Cyber-Attack Classification Using AI & ML Techniques

Abhale B A[1], Dr. Rokade P.P.[2] Savita Khatal[3], Nikita Jagdale[4], Rutuja Gaikwad[5] Chaitali Rane[6]
Information Technology & Engineering,
S.N.D. COE & RC Yeola, Maharashtra, India.

.

***ABSTRACT**:The rapid growth of network infrastructure and online services has increased the vulnerability of systems to cyberattacks. To mitigate such risks, our project focuses on classifying cyberattacks using machine learning algorithms. The project is designed as a web-based application, integrating both front-end and back-end technologies to provide an efficient and user-friendly interface. The user provides information to the front-end, including IP addresses, port numbers, data sizes, and protocols, while the back-end processes this data to classify the type of attack using pre-defined rules and AI-based logic. We employ a simple rule-based classifier in the first phase, with the option of incorporating more advanced machine learning algorithms such as decision trees, random forests, and neural networks in future iterations. Our system classifies common types of cyberattacks such as DDoS, SQL injection, SSL injection, and UDP flood attacks. This project aims to be a foundation for further research and enhancement in AI-driven cyber defence systems.*

Keywords
Cybersecurity, Machine Learning, Cyberattack Classification, Web Application, Front-end Development, Back-end Development

1.INTRODUCTION

In today’s digital era, where organizations, governments, and individuals increasingly rely on networked systems, cybersecurity has become more crucial than ever before. [4] As the internet continues to expand, so too does the volume of cyber threats. Attackers are employing more sophisticated techniques, making it difficult for traditional security measures to keep up. Whether targeting sensitive corporate data, governmental operations, or personal information, cyberattacks now have the potential to cause widespread disruption and financial loss. Consequently, safeguarding digital infrastructures is a top priority, and organizations must adopt innovative approaches to detect and neutralize these attacks. [12] Traditionally, cybersecurity systems have relied on signature-based detection techniques and rule-based logic to identify and thwart attacks. These methods involve defining specific patterns or rules that must be met to trigger an alert. While effective against known threats, they are inherently reactive, relying on past data to address future concerns. As new, previously unseen threats emerge, such as zero-day attacks or advanced persistent threats (APTs), these static defences become inadequate. New approaches, particularly those rooted in artificial intelligence (AI) and machine learning (ML), offer promising alternatives for detecting and classifying cyber threats in real-time. Challenges in Cybersecurity: The cybersecurity landscape has grown increasingly complex due to the diverse nature of cyberattacks and the enormous volume of data that needs to be analyzed. Cyberattacks come in various forms, including Distributed Denial of Service (DDoS) attacks, SQL injection, phishing, malware distribution, ransomware, and more. Each type of attack has its own characteristics, making detection particularly challenging for security systems. [2],For instance, a DDoS attack involves overwhelming a target system with an influx of network traffic, whereas phishing attacks exploit human error to compromise sensitive information. The variability in attack strategies requires systems that can accurately and efficiently identify these differences in real time.[1]

2. LITERATURE REVIEW

Ernesto Damiani,Chan Yeun proposed a comprehensive review of current literature on Explainable Artificial Intelligence (XAI) methods for cyber security applications[1]. Ebu Yusuf Güven, Sueda Gülgün, Ceyda Manav, Behice Bakır proposed development of learning intrusion detection systems in order to detect sophisticated and undetected threats[2].

Bhavna Dharamkar, Rajni Ranjan Singh proposed the rest of this paper is organized as follows: Section 2 describes the related work. Section 4 explains proposed method. Section 5 focuses on the experimental results analysis finally, results are summarized and concluded in Section 6[3].Ismail Mohmand, Hameedhussain, Ayaz Khan proposed the Prediction Technique for DDoS Attacks This paper, used a machine learning approach for DDoS attack types classification and prediction[4]

İsa Avcı, Murat Koca proposed Machine Learning Techniques Random Forest (RF), K-Nearest

Neighbors (KNN), and the Decision Tree (DT) systems for intrusion detection are explored[5]. In addition, feature selection techniques are employed for the selection of important features.Kamran Shaukat, Ibrahim A. Hameed,Suhual Luo,Vijay Varadharajan, Min Xu proposed challenges that ML techniques face in protecting cyberspace against attacks, by presenting a literature on ML techniques[6]. Hamed Alqahtani, Asra Kalim proposed in this paper is on cyber threats, DDos attack , dataset, cyber security[7]. This paper explain the threats in cyber security.Rahul Chourasiya, , Vaibhav Patel proposed the ml techniques that can classify the attack[8].

3.PROPOSED METHODOLOGY

System Implementation Plan for AI-Based Cyberattack Classification: The implementation of an AI-based cyberattack classification system is a multi-phase process involving data collection, preprocessing, feature engineering, model training, and real-time integration. This system is designed to detect and classify network attacks such as Distributed Denial of Service (DDoS), SQL injection, phishing, and malware in real-time, using machine learning techniques. Below, we outline the key stages of implementation, detailing how each phase contributes to a robust and effective system that adapts to evolving cyber threats.[3]

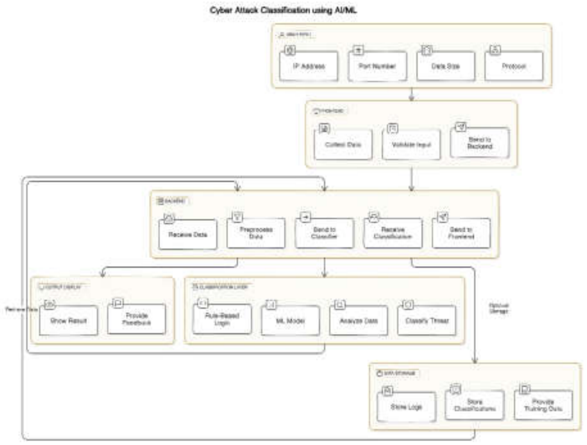

Fig1.Working Architecture

Data Collection:

Data collection is the foundational step of the cyberattack classification system, as high-quality data is crucial for the success of any machine learning model. In this project, the primary goal is to gather relevant network traffic data, including key parameters such as IP addresses, port numbers, packet sizes, and protocols used in communications. These inputs are critical for analysing network behaviour and identifying potential attacks. [13] There are several ways to collect network data for this system. One approach is to use publicly available datasets such as NSL-KDD and CICIDS2017, which contain labelled data on a variety of network attacks.[15] These datasets have been curated for research purposes and provide a solid baseline for training models to detect specific attack types. Another approach is to capture live network traffic from an organization's own infrastructure. This can be done using packet capture tools like Wireshark or tcpdump, which monitor and record network activity in real time. Using live traffic data has the added benefit of capturing the unique characteristics of an organization's network, ensuring that the model is well-suited for the specific environment it will protect.[9]

Data Preprocessing:

Once the raw data is collected, it must be pre-processed before being fed into machine learning algorithms. This step is essential for transforming the raw network traffic data into a format that is suitable for model training and classification. The first task in data preprocessing is cleaning the data. This involves removing irrelevant or duplicate data points, handling missing values, and filtering out any noise that might interfere with the classification process. Normalization is another crucial step. Since network traffic data can vary significantly in scale, with some features like packet size ranging from a few bytes to several megabytes, normalization ensures that all features contribute equally to the model's performance. This prevents features with larger scales from dominating the training process, which could skew the classification results. Labelling the data is also a key aspect of preprocessing. For supervised machine learning models, each data point needs to be labelled with the type of attack it represents (e.g., DDoS, SQL injection, phishing). [10] For live traffic data, this can be done by cross-referencing known attack signatures or patterns, or by using expert analysis to identify malicious behaviour. Ensuring that the data is accurately labelled is critical for training an effective model.

Feature Engineering:

Feature engineering is the process of selecting and constructing the most relevant attributes (features) from the raw network data to improve model performance. In this project, the system relies on key network parameters, such as IP address, port number, packet size, protocol type, and additional characteristics like session duration and packet frequency, to classify attacks. During this stage, domain expertise is used to identify the features that are most indicative of malicious behavior. For example, a high volume of small packets sent in rapid succession might indicate a DDoS attack, while unusual access to SQL databases could signal an SQL injection attempt. [10]Flow statistics, which summarize the behaviour of a connection or data flow, are also valuable features for distinguishing between normal and malicious traffic. To optimize performance, dimensionality reduction techniques such as Principal Component Analysis (PCA) may be applied to reduce the number of features. This helps prevent overfitting and speeds up the model training process by focusing on the most informative features. Dimensionality reduction ensures that the system can analyse incoming traffic efficiently, even under high network load.

Model Selection and Training:

Choosing the right machine learning model is a critical part of system implementation. Several types of models are suitable for the task of cyberattack classification, each with its strengths and trade-offs. Some popular models include decision trees, random forests, neural networks-mean. In this project, a combination of models may be used, depending on the specific type of attack being classified. Random forest models, for example, are well-suited for detecting known patterns in network traffic because they are robust to overfitting and can handle large datasets with many features. [5]Neural networks, particularly deep learning models, can be used to identify more complex attack patterns that might not be immediately obvious from simple rules. Once the models are selected, they are trained using the labeled datasets from the data collection phase. [15] This involves feeding the network traffic data into the models and adjusting the model parameters to minimize error rates in classification. The training process typically requires several iterations to ensure the models generalize well to unseen data. During this phase, techniques such as cross validation and grid search are used to fine-tune the models and identify the best-performing algorithms.

Real-Time Implementation:

With the models trained and validated, the next step is to integrate them into a real-time network monitoring system. This involves setting up a pipeline that continuously captures network traffic, preprocesses the data, and feeds it into the machine learning models for classification. The system should be capable of handling incoming traffic with low latency, ensuring that potential threats are detected and classified in real-time. One of the key challenges in real time implementation is minimizing the delay between traffic capture and threat classification. The system needs to process data quickly enough to alert administrators or automated response mechanisms to potential threats before significant damage can be done. Techniques such as parallel processing and hardware acceleration (e.g., using GPUs) can be employed to improve the speed and efficiency of the classification process.

Continuous Monitoring and Updates:

Cyberattack techniques are constantly evolving, so it is important that the system is designed to adapt over time. Continuous monitoring allows the system to identify new types of attacks that were not present in the initial training data. As the system encounters new forms of malicious behaviour, the data can be labelled and added to the training dataset, enabling the model to learn from these new examples. This ongoing process of updating the training data and retraining the models ensures that the system remains effective against emerging threats. Additionally, the system's performance needs to be monitored regularly to identify any areas where it may be underperforming. This can involve tuning model parameters, adjusting feature selection, or incorporating new machine learning techniques to improve accuracy.[6]

Alerting and Response Mechanism:

Finally, the system needs to include an alerting and response mechanism to notify administrators when an attack is detected. This can be done through automated alerts, such as email notifications or integration with security information and event management (SIEM) systems. The response mechanism may also include automated actions, such as blocking traffic from suspicious IP addresses or shutting down compromised systems to prevent further damage. The effectiveness of the alerting system depends on minimizing false positives while ensuring that legitimate threats are flagged. Too many false alarms can lead to alert fatigue, where administrators become desensitized to the warnings, potentially missing critical alerts. Therefore, the alerting mechanism must be fine-tuned to ensure it only triggers when there is a high likelihood of malicious activity. Implementing a cyberattack classification system using AI and machine learning involves multiple stages, from data collection to real-time monitoring and continuous updates.

4.Machine Learning Techniques

The implementation of an AI-based cyberattack classification system to enhance the system's ability to detect and classify attacks, a combination of techniques is applied, ensuring that both known and unknown threats are accurately identified. The primary models considered in this system include Decision Trees, Random Forests, K-Means Clustering, and Neural Networks, each of which plays a vital role in improving the system's accuracy, efficiency, and adaptability.

Decision Trees for Network Feature Classification:

Decision Trees are one of the foundational models in machine learning and offer a transparent, easy-to-understand approach to classification tasks. For cyberattack detection, Decision Trees are particularly useful in mapping out various network features such as IP addresses, port numbers, packet sizes, and protocols. These features are critical indicators of network behaviour, and Decision Trees help to break down this complex data into manageable steps. [14]

A Decision Tree works by splitting the data based on certain conditions or rules, ultimately classifying network traffic as either normal or malicious. At each node of the tree, a decision is made based on a particular feature of the network data. For instance, a node might evaluate whether a certain IP address falls within a known range of malicious addresses or whether the packet size is unusually large. Based on the outcome, the data is routed down a specific branch of the tree until it reaches a final classification (e.g., DDoS attack, phishing attempt, or normal traffic).

Random Forests for Enhanced Classification Accuracy: Random Forests build on the foundation of Decision Trees by using an ensemble learning approach. Instead of relying on a single Decision Tree, Random Forests generate multiple trees, each trained on a random subset of the data. By considering multiple trees, Random Forests reduce the likelihood of overfitting and provide a more generalized model that can perform well on new, unseen data. [11] In the context of cyberattack classification, Random Forests are particularly effective in handling noisy and imbalanced data, which is common in real-world network traffic. For example, a single Decision Tree might classify a particular type of traffic as normal, while another tree trained on different features might classify the same traffic as malicious.

This method not only increases classification accuracy but also ensures that the system can detect more complex patterns in network behaviour. Random Forests are particularly valuable in scenarios where there is a high degree of variability in the types of attacks encountered. [14] For instance, subtle differences in packet frequency or session duration might be difficult for a single model to detect, but a Random Forest can capture these nuances by leveraging the collective knowledge of multiple trees.

K-Means Clustering for Anomaly Detection:

K-Means Clustering is an unsupervised learning technique used to group data points into clusters based on their similarities. In the context of cyberattack classification, K Means Clustering can be applied to detect anomalous behaviours in network traffic that do not fit the typical patterns of normal operation. While Decision Trees and Random Forests focus on classifying known attacks, K-Means is useful for detecting new, previously unseen types of attacks that deviate from normal network behaviour. The process of K-Means Clustering involves dividing the data into a predetermined number of clusters.

Each cluster represents a group of network traffic with similar characteristics, such as packet size, protocol usage, or session duration. Once the clusters are defined, new incoming data points are assigned to the nearest cluster based on their features. If a data point does not closely resemble any of the existing clusters, it may indicate anomalous behaviour that warrants further investigation. 222320 VOLUME 10, 2024 For instance, K-Means Clustering can be used to identify sudden spikes in network traffic that might suggest the onset of a DDoS attack. Similarly, it can detect unusual access patterns to sensitive resources, which could indicate an SQL injection or insider threat. [8]

By continuously monitoring the clusters and flagging any outliers, the system can provide early warning signs of potential attacks. K-Means Clustering is particularly effective for detecting zero-day attacks, which are new, previously unknown threats that have no established signatures. Since these attacks cannot be identified by rule-based systems or supervised learning models, K-Means Clustering offers an additional layer of defence by focusing on abnormal behaviour rather than known attack patterns.

Neural Networks for Complex Attack Detection:

Neural Networks, particularly deep learning models, offer powerful tools for detecting more complex cyberattacks that may not be easily captured by traditional machine learning models. Neural Networks are designed to automatically learn from large amounts of data, making them particularly suited for identifying intricate patterns in network traffic. The architecture of a Neural Network consists of multiple layers of interconnected nodes (neurons), where each node attack based on packet size or IP address alone, a Neural Network can learn from the relationships between multiple features, such as how certain types of packets are transmitted together or how traffic patterns change over time. This makes Neural Networks particularly effective for detecting more sophisticated and stealthy attacks, which often blend into irregularities, or deviations in protocol usage.

Deep learning models excel at identifying attack patterns that are not immediately obvious from the raw data. For instance, while a traditional machine learning model might classify an attack based on packet size or IP address alone, a Neural Network can learn from the relationships between multiple features, such as how certain types of packets are transmitted together or how traffic patterns change over time. This makes Neural Networks particularly effective for detecting more sophisticated and stealthy attacks, which often blend into normal network traffic to avoid detection.

5. Experimental Evaluation:

This section defines the performance metrics in terms of intrusion detection and discusses the outcome by conducting experiments on cybersecurity datasets with different categories of attacks. If TP denotes true positives, FP denotes false positives, TN denotes true negative, and FN denotes false negatives, then the formal definition of below metrics
Are[7]:

$$\text{Precision} = \frac{TP}{TP+FP} \quad \ldots\ldots\ldots\ldots \quad (1)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad \ldots\ldots\ldots.. \quad (2)$$

$$\text{F-score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad \ldots\ldots.. (3)$$

$$\text{Accuracy} = \frac{TP + TN}{TP+TN+FP+FN} \quad \ldots\ldots..4)$$

Conclusion:

Cyber security is increasingly crucial as traditional systems struggle to detect new and polymorphic attacks. Machine learning (ML) techniques are becoming essential in enhancing security measures, evidenced by a growing body of research in the last decade. This paper presents a comprehensive survey of the intersection between ML and cyber security, focusing on applications like intrusion detection, spam detection, and malware detection for both computer networks and mobile devices.

It highlights that unique characteristics of cyber threats make it challenging for any single ML model to address all attack types effectively. Factors such as detection rate, time complexity, and accuracy must be considered when selecting models. The paper also introduces foundational concepts in both cyber security and ML, making it accessible to beginners.

Reference:

[1] Milenkoski, A., Vieira, M., Kounev, S., Avritzer, A., Payne, B.D.: Evaluating computer intru sion detection systems: a survey of common practices. ACM Comput. Surv. (CSUR) 48(1), 1–41 (2015)

[2] G. Karatas, O. Demir, and O. K. Sahingoz, Increasing the performance of machinelearning-based IDSs onanimbalancedandup-to-date dataset, IEEE Access, vol. 8, pp. 3215032162, 2020.

[3] N. Martins, J. M. Cruz, T. Cruz, and P. H. Abreu, Adversarial machine learning applied to intrusion and malware scenarios: A systematic review, IEEE Access, vol. 8, pp. 3540335419, 2020.

[4] Shailendra Singh, Sanjay Silakari "An Ensemble Approach for Cyber Attack Detection System: A Generic Framework" 14th ACIS, IEEE 2013. Pp 79-85.

[5] Xin, Y., et al.: Machine learning and deep learning methods for cybersecurity. IEEE Access 6, 35365–35381 (2018)

[6] X. Li et al., "Smart Community: An Internet of Things Application," IEEE Commun. Mag., vol. 49, no. 11, 2011, pp. 68–75.

[7]Witten,I.H.,Frank,E.,Trigg,L.E.,Hall,M.A.,Holmes,G.,Cunningham,S.J.:Weka:practical machine learning tools and techniques with java implementations (1999)

[8] T. Su, H. Sun, J. Zhu, S. Wang, and Y. Li, BAT: Deep learning methods on network intrusion detection using NSL-KDD dataset, IEEE Access, vol. 8, pp. 2957529585, 2020.

[9] Tong, W. et al.: A survey on intrusion detection system for advanced metering infrastructure, In: Sixth international conference on instrumentation & measurement, computer, communication and control (IMCCC), IEEE, Harbin, China, July 2016, pp. 33-37

[10] Abdulaziz Aborujilah1 and Shahrulniza Musa2 "Cloud-Based DDoS HTTP Attack Detection Using Covariance Matrix Approach" Hindawi Journal of

Computer Networks and Communications Volume 2017, Article ID 7674594, 8 pages

[11] S. T. Miller and C. Busby-Earle, Multi-perspective machine learning a classifier ensemble method for intrusion detection, in Proc. Int. Conf. Mach. Learn. Soft Comput. (ICMLSC), 2017, pp. 712.

[12]Sarker,I.H.,etal.:Cybersecuritydatascience:anoverviewfrommachinelearningperspective (2020)

[13] A. A. Abdulrahman, and M. K. Ibrahem, “Toward constructing a balanced intrusion detection dataset based on CICIDS2017,” Samarra J. Pure Appl. Sci., vol. 2, no. 3, 2020.

[14] Y. Xin, L. Kong, Z. Liu, Y. Chen, Y. Li, H. Zhu, M. Gao, H. Hou, and C. Wang, Machine learning and deep learning methods for cybersecu rity, IEEE Access, vol. 6, pp. 3536535381, 2018.

[15] J. a. H. Friedman, Trevor and Tibshirani, Robert, The elements of statistical learning vol.1: Springer series in statistics Springer, Berlin, 2001.