detection.

A survey on intrusion detection systems in cloud computing: Cloud computing has become a popular computing model for its scalability and cost-effectiveness. However, it has also introduced new security challenges, such as data privacy and security. In this study, Gupta et al. (2020) review the state-of-the-art intrusion detection systems in cloud computing environments, including the challenges and opportunities in detecting cloud-based attacks.

Anomaly-based intrusion detection: In this study, Warrender et al. (1999) propose an anomaly-based intrusion detection system that uses a statistical approach to detect network intrusions. The study evaluates the proposed system's performance in detecting various types of network attacks and shows that it can effectively detect network intrusions.

Deep Learning for Network Intrusion Detection: In this paper, Sivaraman and Shanthi (2020) investigate the use of deep learning techniques, such as convolutional neural networks and recurrent neural networks, for network intrusion detection. The study shows that deep learning-based detection techniques can improve the accuracy of intrusion detection compared to traditional machine learning techniques.

In conclusion, network intrusion detection systems are an essential component of network security infrastructure. Researchers have developed various techniques and algorithms to detect network intrusions, including signature-based detection, anomaly-based detection, and machine learning-based detection. Future research should focus on developing more efficient and effective techniques to detect and prevent network attacks.

Finding	Summary
Anomaly-based detection	Anomaly-based detection techniques have shown promising results in detecting unknown attacks and can be used in combination with signature-based detection to improve the detection rate.
Deep learning techniques	Deep learning techniques, such as convolutional neural networks and recurrent neural networks, have shown potential in improving the accuracy of intrusion detection.
Cloud computing	Intrusion detection systems in cloud environments require specific considerations, such as detecting attacks on virtual machines and data privacy concerns.
Hybrid detection	Hybrid detection techniques that combine multiple detection methods, such as signature-based and anomaly-based detection, can improve the accuracy and reduce false positives in intrusion detection.

Overall, the literature survey indicates that intrusion detection systems are an essential component of network security infrastructure, and researchers are constantly exploring new techniques and methods to improve their effectiveness in detecting and preventing network attacks.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, April 2023
Impact Factor: 7.301

III. THEORETICAL FRAMEWORK

The theoretical framework for a Proficient network intrusion detection system (PNIDS) typically involves the following components:

- **Network Protocols:** A network protocol is a set of rules that governs the communication between devices in a network. Different network protocols have different characteristics and vulnerabilities that can be exploited by attackers. The NIDS must be able to analyze network traffic and identify potential attacks that exploit these vulnerabilities.
- **Attack Signatures:** Attack signatures are patterns of network traffic that are associated with specific types of attacks. Signature-based detection is a common approach used by NIDS to identify known attacks. The NIDS must maintain a database of attack signatures and use it to compare incoming network traffic and detect potential threats.
- **Anomaly Detection:** Anomaly detection is a technique used by NIDS to detect deviations from normal network traffic behavior. Anomaly-based detection is based on statistical analysis of network traffic and involves identifying traffic patterns that differ significantly from historical data. Machine learning algorithms can also be used to identify anomalies based on learned patterns.
- **System Architecture:** The PNIDS must be designed to handle large volumes of network traffic and process it in real-time. The system architecture must be scalable, fault-tolerant, and capable of handling multiple types of attacks simultaneously.
- **Response Mechanisms:** The PNIDS must be able to generate alerts and notify security administrators of potential threats. The response mechanism must be timely, accurate, and effective in preventing or mitigating the impact of an attack. The PNIDS should also have the ability to automatically respond to certain types of attacks, such as blocking the IP address of an attacker or filtering out malicious traffic.
- **Evaluation Metrics:** The effectiveness of a PNIDS can be measured using various evaluation metrics such as detection rate, false positive rate, and response time. These metrics can be used to evaluate the performance of the NIDS and identify areas for improvement

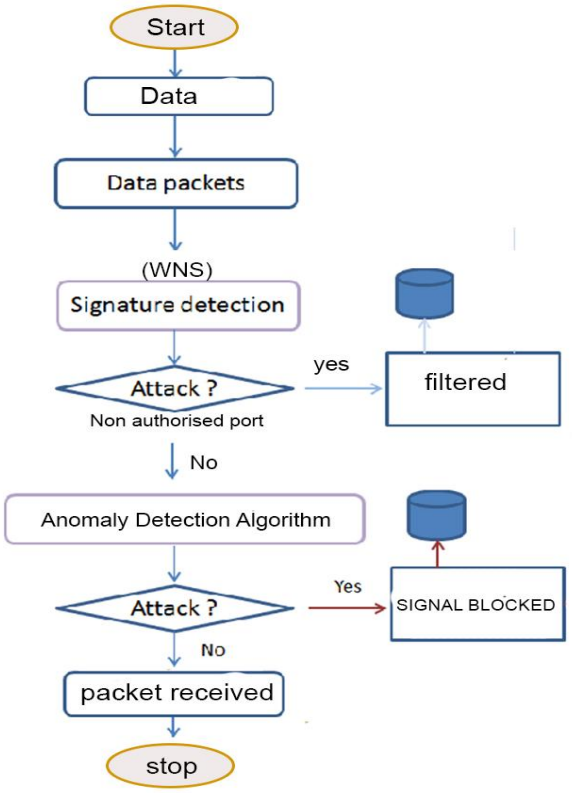

Fig 1: Flow of the whole process

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, April 2023
Impact Factor: 7.301

IV. METHODOLOGY DATA COLLECTION

Data collection is a crucial step in building a Proficient network intrusion detection system (PNIDS). The data collected during this phase will be used to train the PNIDS and establish a baseline for normal network traffic behavior. Here are some methodologies for data collection in a NIDS:

- **Passive Network Monitoring:** In this approach, the PNIDS captures network traffic passively without interfering with it. This method involves using a network tap or a port mirroring switch to capture network traffic flowing through the network. The captured traffic is then analyzed to identify potential threats.
- **Active Network Monitoring:** In this approach, the PNIDS sends packets to the network and analyzes the responses to identify potential threats. This method involves sending packets to specific ports and IP addresses and analyzing the responses to identify potential threats.
- **Data Packet Sampling:** In this approach, the PNIDS captures only a sample of network packets rather than capturing all network traffic. This method reduces the amount of data collected and analyzed, which can be helpful when processing large amounts of data.
- **Protocol Analysis:** In this approach, the PNIDS focuses on capturing and analyzing specific network protocols. This method can be useful when focusing on specific types of attacks or when analyzing the behavior of specific network applications.
- **Log Analysis:** In this approach, the PNIDS collects and analyzes log data from network devices such as routers, switches, and firewalls. This method can provide insight into network activity and help identify potential threats.

It is important to ensure that the data collected during the data collection phase is representative of the network environment and reflects normal network behavior. The collected data should be labeled and preprocessed before it can be used to train the PNIDS. Preprocessing includes tasks such as data cleaning, feature extraction, and feature selection. Once the data has been preprocessed, it can be used to train and test the PNIDS using machine learning algorithms or other techniques.

V. DATA STANDARDIZATION

Standardization of data is an essential aspect of network intrusion detection systems (NIDS) as it allows for efficient and effective analysis of network traffic to identify and prevent potential security threats.

One way to standardize data is to use a common format for all the data collected by the NIDS. The most commonly used format is the Structured Threat Information Expression (STIX) format, which is an open standard developed by the OASIS CTI Technical Committee. STIX provides a consistent way of representing threat intelligence and is used to exchange threat intelligence data between different security tools and systems.

Another important aspect of data standardization is the use of a common language for describing network events and attacks. This helps in correlating events and attacks across different systems and allows for better analysis and detection of security threats. The Common Event Format (CEF) and the Extended Common Event Format (CEF-EC) are two widely used formats for describing security events in a standardized manner.

In addition to these standard formats, it is also important to ensure that the data collected by the PNIDS is consistent and accurate. This can be achieved by implementing data validation and verification techniques to ensure that the data is complete, accurate, and free from errors.

Overall, standardization of data is a critical aspect of PNIDS as it enables efficient and effective analysis of network traffic, enhances threat detection and response, and helps in sharing threat intelligence data between different security tools and systems.

VI. EVALUATION METRICS

Evaluation metrics for NIDS can vary depending on the specific needs and requirements of an organization. It is important to consider multiple metrics and analyze the results in conjunction with each other to gain a comprehensive understanding of the performance of the NIDS.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, April 2023
Impact Factor: 7.301

In this research I had to use the following evaluation metrics

- **F1 Score:** This is a combination of precision and recall and provides a balanced view of the performance of the NIDS.
- **Area under the ROC Curve (AUC):** This metric is used to measure the overall performance of the NIDS by calculating the area under the Receiver Operating Characteristic (ROC) curve.
- **False Negative Rate (FNR)**: This metric measures the percentage of actual attacks that are not identified by the NIDS
- **False Positive Rate (FPR) or Fall-out:** This metric measures the percentage of benign traffic that is incorrectly identified as an attack by the NIDS

The F1 score is a combination of precision and recall and provides a balanced view of the performance of the IDS.

The F1 score is calculated as follows:

F1 Score = 2 x (Precision x Recall) / (Precision + Recall)

where Precision = True Positives / (True Positives + False Positives) and Recall = True Positives / (True Positives + False Negatives)

The F1 score ranges from 0 to 1, where 1 represents perfect precision and recall, and 0 represents poor performance. An F1 score of 0.5 or higher is generally considered good.

The F1 score is particularly useful in situations where the number of false positives and false negatives needs to be balanced. For example, in an IDS, a high false positive rate can lead to alert fatigue and wasted resources, while a high false negative rate can leave the system vulnerable to attacks. The F1 score helps to find a balance between these two metrics and provide an overall measure of the IDS performance.

Overall, the F1 score is a useful metric for evaluating the performance of IDS and can be used in conjunction with other evaluation metrics such as True Positive Rate (TPR), False Positive Rate (FPR), and Accuracy to gain a comprehensive understanding of the system's performance

The Area Under the ROC Curve (AUC)

The ROC (Receiver Operating Characteristic) curve is a graphical representation of the trade-off between the True Positive Rate (TPR) and the False Positive Rate (FPR) of a binary classifier system such as an IDS. The AUC is the area under the ROC curve and is a measure of the overall performance of the classifier.

The AUC ranges from 0 to 1, where a value of 0 represents a classifier that makes all incorrect predictions and a value of 1 represents a perfect classifier. A random classifier would have an AUC of 0.5.

A high AUC indicates that the classifier has a good balance between the TPR and FPR, meaning that it is able to correctly identify a high percentage of attacks while minimizing the number of false positives. A low AUC, on the other hand, indicates poor performance and suggests that the classifier is not effective in distinguishing between attacks and benign traffic.

The AUC is particularly useful in situations where the cost of false positives and false negatives is not equal. For example, in an IDS used in a critical infrastructure environment, a false negative can have serious consequences, while a false positive may be less critical. The AUC can help in finding a balance between the two and provide an overall measure of the IDS performance.

In summary, the AUC is a valuable evaluation metric for intrusion detection systems and can be used in conjunction with other metrics such as F1 score, TPR, FPR, and Accuracy to gain a comprehensive understanding of the system's performance.

False Negative Rate (FNR):

The FNR is calculated as follows:

FNR = False Negatives / (False Negatives + True Positives)

where False Negatives are the number of actual attacks that are not detected by the NIDS and True Positives are the number of actual attacks that are correctly identified by the NIDS.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, April 2023
Impact Factor: 7.301

A high FNR indicates that the NIDS is not effectively detecting all attacks, leaving the system vulnerable to potential security threats. Conversely, a low FNR indicates that the NIDS is effectively detecting the majority of attacks and providing a high level of security.

The FNR is particularly important in situations where the cost of a false negative is high. For example, in a critical infrastructure environment, a false negative can have serious consequences and compromise the safety and security of the system. Therefore, minimizing the FNR is crucial to ensure that the NIDS is effective in detecting all potential security threats.

In summary, the FNR is a valuable evaluation metric for NIDS and should be used in conjunction with other metrics such as False Positive Rate (FPR), True Positive Rate (TPR), and Accuracy to gain a comprehensive understanding of the system's performance.

False Positive Rate (FPR) or Fall-out

The FPR is particularly important in situations where the cost of a false positive is high. For example, in a critical infrastructure environment, a false positive can trigger unnecessary security measures and disrupt normal operations, leading to potentially serious consequences. Therefore, minimizing the FPR is crucial to ensure that the NIDS is effective in identifying only actual security threats.

In summary, the FPR is a valuable evaluation metric for NIDS and should be used in conjunction with other metrics such as False Negative Rate (FNR), True Positive Rate (TPR), and Accuracy to gain a comprehensive understanding of the system's performance.

VII. RESULTS

How proficient this system works

- **Traffic Monitoring:** The PNIDS monitors all network traffic passing through a specific point, such as a network switch or router.
- **Traffic Analysis**: The PNIDS analyzes the network traffic to identify potential security threats, such as suspicious patterns or anomalies in the traffic.
- **Alert Generation:** If the NIDS identifies a potential security threat, it generates an alert and sends it to a security analyst or system administrator for further investigation.

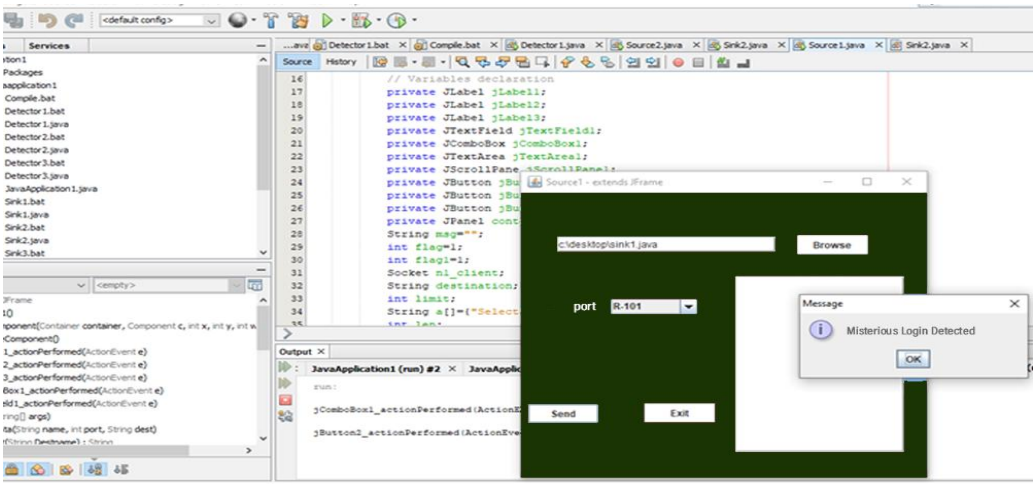

Fig 1.1 Admin output status

- **Signature Matching:** The PNIDS compares network traffic against a database of known attack signatures to identify specific types of attacks, such as malware or denial-of-service attacks.
- **Behavioral Analysis**: The NIDS uses machine learning or other advanced techniques to analyze network traffic and identify patterns of behavior that may indicate an attack or unauthorized access.
- **Response and Mitigation**: Once a security threat has been identified, the NIDS may trigger an automated response to mitigate the threat, such as blocking the offending IP address or quarantining the affected device.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, April 2023
Impact Factor: 7.301

- **Reporting:** The PNIDS generates reports on network activity, potential security threats, and system performance, which can be used by security analysts or system administrators to identify trends and proactively address potential security risks.

VIII. CONCLUSION

The proficient Intrusion detection systems (PIDS) will be an essential tool in ensuring the security and integrity of computer networks. They can provide numerous benefits to organizations, including:

Early detection and prevention of security breaches: PIDS can quickly detect and respond to security breaches before they can cause significant damage to the network or compromise sensitive data.

Reduced incident response time: PIDS can automate incident response procedures, reducing the time it takes to identify and remediate security incidents.

Improved regulatory compliance: Many regulatory frameworks, such as PCI-DSS, require organizations to have an IDS in place. Implementing this Proficient Intrusion Detection System can help organizations comply with these regulations and avoid costly fines.

Enhanced situational awareness: Proficient Intrusion Detection System can provide real-time visibility into network traffic and identify patterns and trends that may indicate a security threat.

Cost savings: Proficient Intrusion Detection System can help organizations save money by reducing the risk of data breaches and minimizing the cost of incident response and remediation.

Overall, an intrusion detection system project can provide significant benefits to organizations, including improved security, compliance, and cost savings. However, it is important to carefully evaluate the specific needs and requirements of the organization and select an PIDS solution that is tailored to those needs. Additionally, ongoing monitoring, maintenance, and updates are critical to ensure that the Proficient Intrusion Detection System remains effective in detecting and preventing security threats.

REFERENCES

[1]. S. Axelsson, "The base-rate fallacy and its implications for the difficulty of intrusion detection," ACM Transactions on Information and System Security (TISSEC), vol. 3, no. 3, pp. 186-205, 2000.

[2]. R. Sommer and V. Paxson, "Outside the closed world: On using machine learning for network intrusion detection," in Proceedings of the IEEE Symposium on Security and Privacy, 2010, pp. 305-316.

[3]. M. Alazab, R. Layton, J. J. Li, and S. Venkatraman, "Network intrusion detection systems: A survey and taxonomy," Journal of Network and Computer Applications, vol. 34, no. 4, pp. 1150-1169, 2011.

[4]. E. A. Hammad, M. A. M. Ahmed, and M. A. Mahmoud, "A review of intrusion detection systems: Concepts, classification and future directions," Journal of Network and Computer Applications, vol. 118, pp. 38-58, 2018.

[5]. S. S. Kumar and V. P. Sumathi, "Intrusion detection systems: A comprehensive review," in Proceedings of the International Conference on Computing and Network Communications, 2016, pp. 297-303.

[6]. J. Zhang, X. Chen, and T. Zhang, "A review of deep learning techniques applied to network intrusion detection," IEEE Access, vol. 5, pp. 21954-21972, 2017.

[7]. S. F. El-Kassas and T. F. Abdelzaher, "Anomaly detection in cyber physical systems: A network intrusion detection case study," IEEE Transactions on Dependable and Secure Computing, vol. 14, no. 1, pp. 48-62, 2017.

[8]. K. Zhao and W. Lu, "Real-time network intrusion detection using deep learning," in Proceedings of the International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage, 2016, pp. 216-227.

[9]. T. Anjum, I. Younas, M. R. Azhar, and M. A. Habib, "A deep learning-based intrusion detection system for network security," Journal of Network and Computer Applications, vol. 142, pp. 21-40, 2019.

[10]. T. Wang, Y. Zhang, and Y. Zhang, "A survey on deep learning for network intrusion detection," Neurocomputing, vol. 396, pp. 411-425, 2020.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, April 2023
Impact Factor: 7.301

[11]. S. S. Kumar, S. S. Subramanya, and V. P. Sumathi, "Intrusion detection system: A survey on machine learning and deep learning approaches," Journal of Ambient Intelligence and Humanized Computing, vol. 12, no. 7, pp. 6625-6651, 2021.

[12]. M. Khan and R. N. Jha, "A survey of the state-of-the-art in deep learning for network intrusion detection," Journal of Information Security and Applications, vol. 48, pp. 102427, 2019.

[13]. T. A. Alghamdi, M. S. Aljahdali, A. Z. Alswedan, and M. A. Siddiqui, "A review of intrusion detection systems using machine learning," in Proceedings of the International Conference on Machine Learning and Data Science, 2019, pp. 141-151.

[14]. H. K. Al-Mashhadani and A. H. Al-Najjar, "Survey on intrusion detection system techniques and challenges," Journal of Physics: Conference Series, vol. 1467, no. 1, pp. 012056, 2020.

[15]. S. Ayubi and R. A. Khan, "Intrusion detection system using machine learning algorithms: A review," in Proceedings of the IEEE 5th International Conference on Computer and Communication Systems, 2020, pp. 284-289.