Malware Detection Using Artificial Intelligence: Techniques, Research Issues and Future Directions

Zahra Jabeen, Khushboo Mishra, Mohit Kumar Mishra, Binay Kumar Mishra

Abstract: *Artificial intelligence (AI) is an effective technology used for upgrading the security posture against a variety of security challenges and cyber-attacks that cyber security teams may use. Malware is a software which aims to access a device without the explicit permission of its owner. Forensics investigations report that many organizations have encountered unusual records, collected by their antiviral security monitoring systems. Most of their arrangements skeptically pass a large amount of diplomatic data through various unethical strategies that make malware identification tougher. However, these procedures have varied limitations that call for an unused inquiry about the track. This study explores the complex relationship between malware detection and AI [1]. This paper provides insights into performance evaluation metrics and discusses several research issues that impede the effectiveness of existing techniques. The study also provides recommendations for future research directions and is a valuable resource for researchers and practitioners working in the field of malware detection.*

Keywords: *Artificial Intelligence, Malware, Cyber Security, Antivirus, Cyber Attack*

I. INTRODUCTION

A collection of technological procedures and practices that guard data, hardware, software, and networks against damage, intrusion, and unauthorized access is referred to as cyber-security. Malware is a set of malicious programming codes or scripts and intrusive software that is built to destroy targeted computer systems and programs or mobile and web applications using different forms including computer viruses, worms, ransomware, rootkits, trojans, dialers, adware, spyware and keyloggers. The first computer-based virus was discovered on Apple II machines called “Elk Cloner” in 1982, developed by a 15-year-old high school student Rich Skrenta. Basit Farooq Alvi and Amjad Farooq Alvi who were two brothers and wanted to prove that PC is not immune, wrote a PC-based stealth virus called “Brain” in 1986 [2].

Manuscript received on 27 July 2024 | Revised Manuscript received on 02 August 2024 | Manuscript Accepted on 15 October 2024 | Manuscript published on 30 October 2024.
***Correspondence Author(s)**

Zahra Jabeen*, Department of Computer Science, Veer Kunwar Singh University, Ara (Bihar), India. Email ID: jabeen.zahra5@gmail.com, ORCID ID: 0009-0009-7194-0888

Khushboo Mishra, Department of Physics, Veer Kunwar Singh University, Ara (Bihar), India. Email ID: kmishra.j94@gmail.com

Mohit Kumar Mishra, Department of Electronics, Manipal University, Jaipur (R.J), India. Email ID: ermishramohit@gmail.com

Binay Kumar Mishra, Department of Physics, Veer Kunwar Singh University, Ara (Bihar), India. Email ID: drmishrabinay@gmail.com, ORCID ID: 0000-0001-8368-3633

These viruses were able to replicate with the use of floppy disks by inserting the infected. Cyber-attacks often target sensitive information and lead to financial losses which severely defames the impacted organizations. When Cyber-attacks target critical infrastructure like healthcare systems, they may even turn into life-threatening consequences. Artificial Intelligence (AI) helps in the advancement of malware detection and prevention while providing opportunities to develop robust, efficient and scalable malware recognition modules. Nonetheless, questions around data privacy, algorithmic accountability, and the potential for misuse have been existing around the corner, it is very clear that the adoption of these advanced technologies is not without growing areas of concern. It is very clear that the traditional techniques are not enough to protect us from the evolving landscape of cyber threats, and therefore turning towards more intelligent systems looks almost inevitable. Not just theoretically, it has real-world implications that can severely impact organizations, individuals, and even nations. AI's tailored user education approach also holds great potential for creating a workforce that is security-aware by customizing training materials to meet unique learning requirements and knowledge gaps This paper addresses the comprehensive report of the technical advancements and their practical applications, also engages critically with the limitations, challenges, and ethical considerations involved in adopting these new technologies for cyber security.

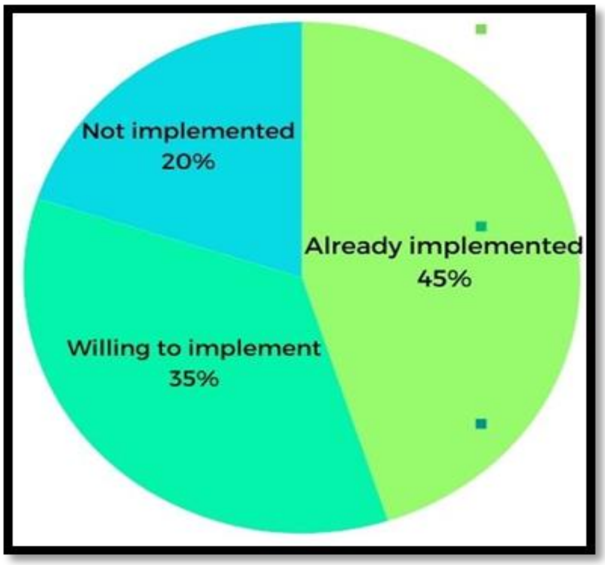

Fig.1. A Survey Report of Organizations' Implementation of MI and AI in their Cyber-Security Systems

II. LITERATURE REVIEW

Aggravation in the digital economy and infrastructure due to technological advancements results in a visible increase in cyber-attacks that could carry grave implications. The majority of cyber-attacks are made up of Malware attacks, accounting for 43% of the total, with 5.6 billion attacks [3]. Some of the Malware that has been discovered in the past are The Nimda, Morris Worm, Sasser, ILOVEYOU, Welchia, Melissa, Code Red, Slammer, Commwarrior-A, Crypto Locker, and Stuxnet. These computer viruses can propagate by download, malicious intent, installation of commercial software, or even by clicking a predefined link and thus affect any government, data center, commercial, laboratory, enterprise, or organizational software application. Identifying, preventing, detecting, addressing, and recording cyber-attacks to avoid future security breaches, advisory organizations like the National Institute of Standards and Technologies (NIST) are regularly advising to use of more proactive and adaptive approaches by moving towards real-time assessments, data-driven analysis, and continuous monitoring. According to the “2021 Sonic Wall Cyber Threat Report,” there was an increase in 62% of global ransomware attacks in 2020, with over 304 million ransomware attacks. Thus there is a need for effective cyber security measures, driving the development of innovative techniques such as Artificial Intelligence (AI) and Machine Learning (ML) algorithms. Distribution of malware from January 2022 to June 2022 was observed as 6% IoT malware, 19% crypto-jacking, 28% malicious intrusions, 53% malicious office and PDFs, 105% ransomware, and 167% encrypted threats.

The mapping of the state of AI-powered malware has also been performed to recognize AI-enhanced attacks carried out by malware. Malware types that conceal themselves from detection using AI techniques could also be identified, to get a better understanding of the maturity of those attacks, and to develop the algorithms and methods involved in those attacks.

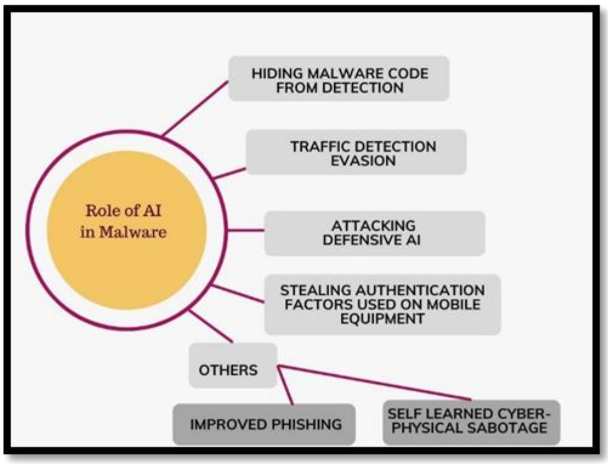
Role of AI in Malware
HIDING MALWARE CODE FROM DETECTION
TRAFFIC DETECTION EVASION
ATTACKING DEFENSIVE AI
STEALING AUTHENTICATION FACTORS USED ON MOBILE EQUIPMENT
OTHERS
IMPROVED PHISHING
SELF LEARNED CYBER-PHYSICAL SABOTAGE

Fig. 2.1. Uses of AI in Malware

According to the research done, we found out that there are several malware detection techniques employed by cyber security professionals, including some most common ones as follows:

A. Signature-Based Detection

This method relies on known digital indicators as a database of known malware signatures to identify suspicious behavior or threats. Whenever a software piece matches a signature in the database, the system alerts it as malicious. This technique is only effective for finding known malware and is not efficient with polymorphic malware. Therefore a list of indicators of compromise (IOCs), is supposed to be maintained in a database, which can eventually be used to identify a breach.

B. Static File Analysis

This technique refers to the examination of the code of a file, without running it, to identify any kind of malicious content. This evaluation of finding malicious objects could be done on file names, strings such as IP addresses, hashes, and file header data [4]. Proficient security teams are now using additional techniques to alert about advanced malware that might go unidentified during normal static analysis.

C. Dynamic Malware Analysis/ Sandboxing

This analysis is a closed system that enables security professionals to analyze and execute suspected malicious code in a safe environment called a sandbox [4]. Intel PT (Processor Trace) is applied in generation sandboxing to access the full execution flow of the potentially malicious artifact and analyze it using a complete “trace” alongside examining changes to virtual memory during execution. The process helps in studying malware without the risk of infection into their system or letting them escape into the enterprise network.

D. Heuristic Analysis

It does malware detection of new or modified malware by analyzing the behavior of a program or code that might not have a known signature. As soon as the software is guessed to exhibit characteristics similar to malware, it alerts it as a potentially malicious object.

E. Machine Learning and Artificial Intelligence

This technology can detect previously unknown threats and adapt to new malware variants while analyzing huge amounts of data and identifying patterns to classify them as favorable or malicious.

F. Checksumming/Cyclic Redundancy Check (CRC)

This technique offers calculation on a collection of data, such as a file, to confirm its integrity and can be effective for identifying corruption in data. CRC is one of the most common checksums used that involves analysis of both the position and value of a group of data.

G. Honeypots

This system is designed to mimic a software application that entices targets to malware. As soon as they get infected by the malware, security professionals can study it and design defenses to address these specific vulnerabilities or threats accordingly for their real system.

Retrieval Number: 100.1/ijeat.A453114011024
DOI: 10.35940/ijeat.A4531.14011024
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

A malware honeypot is similar to an application programming interface (API) that draws out malware attacks in a controlled and non-threatening environment.

H. Intrusion Detection and Prevention Systems (IDS/IPS)

Suspicious activities on network traffic are monitored and thus alert the administrators of potential security breaches, while detected threats are blocked by IPS. Both can be host-based (HIDS/HIPS) or network-based (NIDS/NIPS) [5].

I. Endpoint Protection Platforms (EPP)

Centralized management and protection of endpoint devices like desktops, laptops, and mobile devices are offered by EPP that includes providing anti-malware, antivirus with other security features like application control, device control, and data loss prevention (DLP).

To improve malware detection, Artificial Intelligence (AI) and Machine Learning (ML) techniques have been put in force, with deep learning being a popular approach to AI.

Advanced deep learning algorithms like Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) are trained on vast datasets, ultimately excelling in identifying and categorizing malware with the closest precision. Furthermore for enhancing training capabilities, generative models such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) are incorporated for creating high-quality synthetic data that mimics real-world patterns. Remarkable progress has been achieved by AI and ML in strengthening measures for network security by offering advanced capabilities to detect and counteract threats such as unauthorized network infiltrations, Distributed Denial of Service (DDoS) attacks, and Advanced Persistent Threat (APT) cyber-attacks with increased efficiency and accuracy [6].

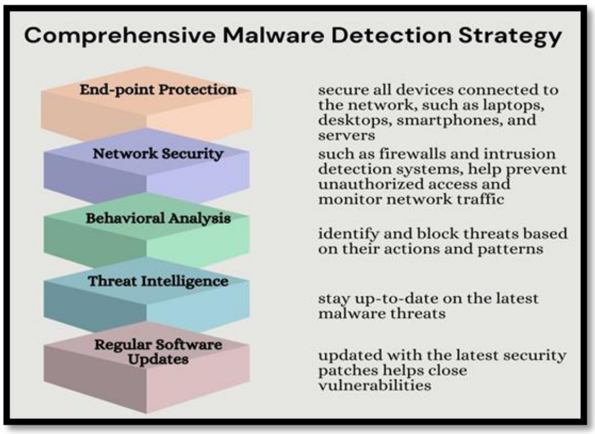
Comprehensive Malware Detection Strategy
End-point Protection
secure all devices connected to the network, such as laptops, desktops, smartphones, and servers
Network Security
such as firewalls and intrusion detection systems, help prevent unauthorized access and monitor network traffic
Behavioral Analysis
identify and block threats based on their actions and patterns
Threat Intelligence
stay up-to-date on the latest malware threats
Regular Software Updates
updated with the latest security patches helps close vulnerabilities

Fig.2.2. A Comprehensive Malware Detection Strategy Involves Multiple Layers of Protection

III. RESEARCH METHODOLOGY

In this study, we surveyed to investigate the current role of applications of AI and ML in malware detection for achieving cyber-security. A collection of various research papers, books, and conference proceedings was analyzed during this survey. We mainly focused on literature published between 2018 and 2024 to ensure coverage of only recent developments in the respective field. The selected literature was then thoroughly reviewed to extract information pertinent to some of our research questions which are as follows [7]:

- What are the major challenges for malware detection using AI?
- What are the developing technologies used by malware authors?
- How is sophisticated malware impacting static and dynamic analysis?
- What are the limitations of existing malware repositories?
- What features are optimal for training an AI model?
- Which AI models are most successful for malware detection and what are their advantages and limitations?

An extensive literature search was conducted in the ACM Digital Library, Google Scholar, IEEE Xplore, and Scopus database. ACM Digital Library provided access to Association for Computing Machinery (ACM) journals, proceedings, and conferences. Similarly, IEEE Xplore was chosen because it provides access to Institute of Electrical and Electronics Engineers (IEEE) conference papers and journals. Both ACM and IEEE are predominant in cyber-security. Scopus provided the most extensive database for the related subject. We also filtered publication topics including Systems and Data Security for Springer Link and Publication Topics for IEEE Explore and Computer Science and Security for Science Direct. Search keywords were mainly malware detection, artificial intelligence, and machine learning joined with offensive, network security, information security, and cyber-attack. An approx. of 235 studies were found during the initial search. Once the search processes were completed, we had to go through a screening process for finding relevant papers based on the paper title at first followed by reading and understanding the abstract and conclusion from screened papers. The results were limited by cutting off snowballing for articles older than 2018 to focus on the latest research in a rapidly evolving field. The articles included were needed to contain descriptions of malware based on machine learning or AI functionality and thereafter the literature was analyzed with respect to the application areas of AI and ML in cyber-security. Thorough analysis identified the most relevant AI/ML techniques and their applications were studied in areas such as intrusion detection, malware detection, security automation, network security, threat intelligence, security management, cyber-attack prediction, vulnerability management, and associated awareness education.

IV. RESEARCH ISSUES

The increasing adoption of AI and ML in cyber security has many associated challenges and barriers to its implementation [7]. Some of the most frequently reported challenges are said to be the lack of understanding of the technology, reported to be 36.9%, shortage of skilled personnel, reported to be 34%, and High costs as a significant barrier to adoption, with 29.1% of total surveyed organizations.

One of the most common issues is that machine learning might generate false positives or false negatives, which ultimately reduces its reliability and efficiency leaving out concerns for the researchers working in the respective field. Major actions that are needed to address challenges faced by this technology are as follows:

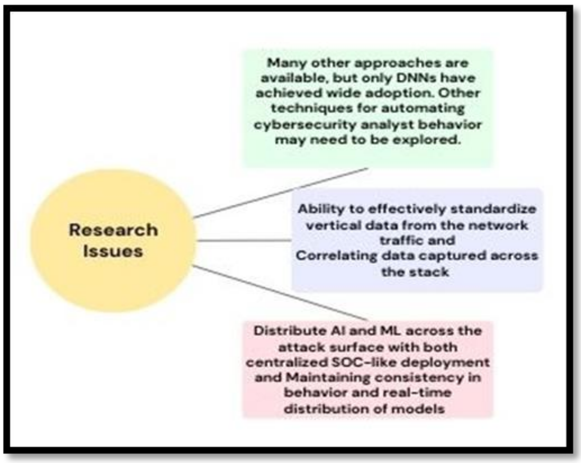
Many other approaches are available, but only DNNs have achieved wide adoption. Other techniques for automating cybersecurity analyst behavior may need to be explored.
Research Issues
Ability to effectively standardize vertical data from the network traffic and Correlating data captured across the stack
Distribute AI and ML across the attack surface with both centralized SOC-like deployment and Maintaining consistency in behavior and real-time distribution of models

Fig.3. Research Issues Faced by Ai in Detecting Malware

V. CONCLUSION AND FUTURE WORK

Artificial Intelligence (AI) and Machine Learning (ML) have already begun to reshape the cyber-security landscape and have the potential to revolutionize the field in multiple dimensions. Windows-based systems are the major targets of malware by cyber attackers. Techniques that are most effective for detecting malware attacks have to be Machine learning and deep learning. By understanding different types of malware accompanied by the engagement of a multi-layered approach to security, organizations can remarkably limit the risk of falling victim to a malware attack. However, AI could be less effective if it relies only on historical data, which might reduce its impacts and further prevent it from adapting to innovative attack methods. Artificial intelligence still is dependent on human knowledge or interference since it finds it difficult to envision contextual differences and can be mistaken in interpreting user behaviors and intentions. Ultimately, despite AI's tremendous potential, overcoming its present limitations requires careful balancing. While focusing on the future of AI and ML in cyber-security, several fascinating research paths are emerging [8][9][10][11][12][13]. The other advanced technologies like blockchain and quantum computing when integrated with AI and ML will provide the potential to promise an avenue for further exploration. Future work must also delve into the ethical considerations, which would aim to formulate strategies for responsible usage and transparent decision-making processes while not just focusing entirely on the technical aspects. Incident response, proactive threat hunting, and disaster recovery are areas that are currently ripe for AI and ML applications, which are opening new frontiers in cyber-security measures. AI's processing power has made it feasible to identify potential threats in advance, and its tailored advice promotes a Cyberwar culture. Thus it is very evident that progress in this field is not going to happen without encountering its challenges. Biases, adversarial flaws, and false positives can undermine effectiveness and confidence. The right mix between AI's advantages and human abilities must be found to optimize its benefits and minimize its disadvantages. The development of artificial intelligence (AI) in cyber-security has been examined in various roles, categories of solutions, particular use cases, and AI approach types.

DECLARATION STATEMENT

Funding	No, I did not receive it.
Conflicts of Interest/ Competing Interests	No conflicts of interest to the best of our knowledge.
Ethical Approval and Consent to Participate	No, the article does not require ethical approval and consent to participate with evidence.
Availability of Data and Material	Not relevant.
Authors Contributions	All authors have equal participation in this article.

REFERENCES

1. Daniel Gibert, Carles Mateu, Jordi Planes., The rise of machine learning for detection and classification of malware: Research developments, trends and challenges, Journal of Network and Computer Applications Volume 153 , 1 March 2020, 102526 https://doi.org/10.1016/j.jnca.2019.102526
2. Kaspersky: A Brief History of Computer Viruses & What the Future Holds
3. Gary Smith, April 10, 2024 : +95 Cyber Security Breach Statistics 2024, station
4. Kurt Baker, Malware Analysis, April 17, 2023 : crowdstrike
5. Perception Point : Malware Detection: 7 Methods and Security Solutions that Use Them
6. Mohamed, Cogent Engineering (2023), 10: 2272358https://doi.org/10.1080/23311916.2023.2272358 https://doi.org/10.1080/23311916.2023.2272358
7. Matthew G. Gaber, Mohiuddin Ahmed, and Helge Janicke. 2024. Malware Detection with Artificial Intelligence: A Systematic Literature Review. ACM Comput. Surv. 56, 6, Article 148 (January 2024), 33 pages. https://doi.org/10.1145/3638552
8. Abdullahi, M., Baashar, Y., Alhussian, H., Alwadain, A., Aziz, N., Capretz, L. F., & Abdulkadir, S. J. (2022). Detecting cybersecurity attacks in internet of things using artificial intelligence methods: A systematic literature review. Electronics, 11(2), 198. https://doi.org/10.3390/electronics11020198
9. Gupta, S., Sabitha, A. S., & Punhani, R. (2019). Cyber Security Threat Intelligence using Data Mining Techniques and Artificial Intelligence. In International Journal of Recent Technology and Engineering (IJRTE) (Vol. 8, Issue 3, pp. 6133–6140). https://doi.org/10.35940/ijrte.c5675.098319
10. R .Sri Devi, M. Mohan Kumar, Cyber Security Affairs in Empowering Technologies. (2019). In International Journal of Innovative Technology and Exploring Engineering (Vol. 8, Issue 10S, pp. 1–7). https://doi.org/10.35940/ijitee.j1001.08810s19
11. Saudi, M. M., Sukardi, S., Abd Aziz, N. A. A., Ahmad, A., & Husainiamer, M. 'Afif. (2019). Malware Classification for Cyber Physical System (CPS) based on Phylogenetics. In International Journal of Engineering and Advanced Technology (Vol. 9, Issue 1, pp. 3666–3670). https://doi.org/10.35940/ijeat.a2711.109119
12. Joshma K J, & Sankar P, V. (2024). Phishing Website Detection. In Indian Journal of Data Mining (Vol. 4, Issue 1, pp. 38–41). https://doi.org/10.54105/ijdm.a1642.04010524
13. Rathore, R., & Shrivastava, Dr. N. (2023). Network Anomaly Detection System using Deep Learning with Feature Selection Through PSO. In International Journal of Emerging Science and Engineering (Vol. 11, Issue 5, pp. 1–6). https://doi.org/10.35940/ijese.f2531.0411523

AUTHORS PROFILE

Zahra Jabeen is a Research Scholar in the University Department of Computer Science from Veer Kunwar Singh University, Ara, Bihar, India. Her work, published in renowned Scopus and UGC journals, has generated critical discourse and earned prestigious accolades. She has completed her Bachelor of Technology (B. Tech) and Master in Technology (M. Tech) in Computer Science & Engineering.

Khushboo Mishra is a Research Scholar in the P.G Department of Physics from Veer Kunwar Singh University, Ara, Bihar, India. She is determined to bring some positive advancements in the society with research findings in her work.

Mohit Kumar Mishra is a Research Scholar in the Department of Electronics and Communication from Manipal University, Jaipur, Rajasthan, India. He aims at proving connections among technical advancements of physical world with the mythological background and has wide knowledge of vedic physics.

Binay Kumar Mishra is working as a Professor in P.G Department of Physics, Veer Kunwar Singh University, Ara, Bihar, India. He has qualified in MS. c., Ph.D , Physics with specialisation in Plasma Physics, Nano Flow and IoT. He has an educational experience of more than 29 years.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of the Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)/ journal and/or the editor(s). The Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP) and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.