Open Access Journal of Computer Science and Engineering
Volume 2, Issue 1, 2025, PP: 1-9
www.aytinpublications.com

Research Article

Data Use Cases, Data Strategy, and Data Analytics of Cyber Incidents for Cybersecurity

Cheryl Ann Alexander[1*] 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.

Abstract

Predicting cyber incidents based on data analytics for cybersecurity is a data-driven issue. Data used in prediction should be trustworthy, comprehensive, and typical without bias. This paper introduces data acquisition, data strategy, cyber incidents, and data analytics. Frameworks for supporting data sources, processing, cybersecurity, and related practices are presented. Data use cases and related technologies for cybersecurity in healthcare are studied. Health data are diverse and huge. Medical records are key data sources in healthcare. The roles and trends of cybersecurity for healthcare include data protection, secure communication, risk management, etc. Challenges in data-driven cyber incident prediction, customizing analytics models for various cyber incidents, and building resilient cyber systems are highlighted. The methods and related information in this paper help enhance cybersecurity in biomedical science and engineering, especially biomedical data engineering

Keywords: Cybersecurity; cyber incidents; offensive data strategy; data analytics; machine learning (ML); deep learning (DL); big data; healthcare

Aricle Information

Recieved: 29 December 2024

Accepted: 17 January 2025

Published: 20 January 2025

Cite this article as:

Cheryl Ann Alexander, Lidong Wang. Data Use Cases, Data Strategy, and Data Analytics of Cyber Incidents for Cybersecurity. Open Access Journal of Computer Science and Engineering, 2025;2(1); 1-9

https://doi.org/10.71123/oajcse.v2.i1.25001

Copyright: © 2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Introduction

The automated and real-time switching of data in cybersecurity from unstructured data to structured helps data analytics of cyber threats. Named entity recognition (NER) can convert unstructured data into structured data. A model of language representation called bidirectional encoder representations from transformers (BERT) has made substantial improvements in various tasks of natural language processing. BERT and its improved version BERT with whole-world masking were applied to the NER task for cybersecurity, and a better performance on the recall, precision, and F1 score was achieved compared to other models (Zhou et al., 2021). A platform, system, or solution leveraging IoT devices needs to support the following security requirements (Adat and Gupta, 2018):

- Authentication: Edge devices need to be authenticated to the cloud and other edge devices, only permitting authorized nodes to communicate and retrieve data.
- Secure service discovery: Services should be discovered and delivered securely, avoiding fake nodes or fake users.
- Secure data aggregation and sharing: Data aggregation (in intermediate layers) and data sharing (between the edge and cloud) should be encrypted.
- Detection of malicious nodes: A mechanism is required to detect malicious nodes because distributed nodes are vulnerable to both internal and external software or hardware attacks.
- Secure virtualization: A secure virtualization environment is needed to avoid virtualization attacks, malicious virtual machines, etc.

Cyber Threat Information and Intelligence (CTII) helps situational awareness of threats and a better understanding of threat actors. Suppose CTII is going to be shared (particularly with external parties). In that case, some

interoperability and cybersecurity problems must be faced, categorized into four layers shown in Table 1 (Rantos et al., 2020). The relationship between systems security requirements and business impacts in Industry 4.0 contexts (if the loss or degradation of data occurs) is shown in Table 2 (Corallo et al., 2020).

Table 1. *Layers of CTII exchange interoperability*

Layers	Details
Technical	• Data transmission & protection
Semantic & Syntactic	• Unambiguous meaning • Data types and formats
Policy & Procedures	• Organization objectives and instructions • Recipients • Business interest
Legal	• Data privacy • Obligations to share • Restrictions on what to share

Table 2. *Security requirements and associated business impacts in Industry 4.0 contexts*

Security Requirements	Business Impacts
Confidentiality: keeping from unauthorized information disclosure	Theft of industrial trade secrets & intellectual property: • Violation of commercial agreements with industrial partners on data confidentiality • Damage to company's reputation • Reduction in the company's competitive advantages
Integrity: keeping from improper alteration	Sabotage of the whole critical infrastructure or specific machines & components: • Damages of working machines • Degradation of product quality • Violation of regulations & standards in pollution & safety • Violation of agreements with clients on the specifications of products. • Threatening workers' lives.
Availability: Protecting data for being accessible and usable on demand	Denial of service (DoS) of networks, system devices, or other resources: • Quality degradation of workpieces • Violation of agreements with clients on the delivery time • Loss of production time

Major sources of cybersecurity threats were investigated and top cyberattacks in the world were identified, which is shown in Table 3 (Shestak and Tsyplakova, 2023). Ten privacy rights derived from Europe's General Data Protection Regulations for owners of online personal information are as follows (Lee et al., 2020):

- Be informed—know who is handling their data.
- Access—access any personal data collected about them.
- Rectification—ask for correcting inaccurate personal data.
- Be forgotten—have their data deleted (preventing further collection).
- Restriction of processing—ask for restricting the process of specific types of personal data.
- Data portability—ask for transferring personal data to a recipient of their choice.
- Right to object—consent or withdraw consent regarding the processing of personal data.
- Opt-out of an automated system—opt-out of the utilization of their data by an automated system.
- Not to be subjected to unsanctioned privacy invasion—be left alone unless they violate regulations or laws.
- Not to be known by unpermitted persons—decide who is permitted to access the personal data and personal online space.

Table 3. *Top-known cyberattacks in the world*

Types	Features or Description
Spams	Sending e-mails with false or hidden information to sell products, perform attacks, or activate phishing, spy- or malware software.
Phishing	Stealing personal information or data via spam and spy- or malware software.
Insiders	Disgruntled employees of an organization
Botnet	A hacker operates a system, coordinates attacks, and dissimilates malware, spam, or phishing.
Hackers (including hacktivists)	Causing serious damages or failures using malware or other instruments.
Cyber-terrorism and cyber-extremism	Attempt to disable, destroy, or use critical infrastructure to compromise national security; destroy the economy of a country; etc.
Foreign intelligence services	Critical infrastructure decommissions, information warfare, etc.
Authors of spy- and/or malware software	Creating or dissimilating computer viruses, worms, and spy- or/and malware software that destroy hard drives and files.
Organized criminal groups	Monetary gain via phishing, spam, spy- or malware to steal personal data and e-fraud.

IT teams and security administers need to make their systems resilient; however, they have the following challenges (Masip-Bruin et al., 2021): 1) static networked configurations and dynamic system audit, 2) evidence-based metrics for trust guarantees and security assurance, 3) requirement for end-to-end solutions for vulnerabilities and risks management, 4) burdensome coordination in multi-actor or multi-vendor supply chains of information and communication technology (ICT) systems, and 5) implausible wide implementation of integrated cybersecurity solutions for created ICT systems.

The objective of this paper is to deal with data acquisition, data strategy, data use cases, and data analytics for cybersecurity. Data use cases and related technologies for cybersecurity in healthcare are presented as a case study. The subsequent sections of the paper are organized as follows: the second section introduces data acquisition, data strategy, cyber incidents, and data analytics; the third section presents frameworks for supporting data sources, processing, cybersecurity, and related practices; the fourth section deals with data use cases and related technologies for cybersecurity in healthcare; and the fifth section is the conclusion.

Data acquisition, data strategy, cyber incidents, and data analytics

Table 4 shows a categorization of cyber incidents (Sun et al., 2019). Howard and Longstaff provided a data acquisition/collection method and comprehensive analysis of security incidents (Howard, 1997; Howard and Longstaff, 1998), shown in Table 5. Challenges in data-driven cyber incident prediction are as follows (Björck et al., 2015; Sarabi et al., 2015):

- Data collection and processing: Predicting cyber incidents for cybersecurity is a data-driven issue. Data used for prediction should be authentic (reliable and accurate), representative (being typical of incidents with no bias), and comprehensive (including everything needed or relevant), which is a challenge. There are often biased datasets in self-reporting incidents detected externally by a third party. Data quality is critical for prediction performance.
- Feature extraction for data representation and representation learning: The performance of prediction methods such as ML depends on the selection and extraction of features that rely on domain-specific knowledge. Critical underlying features or factors hidden behind the data are sometimes overlooked.
- Security problem modeling: There is much work to do in defining and refining the problems of various incidents and much space to improve modeling performance for cyber resilience.
- Cyber incident analysis: It is often hard to extract relevant information (with enough amounts and high quality) of security incidents and perform inferring.
- Model customization: It is often hard to use existing data mining (DM) and ML models and algorithms directly without customization in handling security incidents. However, it is a challenge to customize models for various cyber incidents because domain-specific knowledge is required.
- Evaluation: It is often difficult to properly evaluate the results of cyber incident prediction due to the lack of knowledge of future incidents.

Table 4. *Cyber incidents*

Categories	Sub-categories
Malicious codes	Malicious software, malicious web, malicious mobile app
Unauthorized access	Account hijacking, data breach, reverse engineering attack
Inappropriate usage	Network misarrangement, intrusion, compromise, vulnerability exploits, underground black keywords
DoS	Distributed denial of service (DDoS), amplification attacks

Table 5. *Howard and Longstaff's taxonomy of cyber (security) incidents*

Aspects	Details
Attackers	Hackers, terrorists, spies, professional criminals, voyeurs, vandals, corporate raiders
Tools/approaches	Distributed tools, toolkits, user commands, information exchange, physical attacks, data taps, autonomous agents, programs/scripts
Vulnerabilities	Configuration, design, implementation
Actions	Scan, probe, authenticate, read, delete, modify, steal, copy, flood, spoof, bypass
Targets	Network, Internetwork, computer, components, processes, data, accounts
Unauthorized results	Theft of resources, information disclosure, increased access, DoS
Objectives	Damages, financial gains, political gains, challenges, status, thrill

Industries with strong regulations such as healthcare and financial services usually implement defensive data strategy; however, offensive data strategy (DalleMule and Davenport, 2017) offers new opportunities because it focuses on improving profitability, revenue, and customer satisfaction. In addition, offensive events are generally more real-time than defensive events. Balancing defensive data and offensive data is balancing data control and flexibility. It is significant to create suitable trade-offs between defense and offense and reach a good balance.

Generally, ML can be used in the analysis of malicious activities/behaviors, the detection of DoS, the analysis of intrusion, etc. Table 6 (Sarker et al., 2020; Sarker, 2023) lists some specific applications of various ML methods in cybersecurity. Imbalanced data destroys the performance of DL in the analytics of cybersecurity. In transfer learning, knowledge is transferred from the source domain to the target domain. There are two primary types of transfer learning: inductive learning and transductive learning. The first one mainly deals with task knowledge transfer; the second one mainly handles data domain transfer. A transfer learning method was proposed to mitigate the imbalanced data problem when using DL in cybersecurity. This is a method based on domain adaptation to train a cyberattack detection model using an extremely imbalanced dataset. How domain adaptation helps achieve better results was also discussed (Wang et al., 2023).

Table 6. *Some applications of ML in cybersecurity*

ML Methods	Example of Applications
k-nearest neighbors (*k*-NN)	Systems of network intrusion detection, reduction of the false alarm rate
Naive Bayes	Building an intrusion detection system for multi-class classification
Support Vector Machine (SVM)	Feature selection, intrusion detection and classification; classifying various attacks (e.g., DoS, probe); evaluating host-based anomaly detection systems.
Decision Tree	Feature selection & building a system for network intrusion detection
Random Forests	Systems of network intrusion detection
Association Rule	Systems of network intrusion detection
General DL	IoT botnet traffic classification, detecting of malicious users and phishing websites
DL: Convolutional	Systems of malware traffic classification
DL: Recurrent, RNN, LSTM	Systems of anomaly intrusion detection and attack classification
Deep and Reinforcement Learning	Systems of malicious activities and intrusion detection

The main properties of behavioral processing methods are compared and shown in Table 7 (Sanchez et al., 2021). Statistical methods are often used in data pre-processing and anomaly detection. Rule-based methods are used for model definition and anomaly detection. Knowledge-based methods are mainly used for behavioral anomaly detection. ML/DL algorithms have been powerful in data analytics. ML/DL-based methods can handle multi-variate and multi-dimensional data. Time series analysis includes various models and algorithms such as ML/DL-based algorithms and statistical algorithms (Sanchez et al., 2021).

Table 7. *The comparison of behavioral processing methods*

Properties	Statistical	Rule-based	Knowledge-based	ML/DL-based	Time series
Simplicity	Yes	Yes	Partial	No	No
Large datasets required	No	No	No	Mainly DL	Yes
Multi-dimensional data	No	No	No	Yes	ML/DL-based
Large training time	No	No	No	Mainly DL	Yes
Fast computation/low resource	Yes	Yes	No	No	No
Expert knowledge required	Yes	Yes	No	No	No
Adaptability	No	Dynamic approaches	No	Yes	ML/DL-based
Decision explainability	No	Yes	Yes	Partial	No
Complex feature correlations	No	No	Partial	Yes	ML/DL-based

Big data cybersecurity includes two aspects: one is big data (analytics) as a security tool or solution; the other is securing big data because is often exposed to cyberattacks. Approaches to securing big data include encryption, access control, context-aware data access, context-aware data storage, and data differentiation. The two aspects are shown in Figure 1 (Rawat et al., 2019).

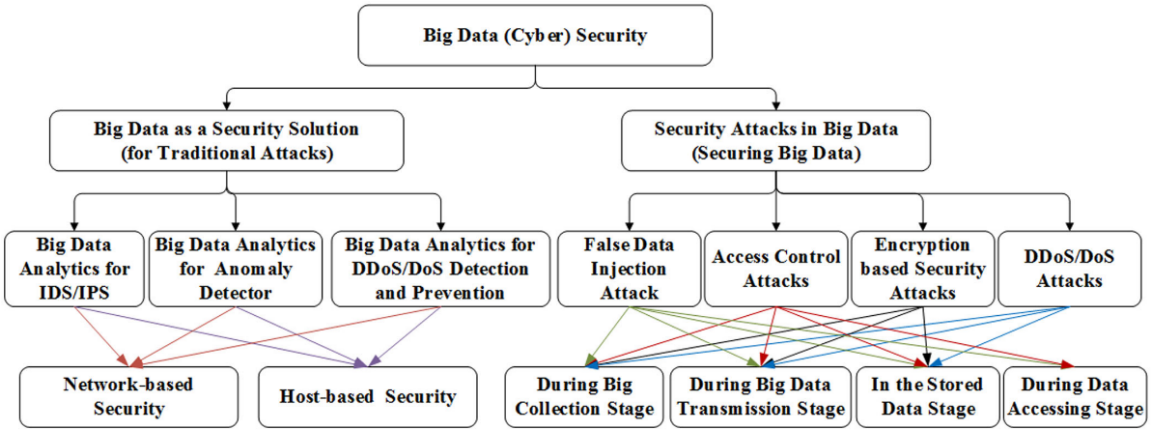

Figure 1. *Big data (analytics) as a security solution and security attacks on big data*

Frameworks for Supporting Data Sources, Processing, Cybersecurity, and Related Practices

A cognitive cybersecurity analysis framework was developed, which streamlined data integration, data processing, and knowledge production. There are the following detailed procedures: 1) the categorization of heterogeneous data streams and fusing into a collaborative information system; 2) the transformation of ingested information into knowledge combined with knowledge formalism, 3) finding the best ensemble model automatedly through the ML pipeline for various tasks of cybersecurity classification. Using more cybersecurity data sources, especially *new data sources* such as technical blogs, further supports data analytics based on ML (Jiang and Atif, 2021).

Table 8 (Sarker et al., 2020) shows a multi-layered framework for smart cybersecurity based on ML. The data flow in the framework is from the capture of raw security event data (L1: layer 1) to L2 (layer 2), then L3 (layer 3), and finally L4 (layer 4).

Table 8. *A framework for smart cybersecurity*

Layers	Functions or Features
L1: Security data collection	Activities (users, network, database, application)
L2: Security data preparation	Data cleansing, normalization, transformation, collation, etc.
L3: ML-based security modeling	• Security feature engineering • Creating similar incident groups/data clustering • Classification or prediction of attacks • Detection of anomalies or malicious behaviors • Association learning & policy rule generation • Model selection or customization
L4: Incremental learning & dynamism	• Recency mining & updating security models • Post-processing & improvement, • Response planning & decision making

A framework for health monitoring/measurement of cyber-physical systems and a framework of the data flow of health measurement were developed. The frameworks are data-driven and shown in Table 9 and Figure 2 (Amarasinghe et al., 2018).

Table 9. *A framework and its components of health monitoring/measurement*

No. of Steps	Steps	Components
1	Data acquisition & feature extraction	• Historical data • Real-time data • Physical data & cyber data • Feature extraction (physical, cyber)
2	State learning & estimation	• Possible state identification & state learning from historical data • Estimating the current state using the learned model • Anomaly detection • Etc.
3	Health evaluation	• Characterizing the "healthy" behaviors (physical, cyber) using the state estimator • Using current state estimation to compare with "healthy" states • Generating a health "score" • Providing a confidence measure on any issue
4	Warnings	• Warning operators about health degradation • Temporal context & spatial context • Confidence of the general warning

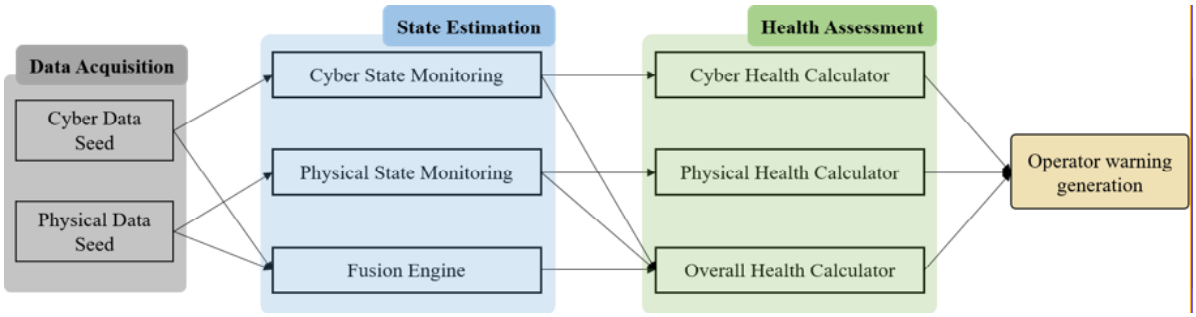

Figure 2. *A framework of the data flow of health measurement*

Data Use Cases and Related Technologies for Cybersecurity in Healthcare

Applications of cybersecurity in healthcare lie in securing healthcare access, protecting health data, improving healthcare outcomes, safe connectivity & securing healthcare network, assisting security teams & coordinating treatment processes, handling complicated treatments with outstanding patient care, safeguarding medical devices & equipment, preventing medical frauds, preventing daily operations from attacks and cyber-criminal threats, risk assessment of healthcare systems, etc. Tools of cybersecurity for healthcare include multi-factor authentication, data loss prevention, anti-theft devices, vulnerability scans, digital forensics, penetration testing, threat intelligence, and disaster recovery plans. The roles and trends of cybersecurity for healthcare are data protection, secure communication, risk management, devices and services, authentication and authorization, device inventory & risk analysis, improved & enhanced services, and security training (Javaid et al., 2023).

Cyber technologies and their risks in healthcare include telemedicine (healthcare providers communicate with patients or other providers via software or unsecured portals), electronic data (risks in exchange and storage), medical devices (e.g., insulin pumps, MRI machines), etc. The risky technologies, vulnerabilities associated with the technologies, current risk management, and recommended strategies are shown in Table 10 (Wasserman and Wasserman, 2022).

Table 10. *Risky technologies, vulnerabilities, management, and strategies in healthcare*

Aspects	Examples or Details
Risky technologies	Telemedicine, electronic data, medical devices
Associated vulnerabilities	Lack of regulations, out-of-date systems, Internet threats, rapid innovation, lack of cyber resources, constant accessibility, interoperability, and focus on medical care over cyber efforts.
Current risk management	Regulatory measures, technical measures, detection and response, device requirements, insurance
Recommended strategies	Group efforts, training, risk management, technical measures, etc.

An expanding attack surface beyond medical IoMT for healthcare systems includes almost all devices such as cameras, HVAC systems, applications, elevator control systems, SaaS, and cloud. Minimizing the attack surface, continuous improvement of data security posture and identity, and enabling zero-trust security will help data access security. Healthcare entities must govern and protect their critical data across an ever-increasing digital ecosystem, including patient data. Staff continuously create, download, extract, copy, and share data which has expanded the potential attack surface as a result and risks such as lasting damage to the healthcare entity's reputation, data loss, and data breach have occurred. Therefore, organizations use data access security to help keep business-critical data guarded. Uncovering any hidden risks and monitoring security as it progresses through integrated reporting and shared dashboards may amplify a healthcare entity's data security posture and identity.

Sources of data in healthcare include insurance claims, EHR, user-generated data, research, and public health data. Proprietary data sources include the American Hospital Directory (AHD) which contains an exclusive database of hospital phone names, phone numbers, addresses, system affiliations, and websites. The AHD is constantly revised as changes occur and new data becomes available. Providers, users, and healthcare planners use health data as data sources to obtain facts that can be gathered and analyzed to yield necessary information such as diagnosis, treatment plans, evaluations, etc.

While health data are diverse and massive, potential sources in practice are medical records, responses in surveys, certificates of vital or other health-related events, and research facts. Unobtrusive data can also be used here. Medical records are key data sources in healthcare and the simplest medical record contains data from all the following categories:

- Personal identification data: name, sex, birth date/age, sex, etc.
- Socio-demographic data: age, sex, occupation, residence place.
- Administrative data: care sites and referrals.
- Clinical data: diagnoses, treatment regimens, investigations, and medical history.
- Behavioral data: observance of recommended regimens (or otherwise).
- Economic data: payment method and insurance coverage.

In today's hospitals and clinics, and public health departments, data from all the categories are found in the medical record, which is a compilation of all treatment and services received. However, not all the information can be found in the same file. Economic and administrative data are typically separated from clinical data. Both file types are usually linked by personal health information and personal identification information. If a patient fails to obtain prescribed medication, does not keep appointments, or other behavioral information, is cross-referenced by linking data in the EHR with pharmaceutical data or calendar data.

Healthcare surveys are used by hospitals, clinics, and other healthcare entities to determine the level of excellence provided by the organization. Multiple variables are typically surveyed on a Likert scale with an occasional comment. However, only a sample of patients return or complete the health survey, although if it is a statistically significant representative sample, findings can be confidently generalized. Health surveys are done by several methods, including verbally through a phone or in person, or the most common method, which is a Likert Scale written questionnaire. Survey data must be checked, organized, and amended for consistency and processed and analyzed by a computer with applications set for analyzing the data. Numerous variables can be collected via this manner of survey, with details such as demographics and experience with the medical center care.

Clinical research datasets offer clinical research data via national and discipline-specific entities. Although most often restricted, management and IT teams can access the data through the appropriate channels. In addition, proprietary research data can be made available through agreements with private organizations. Public health data can also be accessed through government agencies such as the CDC, CMS, and MEPS (Medical Expenditure Panel Survey). These organizations publish data gained through healthcare facilities and other surveys such as population surveys and are critical to obtaining epidemiological data such as mortality and morbidity reports, disease prevalence, prevention, etc.

User-generated data is becoming more important with the rise of the empowered patient and abundant availability of wearables capturing multiple new metrics about the patient such as blood pressure, temperature, pulse, oxygen saturation, medication adherence, etc. This data is critical to the provider who makes the decisions regarding healthcare interventions. Population health data and samples of unobtrusive data allow providers access to multiple sources of information that can also be used to make treatment decisions.

Clinical trials and ongoing patient care provide a rich source of clinical data, useful for most clinical or medical research. Gathered at the beginning of care at a healthcare

facility, the EMR is unavailable to outside researchers. Demographic and administrative data, diagnostic data (e.g., laboratory tests, pharmaceutical data, etc.), hospitalization data, insurance data, and physiological monitoring data (e.g., vital signs, graphic charts, medical equipment rates and orders, etc.) are key data found in patient medical records. The EMR keeps a record of each patient visit through the facility. Documenting clinical information and health facility workflow data (i.e., appointments, etc.), the EMR is the richest source of data in healthcare, therefore, making inter-operability a necessary function for the EMR, which has become easier over recent years. It is also important to note that included in the EMR are administrative data, composed primarily of discharge data and reported to national agencies such as AHQR, CMS, and others.

Claims data describes insurance claims that are between the insured patient and the healthcare provider. Claims data breaks into four categories outpatient, inpatient, enrollment, and pharmacy. The government or commercial health firms (e.g., United Healthcare, Blue Cross Blue Shield, Humana, etc.) are sources of the data. Providers submit claims based on ICD-10 data (i.e., diagnostic codes based on treatment plans. Payment is facilitated based on the diagnostic codes. Claims data is used very often in treatment decisions. Other sources of information include disease registries where agencies track a narrow range of critical data for specific conditions such as cancer, diabetes, Alzheimer's disease, cardiovascular disease, lung diseases, etc. These registries often provide essential information key to treating the patient.

Conclusion

IT teams and security administrators should make their systems robust; however, there are challenges which include networked configurations and dynamic system audits, evidence-based metrics for trust guarantees, security assurance, etc. Predicting cyber incidents for cybersecurity is a data-driven issue, but there are challenges in data-driven cyber incident prediction. Data utilized in prediction needs to be trustworthy, comprehensive, and typical with no bias. Typically, ML can be used in the analysis of malicious activities, detection of DoS, the analysis of intrusion, etc. Enterprises with strong regulations such as healthcare and financial services generally implement a defensive data strategy; however, an offensive data strategy proposes new opportunities because it emphasizes refining profitability, revenue, and customer satisfaction. The roles and trends of cybersecurity for healthcare include data protection, secure communication, risk management, etc. Defending health data, safe connectivity/securing healthcare networks, etc. build robust cybersecurity in healthcare. The methods and related information in this paper help enhance cybersecurity in the areas of biomedical science and engineering, especially biomedical data engineering.

Acknowledgements

The authors would like to express thanks to Technology and Healthcare Solutions, USA for its help and support.

Declaration of the use of AI tools

The authors declare that they did not use AI tools in writing this paper.

Conflict of interest

The authors would like to announce that there is no conflict of interest.

Ethics

In this article, ethical principles related to scientific research articles are observed. The corresponding author confirms that both authors have read, revised, and approved the paper.

References

1. Adat, V., & Gupta, B. B. (2018). Security in Internet of Things: issues, challenges, taxonomy, and architecture. *Telecommunication Systems*, *67*, 423-441.
2. Amarasinghe, K., Wickramasinghe, C., Marino, D., Rieger, C., & Manicl, M. (2018, August). Framework for data driven health monitoring of cyber-physical systems. In *2018 Resilience Week (RWS)* (pp. 25-30). IEEE.
3. Björck, F., Henkel, M. Stirna, J. and Zdravkovic, J. (2015). Cyber resilience—Fundamentals for a definition, in *New Contributions in Information Systems and Technologies*. Cham, Switzerland, Springer, 311–316.
4. Corallo, A., Lazoi, M., & Lezzi, M. (2020). Cybersecurity in the context of industry 4.0: A structured classification of critical assets and business impacts. *Computers in industry*, *114*, 103165.
5. DalleMule, L., & Davenport, T. H. (2017). What's your data strategy. *Harvard business review*, *95*(3), 112-121.
6. Howard, J. D. (1997). An analysis of security incidents on the internet 1989-1995. Ph.D. dissertation, Carnegie Mellon University.
7. Howard, J. D., & Longstaff, T. A. (1998). *A common language for computer security incidents* (No. SAND98-8667). Sandia National Lab.(SNL-NM), Albuquerque, NM (United States); Sandia National Lab.(SNL-CA), Livermore, CA (United States).

8. Javaid, M., Haleem, A., Singh, R. P., & Suman, R. (2023). Towards insighting cybersecurity for healthcare domains: A comprehensive review of recent practices and trends. *Cyber Security and Applications*, 100016.

9. Jiang, Y., & Atif, Y. (2021). A selective ensemble model for cognitive cybersecurity analysis. *Journal of Network and Computer Applications*, *193*, 103210.

10. Lee, J. K., Chang, Y., Kwon, H. Y., & Kim, B. (2020). Reconciliation of privacy with preventive cybersecurity: The bright internet approach. *Information Systems Frontiers*, *22*, 45-57.

11. Masip-Bruin, X., Marín-Tordera, E., Ruiz, J., Jukan, A., Trakadas, P., Cernivec, A., Lioy, A., López, D., Santos, H., Gonos, A., Silva, A., Soriano, J., & Kalogiannis, G. (2021). Cybersecurity in ICT supply chains: Key challenges and a relevant architecture. *Sensors* (Basel, Switzerland), *21*(18), 6057.

12. Rantos, K., Spyros, A., Papanikolaou, A., Kritsas, A., Ilioudis, C., & Katos, V. (2020). Interoperability challenges in the cybersecurity information sharing ecosystem. *Computers*, *9*(1), 18.

13. Rawat, D. B., Doku, R., & Garuba, M. (2019). Cybersecurity in big data era: From securing big data to data-driven security. *IEEE Transactions on Services Computing*, *14*(6), 2055-2072.

14. Sanchez, P. M. S., Valero, J. M. J., Celdran, A. H., Bovet, G., Perez, M. G., & Perez, G. M. (2021). A Survey on device behavior fingerprinting: Data sources, techniques, application scenarios, and datasets. *IEEE Communications Surveys & Tutorials*, *23*(2), 1048–1077.

15. Sarabi, A., Naghizadeh, P., Liu, Y., & Liu, M. (2015, June). Prioritizing Security Spending: A Quantitative Analysis of Risk Distributions for Different Business Profiles. In Proc. Workshop Econ. Inf. Security (WEIS), 1–12.

16. Sarker, I. H. (2023). Multi-aspects AI-based modeling and adversarial learning for cybersecurity intelligence and robustness: A comprehensive overview. *Security and Privacy*, *6*(5), e295.

17. Sarker, I. H., Kayes, A. S. M., Badsha, S., Alqahtani, H., Watters, P., & Ng, A. (2020). Cybersecurity data science: an overview from machine learning perspective. *Journal of Big data*, *7*, 1-29.

18. Shestak, V. A., & Tsyplakova, A. D. (2023). Criminological Features of the Cybersecurity Threats. *Law, State and Telecommunications Review*, *15*(2), 187-203.

19. Sun, N., Zhang, J., Rimba, P., Gao, S., Zhang, L. Y., & Xiang, Y. (2019). Data-Driven cybersecurity incident prediction: A survey. *IEEE Communications Surveys & Tutorials, 21*(2), 1744–1772.

20. Wang, H., Singhal, A., & Liu, P. (2023). Tackling imbalanced data in cybersecurity with transfer learning: A case with ROP payload detection. *Cybersecurity*, *6*(2), 1-15.

21. Wasserman, L., & Wasserman, Y. (2022). Hospital cybersecurity risks and gaps: Review (for the non-cyber professional). *Frontiers in Digital Health*, *4*, 862221.

22. Zhou, S., Liu, J., Zhong, X., & Zhao, W. (2021). Named entity recognition using BERT with whole world masking in cybersecurity domain. 2021 IEEE 6th International Conference on Big Data Analytics (ICBDA), *Big Data Analytics (ICBDA), 2021 IEEE 6th International Conference On,* 316–320.