ISSN: 2832-790X
DOI: 10.33552/OJRAT.2024.03.000551
Online Journal of Robotics & Automation Technology
Iris Publishers

Research Article

AI/ML, Data Science, and Automation in Cybersecurity: Methods and Applications in Healthcare

Cheryl Ann Alexander[1]* and Lidong Wang[2]

[1]*Institute for IT Innovation and Smart Health, Mississippi, USA*

[2]*Institute for Systems Engineering Research, Mississippi State University, Mississippi, USA*

***Corresponding author:** Cheryl Ann Alexander, Institute for IT Innovation and Smart Health, Mississippi, USA

Received Date: May 10, 2024

Published Date: June 06, 2024

Abstract

The support of artificial intelligence (AI)/machine learning (ML), data science, and automation in cybersecurity are introduced in this paper, respectively. Their specific methods or tools in cybersecurity are discussed. The applications of AI/ML, data science, and automation in healthcare cybersecurity are presented as case studies, respectively. Both defenders and cyber criminals have access to AI/ML, data science, and automation. There are challenges in practicing robust cybersecurity. It is significant to maintain strict cybersecurity policies and use advanced or updated technologies for the detection, prevention, or mitigation of cyberattacks and cyber incidents. Healthcare depends on cybersecurity to keep personal health information (PHI) safe because healthcare data is either in motion or at rest. Cybersecurity in healthcare then becomes a primary concern to meet HIPAA standards and hospital standards. Healthcare information travels more often than ever before as providers are using mobile devices, telemedicine, and mobile imaging and diagnostics.

Keywords: cybersecurity; artificial intelligence (AI); machine learning (ML); deep learning (DL); data science; cybersecurity automation; healthcare

Introduction

Detecting threats and protecting systems and their data sources, including intrusion detection/prevention systems, identity and access management, fraud detection/anti-fraud, data loss prevention, antivirus/antimalware, and risk and compliance management is a function of AI-based tools. There are challenges in the training of artificial intelligence (AI)/machine learning (ML) models for complex data due to

1. data from various sensors (possible problems in data fusion),
2. noisy data streams, and
3. data with various modalities.

Data science has the potential in cybersecurity. It offers tools for synthesizing a high volume of data quickly, detecting unforeseen patterns, etc. The smart adversary is a new threat model, in which an adversary employs sophisticated techniques for attacks. The attacks can target training data, testing data, and model parameters in AI/ML. The smart adversary can erode confidence and trust in an AI/ML system, target a specific class for misclassifications, and evade automatic detection by obfuscation. ML should utilize various feature vectors and objective functions to minimize the impacts of the smart adversary [1].

The utilization of an automated method for cybersecurity threat intelligence was studied. The selected use case is an international leading organization in cybersecurity, demonstrating new dynamic

ways to support decision-making at all levels (operational, tactical, and strategic) while being under attacks. An integrated architecture was developed that combines cyber threat intelligence (CTI) and dynamic risk assessment and management (DRA/DRM). It is based on ontologies, Semantic Web Rule Language (SWRL) rules, and the utilization of a reasoner [2]. Table 1 [3] lists some attacks, descriptions, and their targeted objects. Table 2 [4] shows some risky sources (data, devices, and technologies) in healthcare, associated vulnerabilities, current risk management, and recommended strategies.

Table 1: Attacks and targeted objects.

Attacks	Descriptions	Targeted Objects
Phishing and social engineering attacks	Utilizing emails or social-engineering methods to get confidential information	Databases
Malvertising	Injecting malware-laden or malicious advertisement into legitimate advertising webpages or networks	Databases
Watering hole attacks (phishing)	Directly attacking a specific group, leading to the impersonation of the attacker to the daily work site of the group	Databases
System misconfiguration exploitation	Detecting the flaws or faults of unpatched yet software and exploiting them to compromise a system	Technical equipment
Vulnerability exploitation	Utilizing software bugs or an unaddressed to take control of them through a dedicated program	Technical equipment
3rd party vendors (backdoor)	Getting into the internal, corporate networks utilizing back doors, embedded with specific functions or programmable implemented features in devices in the internal network	Technical equipment
Man-in-the-Middle (network spoofing)	An attacker plays between the application and the user, intercepting data packets or impersonating a page.	Internal network, ICT (information and communications technology) equipment

Table 2: Risky sources, associated vulnerabilities, current risk management, and recommended strategies.

Aspects	Details/Examples
Risky data, devices, and technologies	Electronic data, medical devices, and tele-medicine
Associated vulnerabilities	Out-of-date systems
	Rapid innovation
	Internal threats
	Interoperability
	Constant accessibility
	Focus on medical care
	Lack of regulation
	Lack of resources
Current risk management	Detection and response
	Technical measures
	Regulatory measures
	Devices requirements
	Insurance
Recommended strategies	Technical measures
	Build into it, incorporating cybersecurity as an integral element
	Group efforts
	Risk management
	Training

The objective of this paper is to introduce the methods of AI/ML, data science, and automation in cybersecurity, and discuss their applications in cybersecurity of healthcare. The subsequent sections of the paper are organized as follows: the second section, the third section, and the fourth section introduce AI/ML, data science, and automation in cybersecurity, respectively. The fifth section, the sixth section, and the seventh section present AI/ML, data science, and automation in cybersecurity of healthcare, respectively. The eighth section is the conclusion.

AI/ML in Cybersecurity

AI continuously monitors network traffic, user behaviors, and system anomalies. AI swiftly recognizes unusual patterns, revealing cyberattacks. An instant response can be proactive for threat projections and can foster an immediate response, which can prevent a breach. Furthermore, new threats can teach AI-powered cybersecurity to adapt and learn when new threats arise. Predictive analytics can forecast possible vulnerabilities and further recommend actions. AI-powered cybersecurity can also understand and learn new threats. AI uses predictive analysis and AI-powered cybersecurity tools that can estimate many vulnerabilities and therefore recommend or predict actions. A comprehensive view of vulnerabilities and potential actions to avoid threats leads AI to correlate the data from a multitude of sources, which can provide a comprehensive point of view for potential threats. This becomes a holistic approach enabling early detection of any advanced or muti-staged attacks. AI streamlines administrative activities, improving efficiency.

Cybersecurity AI acts by analyzing data stream within the system and constructing some interpretation of normal or abnormal for users, computers, or other devices. However, cybersecurity AI is very competent in recognizing the difference between normal network activities and the works of a malicious actor, but AI can issue an immediate response to stop the spread of an attack. AI applications in cybersecurity, healthcare, etc. were introduced. A SWOT analysis of AI is shown in Table 3 [5]. Intelligence, security, risks/threats, privacy, etc. are included in the table.

Table 3: SWOT analysis of AI.

Strengths	Weaknesses
• Intelligence • Smart AI apps • Daily applications • Reduction of errors • Unbiased decision • Available 24/7 • Adaptability • AI utilized in risky situations • Limitless functions (depending on programming)	• Risk of losing data • Computation issues • Lack of efficient algorithms • Threats • Fewer than ideal samples for algorithm development • Lack of ability to think for oneself
Opportunities	**Threats**
• Protect the privacy on sensitive data • Development of novel tools, reducing the complexity • Uplift the 3D immersive experience • Granting funds from various sources • Improvement in the performance, reducing the training time, and enhancing robustness with existing AI models	• Security threats of production AI • Privacy issues • Misuse • Cyber-syndrome • Legal risks • Personal data abuse (ethical issues) • AI is used to do terrible things (e.g., lethal autonomous weapons)

Malicious actors also use AI to design and execute attacks; ransomware, rapid exploitation of vulnerabilities, developing phishing emails, deep target reconnaissance, developing complex malware code, automated attacks, etc. Cyber criminals have developed complex malware codes, automation of attacks, etc. Malicious actors have also multiplied AI capacity, including human impersonation, password guessing, building better malware, penetration testing tools, and stealth attacks. AI-generated phishing emails may be opened at a higher rate due to the ability of AI to target users and recognize patterns. For example, ChatGPT, an AI-powered chatbot can be used in the development of malware and ransomware.

An extensive view of ML algorithms was provided, focusing on how they can be used for intelligent data analytics and automation in cybersecurity. Potential real-world use cases were explored where automation, data-driven intelligence, and decision-making enable more proactive cyber protection than traditional methods. Various ML methods used in cybersecurity are shown in Table 4 [6].

Table 4: Various ML methods and their tasks in cybersecurity.

Methods	Examples of Tasks
k-nearest neighbors (KNN)	Creating an intrusion detection system
	Reducing the false alarm rate
K-means and KNN	Creating an intrusion detection system
Support Vector Machine (SVM)	Selecting features of security, detecting & classifying intrusions
	Classifying cyberattacks such as probing and DoS
Association Rule	Creating an intrusion detection system (IDS)
Random Forest (RF)	Detecting DoS
	Detecting cyber anomalies
	Intrusion detection system
Decision Tree	Selecting features of security, creating an IDS
Naive Bayes Classifier (NBC)	Detecting anomalies
Principal component analysis (PCA)	Processing security data with high dimensionality
Ensemble learning	Detecting cross-site scripting (XSS) attacks
Q-Learning	Detecting distributed denial-of-service (DDoS)
Deep learning: convolutional neural network (CNN)	Classification of malware traffics
Deep learning: LSTM, recurrent neural network (RNN)	Detecting/classifying anomaly intrusions & attacks
Multi-CNN	Constructing an IDS
CNN + long short-term memory (LSTM)	Detecting/mitigating phishing & Botnet attacks

Data Science in Cybersecurity

Data science is a multidisciplinary field, combining math, statistics, computer science, AI/ML, specialized programming, sophisticated analytics, and domain expertise to extract data and expose valued insights from the data. By using various tools and techniques, data can be studied and analyzed in massive volumes, finding unseen patterns, and drawing meaningful insights from the data. Listed below are some of the major applications of data science in cybersecurity:

a. ML is a data science tool used for anomaly detection where ML algorithms are used to analyze massive amounts of datasets and identification of any abnormal patterns or behaviors.

b. ML applies predictive cybersecurity as ML models may be taught to predict prospective cyber threats before they occur.

c. Data science has a critical role in analyzing the scope of a security incident response and identification of any compromised data. ML models can assist in tracing the origin of the attack and assist in mitigating the damage.

How cybersecurity data science is applied to the data-driven process for intelligent decision-making in smart cybersecurity services and systems was studied. A generic multi-layered framework of a cybersecurity data science model was developed based on ML. In the framework, data are captured from various sources, and data analytics complement the latest data-driven patterns to provide smart cybersecurity solutions. The framework is shown in Figure 1 [7].

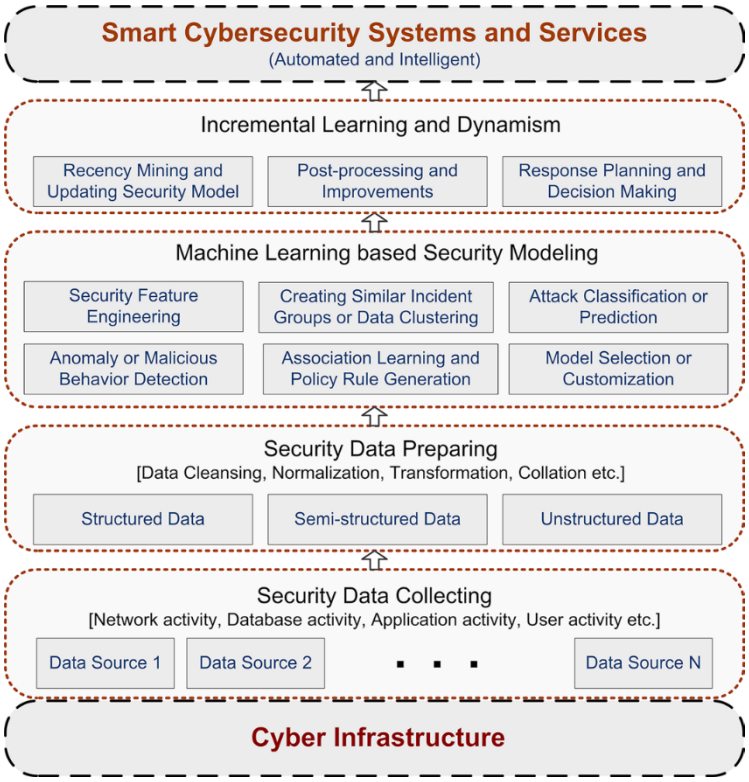

Figure 1: An ML-based framework for smart cybersecurity.

Automation in Cybersecurity

The integration of automated infrastructure management (AIM) into cybersecurity was discussed. AIM's automation of infrastructure management processes improves the effectiveness and speed of cybersecurity measures [8]. A risk analysis method was developed, and major steps were followed that included deciding the automation level, finding cyberattack targets, discovering cyberattack methods, defining cyberattack consequences, and performing risk ratio evaluation.

The cybersecurity automation system is an advanced system driven by AI/ML, involving the automation of cybersecurity procedures so that they are faster and work more effectively. Streamlining manual and often arduous tasks, cybersecurity automation systematizes the workflow. Because cybersecurity automated systems can analyze real-time data, they provide efficient protection against cyberattacks and require only the implementation of automated systems. However, automated compliance observation is utilized to monitor networks and systems for compliance with standards and regulations in healthcare, which helps identify and handle any potential compliance problems. Automated intelligence collection, penetration testing, and AI/ML, etc. are excellent models of automation in cybersecurity.

Through the conduction of malware analysis, detection of data exfiltration, implementation of vulnerability scanning technology, and blocking of common installation paths, defenders can permanently leverage automation. However, malicious actors can still use automated software to recognize invaluable data such as credit cards and passwords. Although most automated cyber intelligence tools were designed to secure systems, embracing systems for security monitoring or alerting, network intrusion detection and prevention, and vulnerability management, is also suggested. Various types of security automation tools include:

a. Security information and event management (SIEM) tools

b. Security orchestration, automation, and response (SOAR) tools

c. Vulnerability management tools-Automatically scan IT resources for vulnerabilities, recognize flaws, classify them, prioritize the risks, and propose remediation activities.

d. Endpoint protection tools-The endpoints include network connections, Internet of Things (IoT) devices, cloud-based applications, PCs, etc. Major categories of endpoint protection tools contain anti-malware solutions, response software, endpoint detection, anti-malware solutions, etc.

AI/ML in the Cybersecurity of Healthcare

Impactful vectors directed at healthcare include wireless technology compromise (especially Bluetooth and Wi-Fi), compromise of vulnerabilities, phishing attacks, compromise of

remote access technologies, credential compromise, etc. Major contributions of AI to healthcare cybersecurity include the following:

a. Identification of vulnerabilities and threats: AI can examine patterns in network traffic and user activity in the healthcare environment to flag any abnormalities from normal actions. Natural language processing (NLP) scans for social engineering attacks or documents and communications to identify signs of emergent cyber threats.

b. Behavioral modeling is used for identifying and responding to breaches using AI to improve identification and isolation of threats much quicker than traditional security methods when a security breach does occur.

c. Medical devices are protected from attacks: If left exposed to remote breaches smart medical devices pose significant threats to patient safety but AI helps address the most challenging barriers to safeguarding these devices.

d. Increasing accuracy and efficiency: AI analyzes patient outcomes and treatment efficiency based on historical data, which helps healthcare providers enhance their practices and provide better patient care quality.

e. Ensure privacy and compliance: There is a need for AI-driven systems to constantly monitor data access and usage, guaranteeing that only approved employees can access the EMR/EHR/PHI, to meet conditions of regulations like HIPAA. AI can also assist in audits and reports, streamlining the process of developing fulfillment to regulatory authorities. Automating data access tracking helps AI reduce administrative burdens connected with compliance documentation.

f. Enhancing third-party risk management: AI rapidly analyzes questionnaires, audit security assessments, documents, and additional materials from third-party payors to weigh their cyber risk levels in real time.

Healthcare use cases of ML for cybersecurity include predictive analytics visualizing cybersecurity threats in a user interface for malware detection, anomaly detection for cybersecurity, etc.

Data Science in the Cybersecurity of Healthcare

Patient information, clinical records, analysis, and interpretation of medical data are analyzed by data science as the collection and analysis occur by data science and ML. Because the purpose of advanced data science and ML algorithms is leveraging advanced analytics, AI/ML works together to gain significant information from healthcare data. Working together, data science and cybersecurity work together to ensure the safety of patient information yet able to harness the power of data for better healthcare outcomes, summarized by the following:

a. Unauthorized retrieval of patient records or unusual data transfers inside a healthcare system can be an anomaly detection.

b. User and entity behavior analytics (UEBA) is a data science method that focuses on observing and analyzing user and entity performance within the healthcare network. UEBA establishes a baseline user profile that detects deviances from normal performance and triggers alerts. Extremely successful in the identification of insider threats and protects from unauthorized access to patient data.

c. Data science can design role-based access control systems, while cybersecurity confirms that unauthorized persons cannot gain access to sensitive patient data.

d. Encrypting data at rest and in transit is crucial to protect patient information. Data science can be used for implementing encryption techniques whereas cybersecurity can ensure their efficacy.

e. Data science facilitates the design of secure data-sharing protocols, certifying that sensitive patient data is only revealed to authorized personnel. Workers, staff, and third parties who do not have the encryption key cannot access the data. This aids healthcare facilities in tracking and auditing data access, improving security and accountability.

f. With the number of IoT and IoMT in healthcare, data science must control and interpret their data for better patient care, while cybersecurity guarantees the defense of these devices from probable attacks.

g. Threat intelligence is the process by which data science gathers data on the latest cybersecurity threats and vulnerabilities. Data science also aids the assimilation of threat intelligence into healthcare cybersecurity. Furthermore, healthcare organizations can stay updated on emergent threats and tailor a security package to address their specific security needs and specific risks. By staying updated on emerging threats, healthcare organizations can tailor their security measures to address these specific risks.

h. Data Science optimizes telehealth services by analyzing patient data for remote diagnostics, while cybersecurity safeguards the protected transmission of patient data in transit over digital channels.

i. Routine audits are the key to complying with regulations like HIPAA. Data science can aid in compliance monitoring and identification of any areas of noncompliance, meanwhile, cybersecurity can ensure data is protected over the entire process.

j. Data privacy is the primary consideration in healthcare and medical data privacy must be guaranteed while detecting fraudulent data and must not lose its usefulness. Data scientists must construct frameworks and systems to protect sensitive data such as insurance and billing information. The need to uncover fraud is essential. A healthcare organization should be updated regularly to stop any data breaches.

k. Equalizing patient privacy for the greater good of public health is yet a complex ethical dispute. There needs to be a collaboration between data science and cybersecurity to strike the right balance between data science collaboration and cybersecurity to protect patient data while granting data sharing for research purposes.

Automation in the Cybersecurity of Healthcare

Automation can fight these tactics if used correctly by defenders. Healthcare entities can utilize automation to build detections for browsing behaviors and collecting visitor logs. When defenders use automated technologies, healthcare organizations can detect threats more rapidly. Unfortunately, malicious actors may also find automation useful in cybercrime efforts. Threat actors can utilize automation to move through phases faster. In the healthcare environment, data privacy and compliance are key. Automation, therefore, reveals itself as an impressive tool. Enhancing security with automation is as follows:

a. Automation in identity and access management (IAM) processes: IAM harnesses biometric recognition, supported by automation, to admit access. Automation also quickens dynamic role-based access controls, adapting user permissions in real time in response to a change in responsibilities or duties.

b. Continuous monitoring and anomaly detection for data security: In scenarios like unauthorized data extractions from electronic health records (EHRs), immediately trigger alarms, guaranteeing a quick response.

c. Automated threat detection and prevention systems: When malicious payloads are detected on MRI or CT scan consoles, the systems can promptly insulate the devices, preserving wider hospital network reliability.

d. Integration of automation with security information and event management (SIEM) systems: Critical warnings are automatically highlighted, certifying timely mitigation. Automation also accelerates the creation of a visual dashboard within SIEMs for real-time threat landscapes.

e. Automated compliance assessments: Automated tools leverage AI to conduct multidimensional compliance checks. NLP can semantically analyze EHR annotations and access logs, identifying potential unauthorized access or data mishandling.

Conclusion

AI can continuously monitor network traffic, user behaviors, and system anomalies. It can swiftly recognize unusual patterns, revealing cyberattacks. ML can make cybersecurity simpler, more effective, more proactive, and less expensive. Data science has a critical role in analyzing the scope of a security incident response and identification of any compromised data. Data science and cybersecurity can work together and ensure the safety of patient information and good healthcare outcomes. Cybersecurity automation can perform real-time data analytics and provide efficient protection against cyberattacks. In the healthcare sector, automation helps detect threats more rapidly. Cybersecurity automation is a useful tool in healthcare for data privacy and compliance. Unfortunately, cyber criminals also have access to AI/ML and automation. It is the responsibility of the healthcare IT team to maintain strict cybersecurity policies and updated technologies.

Acknowledgements

The authors would like to thank Technology and Healthcare Solutions, Mississippi, USA for its support.

Conflict of interest

Authors declare no conflict of interest.

References

1. Yener B, Gal T (2019) Cybersecurity in the era of data science: Examining new adversarial models. IEEE Security & Privacy 17(6): 46-53.
2. Riesco R, Villagrá V A (2019) Leveraging cyber threat intelligence for a dynamic risk framework: Automation by using a semantic reasoner and a new combination of standards (STIX™, SWRL and OWL). International Journal of Information Security 18(6): 715-739.
3. Tubis AA, Werbińska Wojciechowska S, Góralczyk M, Wróblewski A, Ziętek B (2020) Cyber-attacks risk analysis method for different levels of automation of mining processes in mines based on fuzzy theory use. Sensors 20(24): 7210.
4. Wasserman L, Wasserman Y (2022) Hospital cybersecurity risks and gaps: Review (for the non-cyber professional). Frontiers in Digital Health 4: 862221.
5. Sharma N, Jindal N (2023) Emerging artificial intelligence applications: metaverse, IoT, cybersecurity, healthcare-an overview. Multimedia Tools and Applications, 1-29.
6. Sarker IH (2023) Machine learning for intelligent data analysis and automation in cybersecurity: current and future prospects. Annals of Data Science 10(6): 1473-1498.
7. Sarker IH, Kayes ASM, Badsha S, Alqahtani H, Watters P, et al. (2020) Cybersecurity data science: an overview from machine learning perspective. Journal of Big data 7: 1-29.
8. Yaseen A (2024) Enhancing Cybersecurity through Automated Infrastructure Management: A Comprehensive Study on Optimizing Security Measures. Quarterly Journal of Emerging Technologies and Innovations 9(1): 38-60.