INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING, MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

Machine Learning in Cyber Security

Aradhya Desai*, Shraddha Khorgade

Department of Computer Science, Dr. D. Y. Patil Arts, Commerce & Science College Pimpri, Pune -18, Maharashtra, India

DOI: https://doi.org/10.51583/IJLTEMAS.2025.1413SP022

Received: 26 June 2025; Accepted: 30 June 2025; Published: 24 October 2025

Abstract — A component of artificial intelligence (AI), machine learning (ML) enables computers to learn from historical data, identify trends, and make judgments with little to no assistance from humans. Protecting computers, smartphones, servers, networks, and data from malicious attacks is the goal of cyber security. There are two ways that machine learning and cyber security can work together: by protecting machine learning systems and by leveraging machine learning to enhance cyber security. This combination has the potential to improve cyber security technologies, detect unknown and novel threats (known as zero-day attacks), and lessen the need for human intervention. Protecting critical data and systems becoming more difficult as technology advances quickly. In order to improve cyber security, this project intends to use machine learning to develop three distinct systems.

Keywords — Machine learning, Cyber security, Deep learning, Network, Attack

I. Introduction

Cyberspace has become a more important source for node-to-node information transport in the modern world. One important way to access a vast number of global resources and information is through cyberspace. In 2017, 48% of people worldwide used the internet; in developing nations, that number rose to 81%. Users, system resources, participant technical expertise, and much more are all part of the enormous array of cyberspace, which goes far beyond the internet. Furthermore, there are innumerable vulnerabilities to cyberthreats and attacks, which are greatly exacerbated by the cyber sphere. The goal of cyber security is to defend cyberspace from threats and cyber attacks by utilizing a variety of tactics, instruments, and protocols.

At the contemporary era of computers and information technology, cybercrimes are growing faster than the cyber security system. Numerous variables, such as inadequate system configuration, inexperienced personnel, and a lack of skills, might contribute to a computer system's susceptibility to threats. The growing cyber threats necessitate further advancements in cyber security measures.

Similar to how web and mobile technologies are developing rapidly, attack tactics are also evolving to breach systems and avoid general signature-based defenses. Machine learning approaches offer potential solutions that can be applied to such challenging and complicated problems because of their capacity to swiftly adjust to new and unknown situations. Numerous machine learning approaches have been effectively applied to address a range of computer and information security problems. Several machine-learning applications in cyber security are discussed and highlighted in this study. Learning by machine: Machine learning algorithms are among the most widely used sophisticated technologies for cybercrime detection. Conventional detection methods have limits and constraints that can be solved by applying machine learning techniques.

Figure 1: Cyber Threats in the Cyber Space

Machine Learning in Cyber Security

Similar to how web and mobile technologies are developing rapidly, attack tactics are also evolving to breach systems and avoid general signature-based defenses. Machine learning approaches offer potential solutions that can be applied to such challenging

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

and complicated problems because of their capacity to swiftly adjust to new and unknown situations. Numerous machine learning approaches have been effectively applied to address a range of computer and information security problems. Several machine-learning applications in cyber security are discussed and highlighted in this study. Learning by machine: Machine learning algorithms are among the most widely used sophisticated technologies for cybercrime detection. Conventional detection methods have limits and constraints that can be solved by applying machine learning techniques.

All of the other conventional techniques for detection are going to be surpassed by machine learning approaches. Cost-effective malware detection training techniques should be used. Additionally, the malware analysts should be able to maintain a high degree of expertise in ML malware detection techniques. In their article, Ambalavanan et al. discussed a few methods for effectively identifying cyber threats. One of the main drawbacks of the security system is that the average user, who has technical security understanding, typically determines the security reliability level of the computing resources.

Cyber security

The information and communication technology (ICT) sector, which is extensive and intricately entwined with our contemporary culture, has seen significant change over the past 50 years. As a result, security policymakers have recently been quite anxious regarding protecting ICT systems and applications against cyber attacks. Cyber security is the process of defending ICT systems against different cyber threats or attacks. Cyber security encompasses a number of elements, including information and communication technology protection measures, the raw data and information it contains, as well as their processing and transmission, related virtual and physical components of the systems, the level of protection that results from the implementation of those measures, and eventually the related field of professional endeavor

Confidentiality-A property called confidentiality is used to stop information from being accessed and revealed to unauthorized people, organizations, or systems.

Integrity- is a quality that prevents information from being altered or destroyed without authorization.

Availability- is a quality that provides an authorized entity rapid and reliable access to information assets and system

II. Methodology-

Machine learning techniques have become more valuable in the field of cyber security because they enable more effective threat identification and response. This article provides a full examination of the methods used to apply machine learning in cyber security, focusing on their benefits, drawbacks, and real-world applications

Data Collection and Preprocessing

Gathering and collecting cyber security datasets for model training and assessment is known as "acquisition of relevant data." Data Transformation and Cleaning: Preprocessing methods to deal with outliers and missing values while maintaining data quality. In order to improve model performance, feature extraction and engineering involves selecting informative features and developing new representations. [Bharadiya, J. P. 2023]

Model Selection and Evaluation

Algorithm Selection: Depending on the task and the properties of the data, selecting suitable machine learning algorithms, such as support vector machines, decision trees, or neural networks. Training and Testing: dividing the dataset into sets for training and testing, making sure sample sizes are suitable, and evaluating the generalization of the model. Performance Metrics: Calculating assessment metrics to gauge the models' efficacy, such as accuracy, precision, recall, F1-score, and area under the curve (AUC). [Bharadiya, J. P. 2023]

Model Training and Optimization

Model Training Techniques: Depending on the data's availability and labeling, supervised, unsupervised, or semi-supervised learning methodologies are used. Hyper parameter tuning: Using methods like grid search, random search, or Bayesian optimization to optimize model parameters in order to improve performance. Regularization and Over fitting Prevention: To avoid over fitting, strategies such as dropout, early halting, and L1 and L2 regularization are used

Deployment and Integration

Real-time monitoring involves integrating models into operational systems to enable ongoing observation and prompt reaction to online threats. Integration with Security Infrastructure: Applying machine learning models to firewalls and intrusion detection systems that are already in place. Model Maintenance and Updates: Putting in place systems to update models with fresh information and adjust to the evolving threat landscape

Spam-

In this method, a dataset of labeled emails—each classified as either spam or non-spam—is used to train a machine learning model. The machine learning algorithm picks up traits and patterns that set spam emails apart from authentic ones during training. These patterns may include particular terms or phrases that are frequently used in spam emails, the existence of particular kinds of

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

URLs or attachments, or traits of the email sender. After training, the model can be used in a production setting to evaluate incoming emails and determine if they are spam or legitimate. The subject line, sender's address, content, and other pertinent metadata are among the features that are evaluated by the model after being extracted from the email.

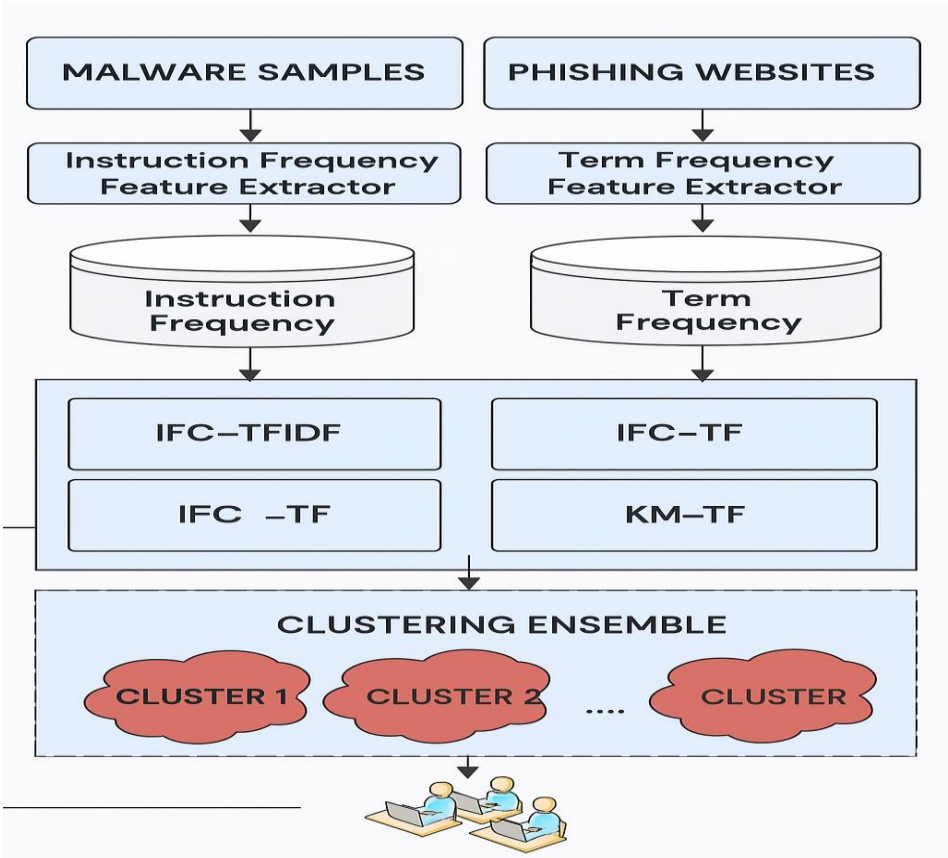

Figure 2: architecture of ACS

Emails were processed using text indexing algorithms for testing. The "header information of all emails and html tags" from the emails' bodies, together with their particular components, were extracted, and all attachments were deleted. After that, all of the superfluous words were eliminated using a stemming algorithm. Ultimately, every item was arranged based on how frequently they appeared in emails. Because of its low false positive rate, LR is a more user-favored alternative, according to this work (typically, consumers would not want their valid emails to be misclassified as garbage). Additionally, when compared to other classifiers under consideration, LR offers the highest precision and comparatively high recall.

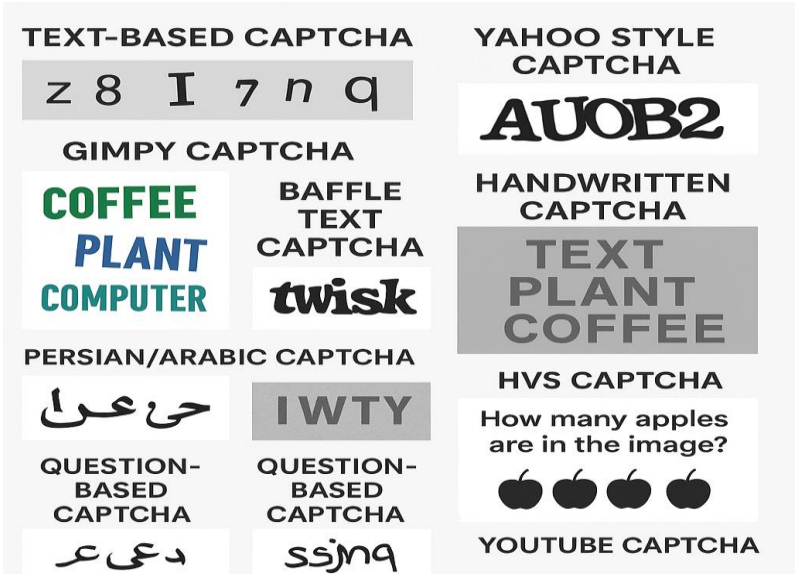

Figure 3: Types of Captcha

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

Intrusion Detection-

Cyber security intrusion detection heavily relies on machine learning. Machine learning algorithms can evaluate vast amounts of data, spot irregularities, and spot possible security breaches instantly by utilizing their skills. An outline of machine learning's use in cyber security intrusion detection is provided below. [Hariri RH 2019]

Data Collection-

In machine learning algorithms to identify trends and generate predictions, data is necessary. Network traffic logs, system logs, user activity data, and security event data from many sensors are all significant data sources in cyber security. [Bharadiya, J. P. 2023]

Feature Extraction-

To represent the traits of normal and aberrant behavior, pertinent features must be extracted from the data once it has been gathered. These characteristics could include file access patterns, login activity, program usage, network traffic patterns, and more.

Model Training-

Machine learning models are then trained using the features that were retrieved. Decision trees, random forests, support vector machines, and deep learning methods like neural networks are examples of frequently used algorithms. Labeled datasets, which accurately classify instances of malicious and normal behavior, are used to train the models. [Densham B 2015]

Intrusion Detection-

Supervised learning is the process of teaching models—like support vector machines, decision trees, or random forests—to identify malicious or benign network traffic using labeled datasets. Deep Learning: Analyzing network traffic and identifying intrusions by using deep neural networks, such as recurrent neural networks (RNN) or convolutional neural networks (CNN). [Hariri RH 2019]

Malware Detection-

Signature-based Detection: Comparing file or code signatures to known malware signatures using pattern matching algorithms. Behavior-based Detection: Using machine learning models to examine how files or code behaves in order to spot questionable or harmful activity. [Hariri RH 2019]

Human-in-the-Loop Machine Learning-

Integrating human insights for model training and validation, augmented threat intelligence combines machine learning algorithms with human knowledge to improve threat intelligence capabilities. User-Centric Security: Using behavior analysis and user input to tailor security protocols and offer preemptive protection against specific threats. [Bharadiya, J. P. 2023]

III. Conclusion

In order to sum up, machine learning methods are starting to prove to be quite helpful in the cyber security sector. Given the quick rise in cyber threats and attacks, traditional detection methods have proven inadequate in tackling the evolving nature of cybercrimes. Machine learning offers a solution by developing intelligent and automated systems that can analyze vast volumes of data, identify trends, and identify any security breaches instantly. Numerous machine learning applications in cyber security, including malware detection, intrusion detection, spam classification, and more, have been discussed in this article. These software applications employ machine learning techniques to enhance danger identification and response times. By being trained on labeled datasets, machine learning algorithms can detect cyber threats and attacks and learn to differentiate between benign and malicious activities. considerable effect on the performance of machine-learning models. It can be challenging to locate relevant and meaningful information, particularly considering how rapidly cyber threats are evolving. Machine-learning models must also be updated and retrained frequently to maximize their effectiveness, adapt to new attack techniques, and confirm their accuracy. Privacy and security concerns are also brought up by the use of machine learning in conjunction with big data and IoT a security. Data privacy and confidentiality must be maintained when using big data to increase the efficacy of machine learning models. Working together on threat intelligence while maintaining the privacy of raw data is now feasible because to the advent of techniques like federated learning.

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

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

4. Bharadiya, J. P., Tzenios, 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.