International Journal of Mathematics and Computer Research

ISSN: 2320-7167

Volume 14 Special Issue 03 March 2026, Page no. – 76-81

Impact Factor: 8.751

DOI: 10.47191/ijmcr/v14iSPC3.16

The Role of Machine Learning in Cyber Security

Bhagyashri A. Patel, Mizna M. Patel

Department of Computer Science, Dr. D. Y. Patil, Arts Commerce & Science College, Pimpri, Pune, Maharashtra, India

ARTICLE INFO

Published Online:
14 March 2026

Corresponding Author:
Bhagyashri A. Patel

ABSTRACT

Cyber security is like a moving target—hackers are always coming up with new tricks, so systems that just look for "known" issues can't keep up anymore. That's where machine learning (ML) comes in. Instead of following fixed rules, ML looks for anything unusual or suspicious in huge amounts of data. This means it can spot new types of threats much faster than traditional methods. In this paper, we explore how ML is being used for things like catching malware, stopping phishing attacks, and detecting intrusions in networks. We also talk about the tough parts, like needing lots of good data, powerful computers, and making sure attackers don't fool the ML systems. Our results show that while ML is a big help, it's not a silver bullet. The best protection happens when ML works together with traditional security tools and the know-how of human experts.

KEYWORDS: Machine Learning, Cyber Security, Threat Detection, Network Intrusion, Malware, Phishing, Anomaly Detection

INTRODUCTION

In this age, the cyberspace is growing faster as a primary source for a node-to-node information transfer with all its charms and challenges. The cyberspace serves as a significant source to access an infinite amount of information and resources over the globe. In 2017, the internet usage rate was 48% globally, later it increased to 81% for developing countries. The vast range of cyberspace encompasses a lot more than just the internet, including users, system resources, participant technical expertise, and much more. Additionally, the cyber sphere significantly contributes to the countless vulnerability to cyber threats and attacks. Cyber security is a collection of many strategies, tools, and procedures intended to protect cyberspace against threats and cyber-attacks. Cybercrimes are expanding more quickly than the current cyber security system in the modern world of computers and information technology. A computer system's vulnerability to threats can be attributed to a number of factors, including a weak system configuration, untrained staff, and a dearth of techniques. More progress must be made in creating cyber security techniques due to the expanding cyber threats. Attack strategies are advancing quickly to penetrate systems and elude generic signature-based defences, much as web and mobile technologies are doing the same. Due to their ability to quickly adapt to novel and unknowable circumstances, machine learning techniques present prospective answers that can be used to resolve such difficult and complex issues. Many different machine learning techniques have been successfully used to tackle a variety of issues in computer and information security. This paper covers and emphasizes several machine-learning applications in cyber security. Machine learning: One of the primarily used advanced methods for cybercrime detection is machine learning techniques. Machine learning techniques can be applied to address the limitations and constraints faced by conventional detection methods.

Cyber-attacks are getting smarter and sneakier, making them tougher to spot with old-school defences like firewalls or antivirus programs. Those tools mostly stop the threats we already know about, but hackers are always cooking up new tricks. Machine learning changes the game by letting computers "learn" from all kinds of data—like what users do, what's in system logs, or what's happening on the network. This way, the system can notice when something just doesn't look right. Unlike fixed rule-based systems, ML can adapt and get better over time.

This paper explores how ML is used in cyber security today. We begin by reviewing previous research, then examine the theory underlying ML approaches, and finally discuss how our study evaluates different ML models using real cyber security datasets.

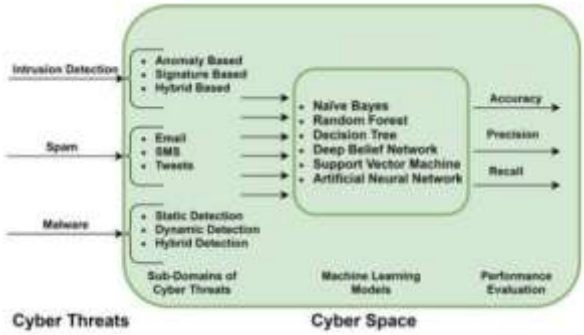

Figure 1: Cyber Threats in the Cyber Space

Cyber security's role as a machine learning capabilities and functions-

Machine Learning in Cyber Security

Cyber threats: Machine learning techniques are playing a vital role in fighting against cyber security threats and attacks such as intrusion detection system, malware detection, phishing detection, spam detection, and fraud detection to name a few. We will focus on malware detection, intrusion detection system, and spam classification for this review. Malware is a set of instructions that are designed for malicious intent to disrupt the normal flow of computer activities. Malicious code runs on a targeted machine with the intent to harm and compromise the integrity, confidentiality and availability of computer resources and services. Sad et al. in discussed the main critical problems in applying machine learning techniques for malware detection. Saad ET al.argued that machine-learning techniques have the ability to detect polymorphic and new attacks. Machine learning techniques will lead to all other conventional detection methods in the future. The training methods for malware detections should be cost-effective. The malware analysts should also be able to keep with the understanding of ML malware detection methods up to an expert level. Ambalavanan et al. in described some of the strategies to detect cyber threats efficiently. One of the critical downsides of the security system is that the security reliability level of the computing resources is generally determined by the ordinary user, who does not possess technical knowledge about security. Attacks such as replay, man-in-the-middle (MiTM), impersonation, credentials leakage, password guessing, session key leakage, unauthorised data update, malware injection, flooding, denial of service (DoS) and distributed denial of service (DDoS), among others, can be carried out against connected systems in the cyberspace. Therefore, in order to recognize and stop these assaults, we need some sort of security standard. Through the offered pre-processed dataset, the machine learning models (ML algorithms) may learn about various cyber assaults in the offline and online modes. The machine learning algorithms identify any indication of an incursion (a cyber-attack) in real time, or in online mode.

Cyber security

Over the last half-century, the information and communication technology (ICT) industry has evolved greatly, which is ubiquitous and closely integrated with our modern society. Thus, protecting ICT systems and applications from cyber-attacks has been greatly concerned by the security policymakers in recent days. The act of protecting ICT systems from various cyber-threats or attacks has come to be known as cyber security. Several aspects are associated with cyber security: measures to protect information and communication technology; the raw data and information it contains and their processing and transmitting; associated virtual and physical elements of the systems; the degree of protection resulting from the application of those measures; and eventually the associated field of professional endeavour. Overall, cyber security concerns with the understanding of diverse cyber-attacks and devising corresponding defines strategies that preserve several properties.

 Confidentiality is a property used to prevent the access and disclosure of information to unauthorized individuals, entities or systems.

 Integrity is a property used to prevent any modification or destruction of information in an unauthorized manner.

 Availability is a property used to ensure timely and reliable access of information assets and systems to an authorized entity.

II. LITERATURE REVIEW

A. Intrusion Detection

Many researchers have shown that ML can detect intrusions by flagging unusual behaviour in network traffic. For example, supervised and unsupervised models can both spot activities that don't match normal patterns, making them effective at catching attacks even without a known signature.

B. Malware Detection

Traditional antivirus software relies on signatures, which can't keep up with new or zero-day malware. Studies show that ML models can identify malware based on behaviour and file characteristics, allowing them to catch unknown threats with high accuracy.

C. Phishing and Spam

Email is still a common attack vector. ML techniques, especially natural language processing (NLP), have been applied to detect phishing by analysing email text, sender details, and metadata. These models outperform simple rule-based spam filters by adapting to new phishing tactics.

III. THEORETICAL FRAMEWORK

1. Supervised Learning

Uses labelled data to train models that can classify new inputs (e.g., normal vs. malicious). Popular algorithms include Support Vector Machines and Random Forests.

2. Unsupervised Learning

Doesn't require labelled data. Instead, it finds "normal" patterns in data and flags outliers as suspicious. Algorithms like K-means clustering are often used for anomaly detection.

3. Deep Learning

A more advanced branch of ML, using multi-layered neural networks to automatically learn complex patterns. Deep learning is effective for tasks such as image-based malware analysis and detecting subtle anomalies in traffic data.

IV. METHODOLOGY

Research Design:

We use a quantitative approach, testing ML models on well-known datasets like NSL-KDD and CICIDS2017.

Steps:

- **Data Collection:** Use benchmark datasets.
- **Model Building:** Train and test algorithms such as SVM, Random Forest, and Deep Neural Networks.
- **Evaluation Metrics:** Compare models using accuracy, precision, recall, and F1-score.

Variables:

- **Independent:** Choice of ML algorithm.
- **Dependent:** Model performance (accuracy, recall, precision, F1).
- **Control:** Same dataset and feature selection process across models.

Phishing Detection

Phishing is aimed at stealing personal sensitive information. Researchers [2] have identified three principal groups of anti-phishing methods: detective (monitoring, content filtering, anti-spam), preventive (authentication, patch and change management), and corrective (site takedown, forensics) ones.

Table 1: Principal Groups of Anti-Phishing Methods

Detective Solutions	Preventive Solutions	Corrective Solutions
Monitors account life cycle Brand monitoring Disables web duplication Performs content filtering Anti-Malware Anti-Spam	Authentication Patch and change management Email authentication Web application security	Phishing site takedown Forensics And investigation

A comparison of phishing detection methods is presented, and it is found that many of the solutions being considered for phishing detection have a high percentage of missed detection. Researchers compared six machine learning classifiers, including Logistic Regression (LR), Classification and Regression Trees (CART), Bayesian Additive Regression Trees (BART), Support Vector Machines (SVM), Random Forests (RF), and Neural Networks (Nets), using 1,171 raw phishing emails and 1,718 genuine emails. Here is a summary of the error rates for each of the classifiers listed above.

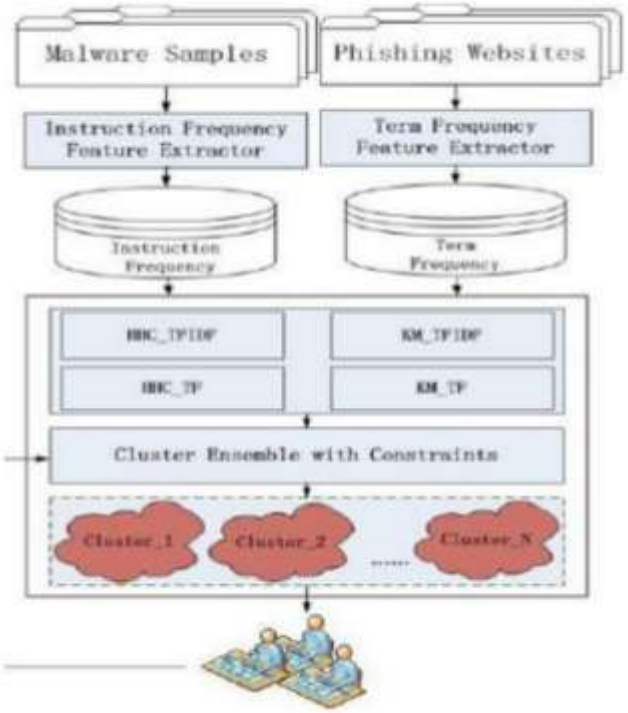

Figure 2: The Architecture of ACS

For experimentation, text indexing techniques were used for parsing the emails. All attachments were removed; "header information of all emails and html tags" from the emails' bodies as well as their specific elements were extracted. Afterwards, a stemming algorithm was applied and all the irrelevant words were removed. Finally, all items were sorted according to their frequency in emails. As a result of this work, it can be concluded that LR is a more preferable option among users due to low false positive rate (usually, users would not want their legitimate emails to be misclassified as junk). Also, LR has the highest precision and relatively high recall in comparison with other classifiers under contemplation.

Breaking Human Interaction Proofs

Chellapilla and Samar [16] discuss how the Human Interaction Proofs (or CAPTCHAs) can be broken by utilizing machine learning. The researchers experimented with seven various HIPs and learned their common strengths and weaknesses. The proposed approach is aimed at locating the characters (segmentation step) and employing neural network [17] for character recognition. Six experiments were conducted with EZ-Gimpy/Yahoo, Yahoo v2, mail blocks, register, Ticketmaster, and Google HIPs. Each experiment was split into two parts: (a) recognition (1,600 HIPs for training, 200 for validation, and 200 for testing) and (b) segmentation (500 HIPs for testing segmentation). On the recognition stage, different computer vision

techniques like converting to grayscale, thresholding to black and white, dilating and eroding, and selecting large CCs with sizes close to HIP char sizes were applied.

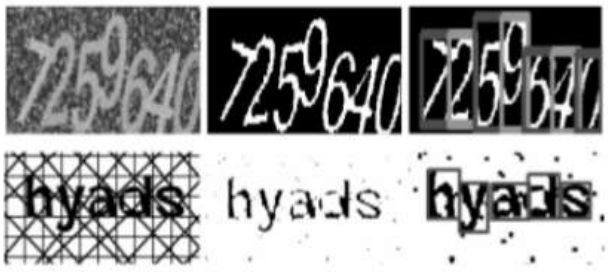

Figure 3: Types of Alphanumerical CAPTCHA

Intrusion Detection

Machine learning plays a significant role in cyber security intrusion detection. By leveraging its capabilities, machine learning algorithms can analyse large volumes of data, detect anomalies, and identify potential security breaches in real-time. Here's an overview of how machine learning is applied in cyber security intrusion detection:

DATA COLLECTION

Machine learning models require data to learn patterns and make predictions. In cyber security, relevant data sources include network traffic logs, system logs, user behaviour data, and security event information from various sensors.

Feature Extraction

Once the data is collected, relevant features need to be extracted to represent the characteristics of normal and abnormal behaviour. These features could include network traffic patterns, application usage, login activities, file access patterns, and more.

Model Training

The extracted features are then used to train machine learning models. Commonly used algorithms include decision trees, random forests, support vector machines, and deep learning techniques like neural networks. The models are trained on labelled datasets, where instances of normal and malicious behaviour are properly classified.

Anomaly Detection

After training, the machine learning models can identify deviations from normal behaviour. When deployed in a real-time environment, they continuously monitor network traffic and system logs, comparing incoming data against the learned patterns. Unusual patterns or behaviours are flagged as potential security threats or intrusions.

Alert Generation

When an anomaly is detected, the system generates an alert or notification to security analysts or administrators. The alert includes details about the detected anomaly, such as the type of intrusion, affected systems, and severity level. Analysts can then investigate and respond accordingly.

Model Adaptation

Cyber security threats are dynamic and constantly evolving. Machine learning models need to be regularly updated and retrained to adapt to new attack techniques. This involves incorporating new data, adjusting model parameters, and fine-tuning the algorithms to maintain their effectiveness over time.

Collaborative Intelligence

Machine learning models can benefit from collaborative intelligence, where multiple models or systems work together to enhance intrusion detection capabilities. By sharing information and insights, models can improve their accuracy and identify sophisticated attacks that may involve multiple stages or components.

Cyber security Advance Techniques

Machine learning has proven to be a powerful tool in bolstering cyber security defences by detecting and mitigating cyber threats. This article provides a comprehensive overview of various machine-learning techniques employed in cyber security, highlighting their capabilities and applications.

Anomaly Detection

Unsupervised Learning: Utilizing algorithms like k-means clustering, DBSCAN, or Isolation Forest to detect deviations from normal patterns and identify anomalous behaviour. One-Class Classification: Employing techniques such as support vector machines (SVM) or auto encoders to build models that classify instances as normal or anomalous.

Intrusion Detection

Supervised Learning: Training models, such as decision trees, random forests, or support vector machines, to classify network traffic as normal or malicious based on labelled datasets. Deep Learning: Utilizing deep neural networks, such as convolutional neural networks (CNN) or recurrent neural networks (RNN), to analyse network traffic and detect intrusions.

Malware Detection

Signature-based Detection: Using pattern matching techniques to compare file or code signatures against known malware signatures. Behaviour-based Detection: Employing machine learning models to analyse the behaviour of files or code and identify suspicious or malicious activities.

Threat Intelligence

Text Mining and Natural Language Processing: Extracting valuable information from textual sources, such as security reports or social media, to identify emerging threats or vulnerabilities Sentiment Analysis: Analysing sentiments expressed in cyber security-related data to gauge public opinion or detect potential risks.

User Behaviour Analytics

Sequential Pattern Mining: Identifying patterns in user behaviour sequences to detect abnormal or potentially malicious activities. Clustering and Profiling: Grouping users based on their behaviour characteristics and detecting deviations from their normal patterns. Adversarial Machine Learning: Generative Adversarial Networks (GANs): Employing GANs to generate adversarial examples and evaluate model robustness against malicious attacks. Defensive Distillation: Implementing techniques to make machine learning models more resilient against adversarial manipulation

Explainable AI in Cyber security

Interpretable Models: Developing machine learning models that provide transparent explanations for their decisions, enabling better understanding and trust in the system's outputs. Rule Extraction Techniques: Extracting human-readable rules from complex machine learning models to facilitate comprehensibility and explain ability.

Federated Learning for Privacy-Preserving Collaborative Security

Collaborative Threat Intelligence: Leveraging federated learning techniques to enable multiple organizations to share insights about emerging threats while preserving data privacy. Privacy- preserving Model Training: Training machine learning models on decentralized data sources without sharing raw data, thus maintaining data confidentiality.

Integration of Machine Learning with Big Data and Iota Security

Data Fusion and Analysis: Harnessing the power of big data to improve the accuracy and effectiveness of machine learning models in detecting and mitigating cyber threats. Secure Iota Ecosystems: Developing machine learning-driven solutions to enhance security measures and anomaly detection in vast Iota networks.

Context-aware and Adaptive Security

Context-aware Threat Detection: Developing machine learning models that consider contextual information, such as user behaviour, network conditions, and system configuration, to improve threat detection accuracy. Adaptive Defence Systems: Creating dynamic defines systems that can adapt and evolve in real-time based on changing threat landscapes, leveraging machine learning to detect and respond to new attack vectors.

Human-in-the-Loop Machine Learning

Augmented Threat Intelligence: Combining human expertise with machine learning algorithms to enhance threat intelligence capabilities, leveraging human insights for model training and validation. User-Centric Security: Incorporating user feedback and behaviour analysis to personalize security measures and provide proactive defines against targeted attacks.

Cyber security Challenges

Machine learning (ML) has been increasingly used in cyber security to detect and prevent various types of cyber threats. While ML offers numerous advantages, it also poses several challenges in the context of cyber security. Here are some key challenges associated with machine learning in cyber security:

Adversarial Attacks

Adversaries can attempt to manipulate or deceive ML models by exploiting vulnerabilities. Adversarial attacks include techniques like data poisoning, evasion attacks, and adversarial examples, where slight modifications to input data can mislead the ML model and compromise its effectiveness. Lack of labelled training data: Building accurate ML models requires large amounts of high-quality labelled training data. In the cyber security domain, obtaining such data can be challenging due to the limited availability of real-world cyber-attack data, as well as the difficulty in labelling it correctly.

Imbalanced Datasets

Cyber security datasets often suffer from class imbalance, where the occurrence of positive (attacks) and negative (normal) instances is disproportionate. Imbalanced datasets can lead to biased ML models that perform poorly in detecting minority classes or exhibit high false positive rates.

Interpretability and Explain Ability

Many ML algorithms, particularly deep learning models, are often considered "black boxes" due to their complex architectures. This lack of interpretability makes it difficult to understand the reasoning behind ML model decisions, hindering the ability to trust and explain their predictions, which is crucial in cyber security.

Concept Drift and Evolving Threats

The cyber security landscape is constantly evolving, with new threats and attack techniques emerging regularly. ML models trained on historical data may struggle to adapt to novel attacks or changing patterns, as they might not have encountered such instances during training.

Scalability and Performance

ML models in cyber security should be capable of handling large-scale, real-time data streams with low latency. Ensuring high performance and scalability can be a challenge, especially when dealing with computationally intensive ML algorithms or when operating in resource-constrained environments.

Privacy and Data Protection

ML models often require access to sensitive and private data for training and inference, raising concerns about data

privacy and compliance with regulations like GDPR. Protecting the confidentiality of user information and preventing unauthorized access to ML models and their training data is crucial. Addressing these challenges requires on-going research and development efforts to improve the robustness, resilience, and effectiveness of ML-based cyber security systems. Solutions may involve developing robust ML algorithms, designing resilient architectures, enhancing data collection and labelling techniques, incorporating explain ability methods, and adapting models to changing threats through continuous learning and monitoring.

V. CONCLUSION AND RECOMMENDATIONS

Our study confirms that ML significantly improves the ability to detect cyber threats compared to traditional methods. Deep learning, in particular, yields strong results in terms of accuracy and adaptability. However, ML is not fool proof. Models require high-quality training data and can be vulnerable to adversarial attacks.

RECOMMENDATIONS:

1. **Hybrid Defence:** Combine ML models with rule-based methods for layered security.
2. **Better Datasets:** Encourage sharing of diverse, updated datasets for training ML systems.
3. **Adversarial ML:** Invest in research to make models more resistant to manipulation.
4. **Human Oversight:** Keep security analysts in the loop—ML should assist, not replace human decision-making.

REFERENCES

1. Anti-Phishing Working Group, "Phishing and Fraud solutions". [Online]. Available: http://www.antiphishing.org/. [Accesses: April 4, 2013].
2. Bharadiya, J. P. (2023), "A Comprehensive Survey of Deep Learning Techniques Natural Language Processing", European Journal of Technology, 7(1), 58 - 66. https://doi.org/10.47672/ejt.1473
3. Bharadiya, J. P. (2023), "Convolutional Neural Networks for Image Classification. International Journal of Innovative Science and Research Technology", 8(5), 673 - 677. https://doi.org/10.5281/zenodo.7952031
4. Bharadiya, J. P., Tselios, N. T., & Reddy, M. (2023), "Forecasting of Crop Yield using Remote Sensing Data", Agrarian Factors and Machine Learning Approaches. Journal of Engineering Research and Reports, 24(12), 29–44. https://doi.org/10.9734/jerr/2023/v24i12858
5. Densham B. ,"Three cyber-security strategies to mitigate the impact of a data breach." Netw Secur. 2015;2015(1):5–8.
6. Hariri RH, Fredericks EM, Bowers KM. "Uncertainty in big data analytics: survey, opportunities, and challenges", J Big Data. 2019;6(1):44.
7. Knowledge Discovery and Data Mining group, "KDD cup 1999". [Online]. Available: http://www.kdd.org/kddcup/index.php. [Accessed: March 3, 2013].
8. L. F. Cranor, S. Egelman, J. Hong, and Y. Zhang, "Phinding phish: An evaluation of anti- phishing toolbars", Technical Report CMUCyLab-06-018, CMU, November 2006.
9. Nallamothu, P. T., & Bharadiya, J. P. (2023),"Artificial Intelligence in Orthopedics: A Concise Review." Asian Journal of Orthopaedic Research, 6(1), 17–27. Retrieved from https://journalajorr.com/index.php/AJORR/article/view/164
10. Qiao L-B, Zhang B-F, Lai Z-Q, Su J-S, "Mining of attack models in ids alerts from network backbone by a two-stage clustering method", In: 2012 IEEE 26th international parallel and distributed processing symposium workshops & Phd Forum. IEEE; 2012. p. 1263–9.
11. S. Abu-Nimeh, D. Nappa, X. Wang, and S. Nair, "A Comparison of Machine Learning Techniques for Phishing Detection", APWG eCrime Researchers Summit, October 4-5, 2007, Pittsburg, PA.