IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

Cybersecurity Threat Detection Using Machine Learning Technique

Prof. Ashwini Mahajan[1], Prof. Komal Naxine[2], Ms. Yashoda More[3]
Assistant Professor, Department of Computer Science and Engineering[1,2]
U.G. Student, Department of Computer Science and Engineering[3]
Tulsiramji Gaikwad-Patil Institute of Engineering & Technology, Mohgaon, Nagpur, Maharashtra, India
ashwini.cse@tgpcet.com, komal.cse@tgpcet.com, achalmore48@gmail.com

Abstract: *The ultramodern world has become fully reliant on cyberspace in all areas of everyday life. Cyber space operations are increasing day by day. moment, the world spends more time online compared to history. With this, the pitfalls of cyberattacks and cybercrimes are increasing. The word' cyber trouble' is known as an unlawful act carried out through the Internet. Traditional styles are unfit to identify zero-day attacks and advanced attacks. To date, mountains of machine learning styles have been created to identify cybercrimes and fight against cyber threats. The end of this exploration work is to put forward the analysis of some of the popular machine literacy styles employed to identify some of the most dangerous cyber pitfalls to cyberspace. We've made a terse overview to measure the performance of these machine literacy styles in the intrusion discovery, spam discovery, and malware discovery based on popularly used and standard datasets.*

Keywords: Cyber trouble; Cybercrime; Performance Evaluation; Machine Learning operation; Intrusion Discovery System; Malware Discovery; Spam Bracket

I. INTRODUCTION

The key threats in cybersecurity are botnets, intrusions, malware, and fileless malware. The malware is a harmful software that makes the programs do things that no program should do within the system. It includes all types of malware, such as Bots, Adware, backdoors, Trojans, Worms, and Fileless malware [14].

The fileless malware can infect a system without making any changes or small changes in the file system, existing wholly in the main memory, rootkit, or registry. The fileless malware is still full of features, such as hard- to-detect traditional malware and more dangerous [12]. A bot is a tiny malicious program that can attack an exploitable system to take over and integrate it into a larger botnet (Robot network) under the command of a bot-master. Abot-master can execute a cyberattack via the botnet, such as spamming, stealing data, hijacking confidential information, and distributed denial of service (DDoS)[11].In addition, intrusion can be categorized into any type of unauthorised or malicious activities in the network, such as network flood attack, malware with messages attached, and man-in-the-middle attack [13]. Any network needs to neutralize cyber threats so that it can protect its infrastructure. We have employed machine learning and deep learning methods to identify and classify cyber threats. We needed benchmark datasets with characteristics such as diversified attacks and imbalanced class distribution to train our models. Fine-tuning of hyperparameters was needed for each Model to train effectively. Our model successfully classifies the aforementioned cyberthreats.

II. MOTIVATION

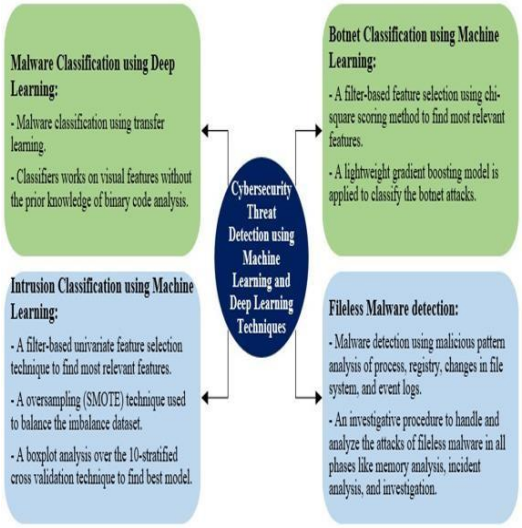
Malware and other attack tools are used by cybercriminals to exploit vulnerable machines. The traditional machine learning malware detection and classification methodologies employ static malware features for the training process. The static analysis extracted features are text-based, i.e., signature [7], Opcode sequence [9], control flow graph [4], bytecode [3], and n-gram [9]; thus, only a subset of malware sample data is used for the training process. Therefore, it will compromise the precision of the machine learning or deep learning models and is time- consuming as opposed to

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

the method that employs full information of malware samples in the form of visual features. The literature to identify malware based on visual features. The machine learning frameworks need to deal with malware datasets and the unavoidable task of feature engineering. Meanwhile, deep learning demonstrates potential outcomes to detect
malware images [1, 2, 5, 10, 15, 16]. Accordingly, we put forward a new MCFT-CNN model to solve the issues mentioned above and trained it on the visual features. We have gained 99.18% accuracy and 5.14ms prediction time on the MalImg dataset [6]. Our model outperforms the current state-of-the-art [14] considerably. Likewise, for intrusion detection as well as botnet detection, we have utilized machine learning algorithms to classify intrusions and botnet attacks efficiently. We have also suggested an incident response and handling process in the event of a fileless malware attack to examine the attack and activity of the fileless malware. The suggested models work much better than existing models in the literature [1–5, 7, 9, 10, 15, 16].

III. THESIS TIMELINE AND AUTHOR CONTRIBUTION

Thus, we propose a novel MCFT-CNN A model to address the above issues and trained with only visual features. Similarly, in intrusion detection and botnet detection, we have used machine learning algorithms to efficiently classify intrusions and botnet attacks. The proposed models perform significantly better than other models available in the literature [1–5, 7, 9, 10, 15, 16].
model in IoT networks (ABBD IoT) [Communicated]. Supervisor: supervises the work to improve the quality, investigates and validates the methods applied, and reviews the drafts to improve them. The thesis writing is in progress, expecting to complete the first and second chapters by October 2021, and submit the PhD thesis in March 2022.
Cybersecurity Threat Detection Using Machine Learning Technique



IV. LITERATURE REVIEW

Growing dependency on digital technologies has significantly increased the number of cybersecurity threats, which present enormous dangers to people, organizations, and infrastructure. Machine learning has been a promising method for the detection of cybersecurity threats, with enhanced accuracy and efficiency in detecting sophisticated attacks.
Current research has shown the ability of machine learning to detect cyber threats such as malware, phishing attacks, and intrusion detection.
Machine learning methods can inspect enormous amounts of information, spot anomalies and trends, and generate forecasts regarding emerging threats, so organizations can address these threats very fast and with a high level of effectiveness.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

Multiple machine learning approaches have already been used to cybersecurity threat detection, from supervised to unsupervised.

Supervised learning algorithm, e.g., support vector machines and random forests, was found to exhibit potential in recognizing known threats, whereas unsupervised learning techniques, e.g., clustering and anomaly detection, are applied in recognizing unknown threats. Deep learning algorithms, like convolutional neural networks and recurrent neural networks, have also been investigated for their application in the detection of threats in cybersecurity, providing higher performance in the recognition of complex threats.

The literature shows that machine learning is an asset when it comes to cybersecurity threat detection since it provides higher accuracy and efficiency in the detection of sophisticated threats. By using machine learning algorithms and methods, organizations can advance their threat detection, increase incident response, and remain ahead of the evolving threats. There are plenty of advantages to using machine learning in the detection of cybersecurity threats. Machine learning models can process vast volumes of data, detect patterns and anomalies, and predict potential threats, allowing organizations to respond efficiently and promptly to potential threats. Machine learning models can also change in response to new patterns of attack, thus becoming more efficient at detecting unknown threats.

Machine learning models also have some weaknesses, such as requiring huge sets of data and the risk of false positives. Despite these restraints, the literature indicates that machine learning is a promising technology for threat detection in cybersecurity, providing enhanced accuracy and efficiency in detecting sophisticated attacks. As the threat horizon keeps changing, the applications of machine learning in cybersecurity will grow more vital for organizations that want to safeguard their valuable data and assets. The literature review emphasizes the promise of machine learning in improving cybersecurity threat detection. Through the use of machine learning algorithms and methods, organizations can enhance their threat detection, minimize the likelihood of cyberattacks, and safeguard their sensitive data and assets. More research has to be conducted to investigate the uses and limitations of machine learning in cybersecurity, such as the creation of more efficient machine learning models and the combination of machine learning with current security systems

V. METHODOLOGY

Machine learning based cybersecurity threat detection with machine learning (ML) is an effective means to detect and mitigate cybersecurity. This is an overview of how it is achieved, delineated by methodology:

1. Data Collection To initiate development of a machine learning based system for detecting cybersecurity threats, a vast amount of data comprising network traffic, system logs, and security-related data should be collected. This data should consist of both benign and malicious activity so that the machine learning model can learn patterns and exceptions.

2. Data Preprocessing

Subsequent data is pre-processed so that it is in an appropriate format for machine learning analysis. Feature extraction entails the extraction of features from the data which are most representative of threats.

3. Feature Engineering

It is the process of choosing and converting raw data into features that are better suited for model training. This can involve methods like dimensionality reduction, feature scaling, and feature selection.

4. Model Selection

The second step is to choose a good machine learning algorithm for threat detection. Well- known algorithms are supervised learning methods like support vector machines (SVMs), random forests, and neural networks, and unsupervised learning methods like clustering and anomaly detection.

5. Model Training

The chosen machine learning model is subsequently trained on the pre-processed dataset.

6. Model Evaluation

It is measured using metrics including accuracy, precision, recall, and F1 score to determine the performance of the model in identifying threats.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

7. Model Deployment

After training and testing the model, it can then be used in a production environment to identify threats in real-time. The model can be built on top of current security systems, including intrusion detection systems (IDS) and security information and event management (SIEM) systems.

VI. CONCLUSION

The traditional security methods are not strong enough to deal with attacks and threats. Machine learning methods are doing their job at both ends: defender-end and attacker-end. We have shown a comparison of the performance of three learning models to identify and classify the intrusion, spam, and malware. We have taken into account commonly used and benchmark datasets to compare the results of the evaluation based on recall, precision, and accuracy. In the above section, we have explained and concluded that we cannot suggest a specific learning technique for each cyber threat detection. Various learning models are applied to distinct cyber threats. Conversely, there exists a large community of authors who have endeavoured to bring out the limitations of machine learning methods. We have seen and proposed that there exists a dire need for a recent benchmark dataset to validate the recent progress in the field of machine learning for cyber threat detection. Existing datasets fall short concerning diversity and complex attacks, and have missing values. There is a requirement for special and tailored learning models, especially developed for security needs. In the future, we will emphasize the study of more learning methods for cyber threat identification.

REFERENCES

[1]. Anderson et al. (2018) explored how hackers are starting to use machine learning the same way cybersecurity experts do — to learn, adapt, and launch smarter attacks. Their paper looked into how algorithms can be trained to bypass spam

[2]. filters or find vulnerabilities faster than traditional methods.

[3]. Papernot et al. (2016) introd uced the idea of “adversarial machine learning,” where attackers fool AI systems by slightly tweaking inputs. It’s like tricking a smart system with carefully crafted data, and it’s a big issue for things like image recognition and fraud detection.

[4]. Huang et al. (2011) gave a solid early warning about how AI and ML could become double- edged swords in the security world. They explained that the same models that detect threats can be reverse-engineered or manipulated to create new ones.

[5]. Tramèr et al. (2016) looked into what happens when attackers try to "steal" the way a machine learning model works sort of like copying a brain.

[6]. Rigaki & Garcia (2018) did a fascinating experiment where malware was trained to change its behavior in real time using ML, making it blend in with normal user traffic. It’s like malware learning how to camouflage itself while watching its surroundings.

[7]. Demetrio et al. (2021) focused on deep learning models used in malware detection and how they can be tricked with adversarial examples. Their research shows that even advanced detection systems can be manipulated with the right input tweaks.

[8]. Bhagoji et al. (2018) took it further by showing how attackers can actually poison a machine learning model — feeding it bad data during training to make it learn the wrong things. Imagine teaching a guard dog to let burglars in.

[9]. Kumar et al. (2020) looked at how social engineering and phishing attacks are now being personalized using machine learning, making fake emails and messages feel frighteningly real and targeted.