Journal of Information Technology and Integrity

Creating a Data Architecture for Cybersecurity in Healthcare Systems

Cheryl Ann Alexander[1]*, Lidong Wang[2], and Vijay Shah[3]

[1] *Institute for IT Innovation and Smart Health, Mississippi, United States.*
[2] *Institute for Systems Engineering Research, Mississippi State University, Mississippi, United States.*
[3] *Division of Business, Accounting & Public Services, West Virginia University at Parkersburg,United States.*

Article Details

Article Type: Research Article
Received date: 19th May, 2025
Accepted date: 28th June, 2025
Published date: 07th July, 2025

***Corresponding Author:** Cheryl Ann Alexander, Institute for IT Innovation and Smart Health, Mississippi, United States.

Citation: Alexander, C. A., Wang, L., & Shah, V., (2025). Creating a Data Architecture for Cybersecurity in Healthcare Systems. *J Inform Techn Int, 3*(1): 107. doi: https://doi.org/10.33790/jiti1100107

Abstract

The cybersecurity architecture forms the structure of a typical cybersecurity system, including its main functions, class, the boundaries of each subsystem, their relationship/independence, etc. There are several tactics to enhancing cyber resilience including empowering security leaders and investing in advanced technologies, testing incident response processes regularly, proactively hunting for threats and sharing intelligence, and hardening and protecting core assets. Artificial intelligence (AI)/machine learning (ML) is used to promote strong cybersecurity programs. Encryption provides a robust safeguard against unauthorized access to sensitive data (at rest or in motion). Interoperability has become a critical component of cybersecurity due to the increase in telehealth modules and remote patient monitoring. Blockchain is also picking up more use among healthcare organizations. This paper introduces data architecture for cybersecurity in general organizations, presents data and cybersecurity in healthcare, and deals with data architecture for cybersecurity in a large medical center as a case study. The case study mainly deals with data capture/collection, data processing (e.g., cleaning, sorting, merging, and data storage), data analytics, data governance and management, and applications and services.

Keywords: Cybersecurity, Big data, Data Transmission, Data Storage, Data Architecture, Information, Cloud, Machine Learning (ML), Blockchain, Encryption, Healthcare, Digital Health

Introduction

The cybersecurity architecture shapes the structure of a system, including main functions, class, boundaries of each subsystem, their relationship/independence, etc. Modeling the transformation of information links in the architecture of a cybersecurity system with discretionary access control was studied. The representation of the system architecture in a security graph and the formal take-grant model for synthesizing the cybersecurity architecture were utilized. Modeling algorithms based on the Ford-Fulkerson theorem and methods for finding the smallest edge section of the graph were developed [1].

Recommender System (RS) is a data filtering tool that can provide users with suggestions after gathering and analyzing users' historical data. An RS helps forecast the best option for a given user based on the user's preferences. RSs help mitigate information overload, provide a personalized experience, and more benefits [2]. RSs help access more accessible content, speed up searches, achieve competitive advantages, etc. However, they can also be affected by the lack of scalability, privacy, sparsity, shilling attacks, gray sheep intrusion, etc. [3,4]. A hopeful direction is to use RSs as navigation assistance tools and attack predictors [5].

Hospitals use cloud computing because it presents scalability so that hospitals can adjust data storage and capacity for processing based on the facility's demands. Because healthcare providers can securely access data anywhere that has an internet connection, cloud-based data architecture is flexible and preferred. However, hospitals must employ vigorous security measures to address security concerns in cloud computing. Encryption, therefore, provides a robust safeguard against sensitive data falling into the hands of malicious actors as it converts data into indecipherable code. End-to-end encryptions prevent eavesdroppers from understanding content and together with authentication fortify data privacy. The code utilized during the encryption should be known to read the concealed data. Encryption is used to protect data at rest and data in motion (e.g., AES-256 for data at rest and TLS for data in motion).

The objective of this paper is to create a data architecture for cybersecurity and deal with the specific applications of a cybersecurity-oriented data architecture in healthcare. The subsequent sections of the paper are organized as follows: the second section introduces a data architecture for cybersecurity in general organizations, the third section presents data and cybersecurity in healthcare, the fourth section deals with a data architecture for cybersecurity in a large medical center, and the fifth section is the conclusion.

A Data Architecture for Cybersecurity in General Organizations

The data flow and analytics architecture of big data for cybersecurity was proposed, shown in Table 1 [6]. There are five layers in the architecture for cybersecurity and there are sub-modules in each layer. For example, the extraction layer contains sub-modules (i.e., batch and streaming). Data is extracted from various data sources. The data sources can be security portals, feeds, or blogs; servers and network appliance logs; intrusion detection systems; cyber simulations platforms; sensors; Netflow; etc. [6].

Layers	Description of Each Layer	Sub-modules & Their Functions of Each Layer
Extraction layer	Extracting data from various sources	• Batch: obtaining data from legacy batches • Streaming: obtaining data from various data streams
Load layer	Loading data into a data lake for further transformation & analysis	• Storage: either on a remote or local platform • Indexing: reducing the time in seeing results
Transformation layer	Cleaning & preparing the stored data	• Cleaning: cleaning raw data & standardizing the data content • Preparation: preparing clean data (e.g., dataset unification, grouping, extrapolation)
Analytics layer	Analyzing clean & organized data	• Machine learning (ML): finding patterns & predicting possible behaviors from the data. • Visualization: helping users better understand results using graphs
Execution layer	Offering various applications & services	• Indicators: allowing the visualization of key performance indicators • Alerts: sending alert messages after identifying anomalous or unusual events

Table 1. The architecture of Big Data analytics for cybersecurity

Due to huge data storage in the cloud, ensuring data security in the cloud is important. A Key Management System is a data protection approach to being usually employed in e-health systems, information security, large-scale enterprises, sensor security, architecture security, access control, etc. Because the approach permits the secret key information to be safely exchanged, the security level can be ensured highly. Making reliable and precise authentication is practicable due to secure data transmission. Attribute-based encryption provides dependable and safe access control [7]. A cloud platform with four modules was developed to offer cloud services for government affairs, healthcare, smart campus, etc. The details of the cloud platform and cloud security are shown in Table 2 [8].

Modules	Items
Elastic computing	• Elastic scalability • Cloud hosting • GPU cloud hosting • Dedicated host • Container services • Enterprise cloud desktop • Standard cloud desktop • Container image services • Bare metal server • Cloud backup for virtual machines
Cloud storage	• File storage • Object storage • Cloud storage gateway • Cloud disk • Cloud disk backup
Cloud networking	• Virtual private cloud • Cloud interconnection • Cloud-dedicated connection • Shared bandwidth • Elastic public IP (Internet Protocol) • IPv6 transition • SSL VPN (virtual private network) • IPSec VPN • Elastic load balancing
Cloud Security	• Cloud firewall • Web application firewall • Cloud hosting security • Vulnerability scanning • Enhanced vulnerability scanning • Database auditing • Log auditing • Cloud bastion host • DDoS (distributed denial-of-service) protection

Table 2. A cloud platform and modules

Resilient cybersecurity is the degree of normal operation that an enterprise can keep during a cyberattack, security incident, data breach, or system fault. Approaches to enhancing cyber resilience include: 1) empowering security leaders, and investing in advanced technologies, 2) testing incident response processes regularly, 3) proactively hunting for threats and sharing intelligence, and 4) hardening and protecting core assets. A cybersecurity architecture design was performed based on situation awareness and decision-making employing a knowledge base of previously spotted unbalanced operations and (potential) breakpoints. The method utilizes specific criteria (e.g., date, time, incidents, situations, and indications for system behaviors) [9].

Mathieu Cousin, Radhika Nallayam, and Wasim Alhamdani introduced how to construct a resilient architecture or security system, respectively. Table 3 [9, 10] shows the comparison of their steps in constructing a resilient security system. A knowledge base of situation awareness and decision-making is built by adding previous unbalanced and breaking point operations (date, time, method, duration of unbalanced operations, recovery procedures, the duration for the recovery procedure, and other information required for making a decision) [9].

Authors	Steps
Mathieu Cousin	• Listing critical assets and making a cyber resilience plan. • Evaluating existing capabilities. • Determining future needs and using new controls. • Adapting and enhancing the resilience strategy
Radhika Nallayam	• Creating a cycle of fast identification and protection. • Performing the automation of detection technologies. • Integrating protective controls. • Performing controls for an automatic response • Performing controls for automatic recovery.
Wasim Alhamdani	• Constructing a knowledge base of situation awareness and decision-making. • Recognizing regular balanced operations. • Detecting unbalanced operations. • Detecting possible breaking points. • Determining recovery procedures. • Determining resilient procedures. • Determining continuity procedures.

Table 3. Comparison of steps to constructing a resilient security system

Data and Cybersecurity in Healthcare

Implementing a complex security system that accounts for present and future needs is one of the greatest challenges facing healthcare administrators and IT teams. Cloud storage is becoming more common among healthcare facilities to keep data secure; however, the cloud has its security risks. Cyber criminals are finding new ways to access sensitive information, and security teams continue to develop more efficient protocols.

There are numerous types of sensitive patient data, including insurance-related data, medical information, and personal data (address, social security number, etc.). Private individual information includes the address, social security number, and phone number, and other personally identifiable data that could be utilized for identity theft and other illicit activities. Healthcare data architects allow for the seamless building of layers of security for patient data in hospitals and other healthcare organizations. Architects must design spaces with flexibility to keep up with the latest data security technological advancements.

Every medical center handles several types of confidential information, such as patient data, billing data, research data, etc. Each type of data carries its own information system. The types of health information systems that can be found in a medical center include the Electronic Medical Record (EMR)/Electronic Health Record (EHR), which deals with patient treatment information; the Master Patient Index (MPI), which contains treatment records from all facilities; remote portals, which deal with distant hardware; Remote Patient Monitoring (RPM), which deal with remote monitoring of networked medical equipment and telehealth; Clinical Decision Support Systems, which provide physician support; and Laboratory Information Systems, which deal with diagnostic data from the laboratory. All these health information systems must be protected and encrypted to prevent malicious actors from accessing the data and inserting malware.

More patient equipment is connected to the Internet, including insulin pumps, pacemakers, defibrillators, etc. Through data encryption, vigorous security can protect what is transmitted between the devices and others, ensuring settings, patient data, and other information are protected from access by malicious actors. RPM uses technology to remotely monitor patient health. RPM facilitates the easy tracking of a patient's vital signs, symptoms, or other diagnostic data by providers outside of the hospital or clinic. Wearable devices, sensors, and other mobile applications are used to collect and transmit data to providers. Various patient data, such as blood pressure, heart rate, blood glucose levels, and medication adherence, are just a few of the data transmitted. Ensuring patient privacy, encryption can protect sensitive data transmitted between patients and providers.

Every medical center has a security architecture with critical pieces of networking such as access controls, disaster recovery plans, encryption, data backup, intrusion detection and prevention systems, and security policies and procedures. These key components can be used to describe the fundamentals of a healthcare system exemplified by its standards and principles that govern the design and evaluation. These security measures align with requirements set forth by HIPAA and NIST CSF to ensure confidentiality, integrity, and availability of protected health information (PHI). Since the increase in telehealth modules and remote patient monitoring, interoperability has become essential because patients and providers must interact from remote locations in telehealth. Health Information Exchanges allow for the easy facilitation of various patient data modules within the healthcare industry. For example, the Sequoia Project, in the United States, runs a nationwide health information exchange called Carequality. This program allows for the exchange of patient data among various EHR systems and connects providers, payers, and other stakeholders.

A VPN connection provides an extra level of protection for confidential data transmitted via the network when alternatives are not practical. The IT team should carefully consider all security and networking issues. Whole disk encryption/full disk encryption can be used to protect against theft. If the computer is stolen, the data will not be accessible. A blank USB drive is necessary to encrypt data on all files going forward. After encrypting the data, it is okay to move files to the newly encrypted USB drive.

Medical cyber-physical systems (MCPS) have the potential to reduce human errors and optimize healthcare services through offering new approaches to monitoring, diagnosing, and treating patients based on the integrated clinical environments (ICEs). Although there are benefits from the MCPS, many ICE's medical devices are not designed to satisfy the requirements of cybersecurity. Therefore, they are vulnerable to cyberattacks. An automatic, intelligent, and real-time system that can detect, classify, and mitigate ransomware attacks in hospital rooms was presented based on ML. In addition, the paradigms of Software Defined Networking (SDN) and Network Function Virtualization (NFV) were also considered to mitigate the ransomware spreading through isolating and replacing infected devices [11].

ML algorithms can be used to support healthcare practitioners in their daily work, such as model-based precision dosing and clinical monitoring. Despite the popularity of AI/ML/DL methods [12] and many successful applications in pharmacy, healthcare, and medical devices, many researchers have yet to uncover more capabilities of such advanced technologies in drug research and development [13]. A multi-step convolutional neural network (CNN) with a stacked long short-term memory architecture was proposed for the attack detection of Internet of Things (IoT) in healthcare applications. Light Spectrum Optimizer was used, and the hyperparameters of the CNN were optimized. The combination of the designed activation function and the best features of the Rectified Linear Unit (ReLU) leads to a lower computation complexity [14]. Cyber risks related to the growth of IoT devices in healthcare were studied by developing a custom cybersecurity IoT environment for healthcare based on reinforcement learning [15].

Digital transformation of healthcare systems should depend on decentralized computer networks and take advantage of the features of blockchain. Decentralization guarantees data transparency and process transparency for all related stakeholders. These are important in populations' healthcare information communications and processing, clinical research management, the tracking and control of medical logistics supply chains, and the control of certified healthcare services organizations [16].

The two main types of healthcare data exchange among entities are business-to-customer (B2C) and business-to-business (B2B). The first type is often performed by offering web-based interfaces for patients, while the second type utilizes the electronic data interchange (EDI) technology between healthcare institutions. The vulnerabilities include poor access controls, weak cryptography, supply chain attacks, repudiable transactions, man-in-the-middle (MiTM), and single points of failure (SPOF). A model was proposed for healthcare data sharing based on the best security practices. The model counters vulnerabilities and automates healthcare processes in a decentralized architecture. It is based on blockchain and zero-knowledge proofs. The Zero-trust concept can be executed using blockchain for the patient data received from sensors or IoT devices, and can be monitored by patients and medical institutions. Various types of blockchain networks are compared in Table 4 [17].

Features	Private Blockchain	Consortium Blockchain	Public Blockchain
Speed	High	High	Low
Agreement	One organization	A chosen set of organizations	All nodes
Central	Yes	Partial	No
Joining terms	Authorized	Authorized	Allow All
Viewing transactions	Public or restricted	Public or restricted	Public

Table 4. Comparison of various types of blockchain

A Data Architecture for Cybersecurity in a Large Medical Center

Charleston Regional Medical Center in the US practices robust cybersecurity for 4,683 employees and all patient visits (yearly over 960,000). There are five functional layers in the cybersecurity data architecture of the Medical Center, including the data capture/collection layer, the data processing layer, the data analytics layer, the applications and services layer, and the data governance layer. Figure 1 shows the five layers and the data flow.

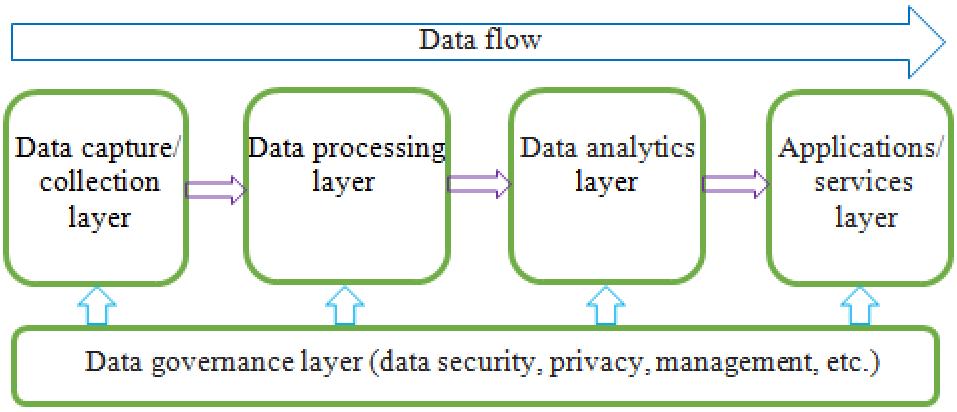
Data flow
Data capture/ collection layer
Data processing layer
Data analytics layer
Applications/ services layer
Data governance layer (data security, privacy, management, etc.)

Figure 1. A Data Architecture for Cybersecurity in the Medical Center

In the data capture/collection layer, the captured or collected data includes structured data (e.g., patient records, mobile apps); semi-structured data or machine-generated data, including clinical results, etc.; and unstructured data (i.e., video, voice, or images) such as R&D labs, home care sensors, etc. In this layer, the acquisition of important data from programs, processes, and devices, including IoMT (e.g., some medical devices, and nursing stations), and scanners (barcoding, fingerprint, RFID) is occurring. However, several factors must be considered such as:

- Cyber simulations platforms;
- Sensors;
- Intrusion detection systems;
- Vulnerability analysis;
- Security portals, blogs, or feeds;
- Net flow;
- Servers and networking appliances logs

The data capture/collection layer also facilitates the connection to the source databases and assists in the extraction of data from these sources. The collected data stream is in numerous formats and at huge volumes, making it almost impossible to impose one data structure. A streaming submodule supervises gathering the data from multiple data streams, utilizing a single data packet each time in sequence. Each packet contains the data source and time reference. The batch obtains data from legacy batches, usually collected over a long period, making batch data extraction a more efficient method for extracting huge volumes of data.

In the data processing layer (e.g., data aggregation and transformation), cleaning, sorting, merging, and data storage occur. When data is collected from heterogeneous sources, various structures and formats require pre-processing to organize the data. Data storage occurs at either a local or remote platform, and this process must be optimal. NoSQL databases are utilized and increase the sensitivity and flexibility of formats. Data indexing reduces the time necessary to see the results. This is especially true for data in a large table. In charge of cleaning and preparing stored data, many variables, data size, and method of accessing the data must be considered. Imperfect tuples may affect the following architectural layers. The data is also prepared in the required formats that will be input for the analysis layer. Data cleaning takes raw data and standardizes it, removing duplicate data, inconsistent heat (or temperature) readings, incomplete values, or meaningless fields, and additional fields that have not yet been considered. Preparation of the clean data occurs in data preparation. This means that data aspects such as reduction, extrapolation, and dataset unification happen in this layer. Although data may be structured or unstructured, it must have a logical and standardized structure.

In the data analytics layer, Big Data analytics, ML, etc., are used to analyze clean, organized data. Various ML and data exploration techniques occur in this layer. Finding anomalous patterns and behavior is one goal for this layer. In the ML module, various ML algorithms are applied to the data with the goal of finding and predicting possible data behaviors. Anomalous behavior fosters the next layer to automate early alerts.

Much happens in the applications and services layer, such as applications, dashboards, and visualization, alerts, etc. In healthcare, visualization, reporting, and real-time monitoring of healthcare dashboards and clinical support happen during this layer. Software (e.g., access control, intrusion/attack detection, diagnostic, medical image processing, patient monitoring, etc.) is necessary. Visualization occurs so that users can better understand data findings. Alerts identify unusual data or anomalies detected in the data analysis, and alerts are sent to the IT team. Indicators foster the visualization of critical performance indicators necessary to obtain a near-real-time status of the data.

Master data management (data immediacy, data completeness, data accuracy, data availability) occurs in the data governance layer. During the data life cycle, data is archived, tested, and delivered in performance applications. During the data life cycle, data deletion and disposal also occur. In data security and privacy management, sensitive data is discovered, vulnerabilities and configurations are assessed, and a security policy is established. There is also change auditing, auditing and compliance, identity and access management, and protection of data in transit. Because patient data must be protected by HIPAA, medical databases for research, e-health, telehealth, EMRs, EHRs, or PHI must have secure data transmission mechanisms using encryption methods. And there is also transport of appropriate events, reporting, storage, and interrogation for investigations such as gunshot wounds, rapes, etc. (i.e., incidents, threat hunting, and forensics).

Conclusion

There are several approaches to enhancing cyber resilience, including empowering security leaders and investing in advanced technologies, testing incident response processes regularly, proactively hunting for threats and sharing intelligence, and hardening and protecting core assets. The Key Management System is a data protection approach that is commonly applied in e-healthcare systems, information security, etc. A cybersecurity architecture forms the structure of a system. Hospitals use cloud computing because of its scalability. This allows hospitals to adjust data storage and capacity for processing.

A security architecture is necessary for every medical center. Key pieces of networking, such as access controls, disaster recovery plans, encryption, data backup, intrusion detection and prevention systems, and security policies and procedures, are descriptive of a fundamental healthcare system that is exemplified by standards and principles that guide the design and evaluation of the system. Interoperability has become a critical component of cybersecurity due to the increase in telehealth modules and remote patient monitoring. Patients and providers must interact from remote locations in telehealth, and a Health Information Exchange facilitates the smooth interface between various data modules within the healthcare industry. Because healthcare providers can access patient data from any place with an Internet connection, a cloud-based data architecture is preferred due to its flexibility. Encryption provides a robust safeguard against unauthorized access to sensitive data (at rest or in motion). Encryption converts data into an indecipherable code, which fortifies data privacy while end-to-end encryptions prevent eavesdropping theft of confidential data. The novel contribution of this paper lies in the construction of an applicable data architecture for cybersecurity in a medical center. Fast Healthcare Interoperability Resources (FHIR) is a standard for exchanging healthcare information electronically. Real-time adaptive threat models and the integration of creating threat models with FHIR can be future research.

Acknowledgements

The authors would like to express their 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. Shumov, A. (2020, September). Modeling transformations of information links in the cybersecurity architecture of systems using algorithms on graphs and the" Take-Grant" Formal Model. *In 2020 International Russian Automation Conference (RusAutoCon)* (pp. 877-881). IEEE.
2. Dilmegani C (2021) Recommendation systems: applications, examples & benefits. AI Multiple. https://research.aimultiple.com/recommendation-system/.
3. Fayyaz, Z., Ebrahimian, M., Nawara, D., Ibrahim, A., & Kashef, R. (2020). Recommendation systems: Algorithms, challenges, metrics, and business opportunities. *Applied Siences, 10*(21), 7748.
4. Mohamed, M. H., Khafagy, M. H., & Ibrahim, M. H. (2019, February). Recommender systems challenges and solutions survey. In *2019 international conference on innovative trends in computer engineering (ITCE)* (pp. 149-155). IEEE.
5. Ferreira, L., Silva, D. C., & Itzazelaia, M. U. (2023). Recommender systems in cybersecurity. *Knowledge and Information Systems*, 1-37.
6. Andrade, R. O., Tello-Oquendo, L., Cadena-Vela, S., Jimbo-Santana, P., Zaldumbide, J., & Yacchirema, D. (2021). Big Data Analytics architecture for cybersecurity Applications. *Latin-American Journal of Computing, 8*(1), 22-37.
7. Ahmad, S., Mehfuz, S., & Beg, J. (2023). Hybrid cryptographic approach to enhance the mode of key management system in the cloud environment. *The Journal of Supercomputing, 79*(7), 7377-7413.
8. Wang, Q., Wang, Z., & Wang, W. (2023). Research on secure cloud networking plan based on industry-specific cloud platform. *IEEE Access*.
9. AlHamdani, W. A. (2020, March). Resilient cybersecurity architecture. In *15th International Conference on Cyber Warfare and Security*. Academic Conferences and Publishing Limited (pp. 23-33).
10. Cousin, M. (2014). Security Think Tank: How to build a resilient defence against cyber attacks. *Retrieved from computerweekly. com: https://www. computerweekly. com/opinion/Security-Think-TankResilience-in-the-face-of-growing-inevitability-of-cyber-attack*
11. Fernandez Maimo, L., Huertas Celdran, A., Perales Gomez, A. L., Garcia Clemente, F. J., Weimer, J., & Lee, I. (2019). Intelligent and dynamic ransomware spread detection and mitigation in integrated clinical environments. *Sensors, 19*(5), 1114.
12. Fleming, N. (2018). Computer-calculated compounds: researchers are deploying artificial intelligence to discover drugs. *Nature, 557*(7707), S55-7.
13. Angehrn, Z., Haldna, L., Zandvliet, A. S., Gil Berglund, E., Zeeuw, J., Amzal, B., ... & Heckman, N. M. (2020). Artificial intelligence and machine learning applied at the point of care. *Frontiers in Pharmacology, 11*, 759.
14. Thulasi, T., & Sivamohan, K. (2023). LSO-CSL: Light spectrum optimizer-based convolutional stacked long short-term memory for attack detection in IoT-based healthcare applications. *Expert Systems with Applications, 232*, 120772.
15. Nadhir, A. M., Mounir, B., Abdelkader, L., & Hammoudeh, M. (2025). Enhancing Cybersecurity in Healthcare IoT Systems Using Reinforcement Learning. T*ransportation Research Procedia, 84*, 113-120.
16. Segal, G., Martsiano, Y., Markinzon, A., Mayer, A., Halperin, A., & Zimlichman, E. (2023). A blockchain-based computerized network infrastructure for the transparent, immutable calculation and dissemination of quantitative, measurable parameters of academic and medical research publications. *Digital Health, 9*, 20552076231194851.
17. Moosa, H., Ali, M., Alaswad, H., Elmedany, W., & Balakrishna, C. (2023). A combined Blockchain and zero-knowledge model for healthcare B2B and B2C data sharing. *Arab Journal of Basic and Applied Sciences, 30*(1), 179-196.