ISSN: 2577-2996

PROGRESS IN MEDICAL SCIENCES, 2022
VOL 6, NO. 1, PAGE 1 – 6
DOI: doi.org/10.47363/PMS/2022(6)E162

REVIEW ARTICLE

Open Access

Data Security in Healthcare with Machine Learning and Biometric Methods: Current Challenges, Solutions, and Future Directions

Kiran Veernapu

USA

ABSTRACT

The healthcare industry is increasingly adopting digital technologies for patient care, management, and research. However, this shift toward electronic health records (EHR), telemedicine, and other digital healthcare solutions presents significant data security challenges. This paper explores the importance of data security in healthcare, identifies key challenges faced by healthcare organizations, and reviews existing solutions to mitigate these risks. Additionally, it discusses emerging trends and future directions for data security in healthcare, including the role of artificial intelligence (AI), blockchain, and regulatory frameworks. The paper aims to provide healthcare organizations, policymakers, and researchers with a comprehensive understanding of the current landscape of healthcare data security and the path forward.

ARTICLE HISTORY

Received January 03, 2022
Accepted January 10, 2022
Published January 31, 2022

KEYWORDS

Data Security, Health Care, Big Data, Data Privacy, Cyber Security, Interoperability, AI and ML for Data Security, Medical Data, Data Warehouse, Machine Learning, Biometrics, Encryption, Quantum Computing, Blockchain

Introduction

In the digital era, healthcare organizations have embraced electronic systems to enhance patient care, reduce administrative costs, and improve overall efficiency. The use of Electronic Medical Records (EMR), telemedicine, wearable devices, and mobile health applications has revolutionized the healthcare industry. However, this increased reliance on digital data systems has made healthcare data a prime target for cyberattacks, putting sensitive patient information at risk.

Healthcare data security encompasses measures, tools, and policies designed to protect health data from unauthorized access, breaches, and misuse. These data security concerns are not only limited to protecting privacy but also involve safeguarding the integrity and availability of healthcare data [1]. The widespread utilization of EMR systems and the need to share data and create data lakes and repositories for analytical reasons reinforced a need for a new generation of data security techniques and innovative solutions. This paper aims to identify the unique challenges associated with healthcare data security, discuss existing data security solutions and strategies, explore future trends in securing healthcare data, and provide actionable recommendations for healthcare organizations and policymakers.

Challenges in Healthcare Data Security

Volume and Complexity of Data

Healthcare data is highly sensitive and diverse, ranging from patient medical records and test results to insurance details and personal identifiers. The sheer volume and complexity of healthcare data make it challenging to secure. Unlike traditional business data, healthcare data requires a high level of granularity, with different access controls and security measures for various types of data (e.g., personal information, medical history, and treatment plans).

A single patient generates 80MB of data each year, in healthcare according to estimates. Culbertson, by 2025 the compound annual growth rate of healthcare data grows by 36%, this growth is faster than many other industries compared with healthcare. Culbertson, also believes that with the enormous growth of the healthcare data, the healthcare industry is not equipped to protect the data [2].

Interoperability and Data Sharing

Interoperability between healthcare systems is essential for providing quality care and enabling seamless communication across hospitals, clinics, insurance companies, and pharmacies [3]. However, the exchange of data across multiple platforms and systems presents a significant security challenge. Data breaches can occur at the points of data exchange, particularly when integrating third-party services or when legacy systems are used in conjunction with modern technologies.

Interoperability is identified as a major goal for IT systems to exchange healthcare data between the providers, and it requires standardized coded data sets that can be easily exchanged and interpreted by all the involved parties. Achieving interoperability requires standardized structured data sets that can be exchanged and interpreted between the healthcare systems [4].

Ransomware and Cyberattacks

Healthcare systems have become frequent targets for cybercriminals, with ransomware being one of the most common

Contact: Kiran Veernapu, USA.

forms of attack [5]. Ransomware encrypts healthcare data, making it inaccessible until the organization pays a ransom to the attacker. The introduction of advanced technologies into the healthcare sector introduced cyber-attacks and put patient health data that protected health information (PHI) at risk [6]. Several known cyber-attacks have been developed to try to hack the data of patients in healthcare.

According to K. S. Bhosale, M. Nenova, and G. Iliev, a recent cyber-attack was posed on public health by a vulnerability called wannacry software virus known as Ransomware, in more than 150 countries. Britain's National Health Service (NHS) many hospitals were disrupted for 4 days [6]. In 2015, the most highly publicized cyber-attack in healthcare was on Anthem, a US-based healthcare from which around 80 million records were stolen.

Lack of Adequate Security Infrastructure

Many healthcare organizations, especially smaller practices or rural hospitals, lack the resources and expertise to implement comprehensive data security measures. Limited budgets, outdated infrastructure, and a lack of cybersecurity personnel exacerbate the vulnerabilities in these organizations, making them more susceptible to data breaches. There are several upcoming countries where there are very limited resources with minimal network infrastructure, minimal support from the government to support the data security initiatives of healthcare.

According to Hood, a survey was conducted in 2019 to understand the security and privacy of telehealth providers to understand the several programs of telehealth perspective to security and privacy of patient data. They identified problems like telehealth session information being stored in a user's computer rather than in a secured network or a remote device, providers are unaware if the telehealth software vendors use data encryption or not [7]. With HIPAA regulations enforcing data privacy and security having an awareness is important. Rural healthcare and telehealth need the required infrastructure to make sure patient's data is protected.

Human Factors and Insider Threats

A significant portion of data breaches in healthcare can be by human errors, such as misconfigured settings, improper data access controls, or poor password management. Additionally, insider threats from healthcare professionals or staff members who misuse their access to sensitive data pose a substantial security risk.

The role of humans in coping with cyberattacks and strengthening cyber defenses in an organization is grouped into the concept of "human factors" in cybersecurity. According to Reid, the role of humans in strengthening the cyber defense system within an organization is critical. His study shows that there are mainly three main categories of threats in healthcare, (i) attacks that result from a misconfiguration of the systems or firewall or network, (ii) ransomware attacks or phishing attacks that are launched to gain access to the data, and the employees of the organization reacts to respond not knowing that it is a phishing attack. (iii) emerging threat of exploiting human vulnerability [8].

Regulatory and Legal Compliance

Healthcare data security is governed by stringent regulations, such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in the European Union. Compliance with these regulations requires constant monitoring, auditing, and updating of security practices, making it a complex task for healthcare providers.

While regulatory compliance is for patient safety, this can be an expensive affair for providers who are in the system. It adds several layers of complexities, and understanding the policies, and implementation of the policies is critical. Healthcare organizations are partnering with cloud computing platforms, the data security is on the cloud platform provider, while enforcing responsibility is on the healthcare provider [9].

Data Security Strategies for Healthcare

Encryption

Encryption is one of the fundamental methods for securing healthcare data, both in transit and at rest. By converting sensitive data into unreadable ciphertext, encryption ensures that even if data is intercepted, it cannot be read without the corresponding decryption key. Most healthcare systems use end-to-end encryption for data transfer between different entities, such as hospitals and insurance providers.

Adedeji et al studied the performance of various encryption and decryption schemes used to secure medical data, the performance was assessed through their execution time, throughput, average data rate, and information entropy [10]. The popular encryption and decryption schemes are discussed below:

- **Advanced Encryption Standard (AES):** AES is considered highly efficient due to its strong security, efficiency, and flexibility [11]. It supports 18,192, and 256 bits. AES 256 is the most secure option.
- **Rivest-Shamir-Adleman (RSA):** RSA is an encryption algorithm that uses two keys, a public key for encryption and a private key for decryption. This is highly used for secure data transfers and digital signature requirements [12].
- **Elliptic Curve Cryptography (ECC):** ECC is a public-key (asymmetry) encryption scheme based on the algebraic structure of elliptic curves over finite fields [10].
- **ChaCha20-Poly1305:** This is a modern encryption algorithm designed for performance and security, especially where performance is critical like Mobile devices or IoT devices. Often used in healthcare mobile apps and wearable devices where high performance is required.

The two modes of data state for healthcare are: Data at Rest and Data in Transit. AES-256 is the preferred choice for encrypting the healthcare data stored in databases, file systems, and servers. For communication and data transfer (Data in Transit), RSA and ECC are commonly used in combination with TLS/SSL protocols.

Access Control and Authentication

Building robust user authentication is the first step of data security. An identifiable username and a secure password to authenticate the user's access is critical. Period password review policies and user accounts as well as system accounts make data more secure [13].

Healthcare systems need to incorporate Role-based access control (RBAC) and multi-factor authentication (MFA) are two critical components of healthcare data security. RBAC ensures that

individuals within the organization can access only the data that is necessary for their job function. For example, a physician need not know the overall revenue of the department, a person with a finance role should have access to the financial information. Identifying and assigning proper roles makes the data more secure. MFA adds an extra layer of security by requiring users to provide two or more forms of identification before accessing sensitive data [14].

According to Shakil et al discussed different levels of biometric data security are implemented to protect healthcare data. Confidentiality, Integrity, Non-repudiation – making sure the user accessing data does not deny their access or identity, Authentication. Cloud computing has become a prominent method for storing healthcare data, but cloud storage needs to be properly secured. Healthcare organizations use private or hybrid cloud models to store and process data securely. Skakil et al discussed a robust solution to implement data security for cloud applications [15].

AI and Machine Learning for Threat Detection

Artificial Intelligence (AI) and Machine Learning (ML) are increasingly being used to detect anomalies and potential security breaches in healthcare systems. These technologies can analyze vast amounts of data in real time, identifying patterns that may indicate malicious activity. For example, AI can flag suspicious login attempts, unauthorized access to sensitive data, or unusual data transmission patterns. Below is the list of a few areas how ML can take care of the proactive defensive mechanism:

- **Anomaly Detection:** Anomaly detection algorithms identify unusual patterns or deviations from normal behavior in the healthcare system [16]. Using ML, the unsupervised learning algorithms monitor the vast amount of network traffic, system logs, and user behavior to identify anomalies. Examples include accessing unusually high-volume patient records, abnormal login times, suspicious device connections, and alerts of the behavior as a potential threat.
- **Phishing Detection:** ML algorithms identify phishing attempts by analyzing the email's content URLS and the sender's behavior. ML-based phishing detection systems can scan emails for suspicious indicators such as fake sender addresses, unusual language, or links leading to fake login pages. According to the HIPAA (Healthcare Insurance Portability Act) journal, 91% of cyber-attacks come from phishing emails. According to Bonagiri et al, AI systems achieve high-efficiency rates ranging from 80% to 90% in identifying phishing attacks [16,17].
- **Malware Detection:** ML algorithms are trained on vast datasets of known malware signatures, system behaviors, and common attack techniques to recognize the characteristics of malware. Instead of relying solely on signature-based detection, ML can analyze file behavior and network traffic for suspicious activity. For example, ML algorithm can detect malware by observing unexpected behavior, such as encrypted files or unauthorized data transmission patterns within the hospital network [18].
- T**hreat Hunting:** ML can assist cybersecurity teams in proactive threat hunting by analyzing large datasets (e.g., network traffic, system logs) for indicators of potential threats. It automates much of the routine analysis and helps security teams focus on more complex, hard-to-detect threats. Threat hunting in healthcare could involve scanning for signs of advanced persistent threats (APTs) or insider threats that have evaded traditional defenses [16,17].
- **Fraud Detection:** ML can help detect fraudulent activities, such as falsifying insurance claims, identity theft, or financial fraud, by learning normal transaction patterns and flagging irregularities. Fraudulent activities in healthcare often involve billing discrepancies, identity theft, or misuse of patient data for financial gain. By analyzing transaction histories, claim patterns, and billing behaviors, ML can flag suspicious activities [17].
- **Ransomware Detection:** ML algorithms can identify patterns associated with ransomware attacks, such as abnormal file encryption or rapid data exfiltration. By using unsupervised learning, these algorithms can spot new or previously unknown types of ransomware. Healthcare organizations are frequent targets of ransomware due to the high value of medical data. ML can help detect ransomware attacks in progress and take actions to prevent the encryption of patient records or medical devices [18,19].
- **Machine Learning Algorithms and Implementation:** C. Feng, S. Wu and N. Liu, studied several ML algorithms trained the systems with sample data like Multi-layer Neural Networks (MNN), Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression (LR) [16]. In their model, MNN and RF algorithms worked well compared to other models. Over comparison of AI-based prediction worked 20% more efficiently than the rule-based model that is existing.

Blockchain Technology

Blockchain offers a promising solution for securing healthcare data. S. Baskar, K. Ramar, and H. Shanmugasundaram proposed a framework that illustrates how blockchain features can be utilized, this framework uses user's social security numbers and show their data with high security. Blockchain technology stores a sequence of transactions as blocks, each block being held by cryptographic keys (Hashes), these keys are saved in shared ledgers. Each node has a replica of the entire chain [20].

"Blockchain is a digital ledger in which transactions made in bitcoin or another cryptocurrency are recorded chronologically and publicly" [21]. Blockchain can be used for securely recording patient records, tracking data provenance, and ensuring that data cannot be tampered with without detection. Health information generated from EMR such as prescriptions, laboratory results, magnetic resonance imaging (MRI), electrocardiogram (ECG) and pathology results, and data from wearable devices would be encrypted and digitally signed before storing through the health blockchain.

Emerging Trends and Future Directions

Integration of AI and Blockchain

The integration of AI and blockchain technologies could significantly improve data security in healthcare. AI algorithms can be used to

identify security threats in real-time, while blockchain can provide an immutable record of patient data interactions. Together, these technologies could offer a new level of security and accountability.

Blockchain for cyber security with AI application technologically is a great combination to explore the data security vulnerabilities, in the health, and pharmaceutical industries. Blockchain with AI applications has been developed exponentially in several leading healthcare systems [21]. Blockchain architecture like Public Blockchain architecture, private Blockchain architecture, and Consortium Blockchain architecture, also Blockchain applications like logiboost, Medblock, OmniPHR are discussed by [22].

Quantum Computing and Data Security

Encryption is a technique used to protect the secrecy of the during the transfer and storage. Private keys and public keys are two categories of keys used in encryption. The use of a private key is also called symmetric encryption where encryption and decryption have the same value. Public key also referred to as asymmetric encryption provides a separate key for decryption compared to encryption. Several encryption models were designed using the key concepts discussed and each has its advantages and disadvantages. From 1994 (Peter Shor) to 2016 (several people from Google and IBM) challenged the processing power and the number of qubits required for quantum number supremacy. To break the AES-256 encryption it would take a quantum computer with 6,681 qubits. Shor's algorithm theoretically could break the RES 256 encryption with about 4096-bit key size. In summary, not all systems use quantum computing to break the encryption [23].

Quantum computing is still in its infancy but has the potential to disrupt current encryption methods. In the future, quantum-resistant encryption algorithms will be necessary to protect sensitive healthcare data from advanced quantum computing attacks. Healthcare organizations must begin preparing for this eventuality by researching and adopting quantum-safe encryption techniques.

Personal Health Data Protection (e.g., wearables)

With the increasing use of wearable devices and mobile health apps, personal health data is becoming more fragmented and diverse. Protecting this data, particularly when it is collected by third-party vendors, presents a significant challenge. Future data security frameworks will need to address how to securely manage data from IoT devices and ensure that it remains confidential and protected.

Wearable medical device has a new kind of health medical carrier, the collection and uploading method is different from the traditional medical data transfer model [24]. The prominent problems with wearable devices are data leakage and privacy protection. The potential solution for these wearable devices is to use the secure network imposed by a satellite network or cloud-based network.

Regulatory Evolution

As cyber threats evolve, so too must regulatory frameworks. Governments and international organizations will continue to update data protection laws to address new challenges, such as data localization, cross-border data sharing, and the security of digital health platforms. Healthcare organizations will need to stay ahead of these changes to ensure compliance.

According to Barlow, P., & Stuckler, D., globalization and health policies change the way national and international health needs are seen. with globalization, the freedom, scope, and mechanisms that governments have to design and implement health policies, 'policy space', can be shaped by a multitude of global actors and institutions with different interests, resources, and power. For example, it is well-recognized that multi-national food corporations, aid organizations, and financial agencies can all influence national health agendas, priorities, funding, and policies [25].

Recommendations for Healthcare Organizations

As healthcare organizations are prime targets for cyber attacks to gain access to the sensitive nature of patient data, robust security measures are essential to protect privacy, maintain the integrity of medical devices and systems, and ensure regulatory compliance. Below are a few key recommendations:

- **Invest in Comprehensive Cybersecurity Infrastructure:** Healthcare organizations should prioritize investments in up-to-date cybersecurity infrastructure, including firewalls, encryption, and secure data storage solutions, Blockchain, and AI.
- **Implement Robust Access Control Policies:** Role-based access control and multi-factor authentication should be standard practices in healthcare settings to ensure that only authorized individuals can access sensitive data.
- **Conduct Regular Security Audits:** Regular security audits and penetration testing can help identify vulnerabilities in healthcare systems and ensure that they are mitigated promptly.
- **Enhance Staff Training:** Continuous security training for healthcare professionals is essential to reduce the risk of human error and insider threats.
- **Stay Informed About Emerging Technologies:** Healthcare organizations should stay informed about emerging security technologies, such as AI, blockchain, and quantum-resistant encryption, to proactively secure data.
- **Collaborate with Industry Experts:** Healthcare providers should collaborate with cybersecurity experts and consult legal teams to ensure compliance with evolving regulations such as HIPAA.

Conclusion

Data security in healthcare is a critical issue that requires attention from all stakeholders, including healthcare providers, technology developers, regulators, and patients. While existing solutions such as encryption, access control, and AI-driven threat detection are effective in mitigating many risks, the ongoing evolution of cybersecurity threats calls for continuous innovation and adaptation. The future of healthcare data security will rely on emerging technologies such as AI, blockchain, and privacy-enhancing technologies to provide robust protection in a rapidly changing landscape. As healthcare systems continue to digitize and evolve, security must remain a top priority to safeguard

patient information, comply with regulations, and ensure the continued trust of the public in healthcare services.

References

[1] M Puppala, T He, X Yu, S Chen, R Ogunti, STC Wong. Data security and privacy management in healthcare applications and clinical data warehouse environment. 2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), Las Vegas, NV, USA. 2016; 5-8.

[2] Nick Culbertson, Former Forbes Councils Member. The Skyrocketing Volume of Healthcare Data Makes Privacy Imperative. 2021; https://www.forbes.com/councils/forbestechcouncil/2021/08/06/the-skyrocketing-volume-of-healthcare-data-makes-privacy-imperative/.

[3] A Jaleel, T Mahmood, MA Hassan, G Bano, SK Khurshid. Towards Medical Data Interoperability Through Collaboration of Healthcare Devices. in IEEE Access. 2020; 8: 132302-132319.

[4] Pine KH. The qualculative dimension of healthcare data interoperability. Health Informatics Journal. 2019; 25: 536-548.

[5] N Thamer, R Alubady. A Survey of Ransomware Attacks for Healthcare Systems: Risks, Challenges, Solutions and Opportunity of Research. 2021 1st Babylon International Conference on Information Technology and Science (BICITS), Babil, Iraq. 2021: 210-216.

[6] KS Bhosale, M Nenova, G Iliev. A study of cyber attacks: In the healthcare sector. 2021 Sixth Junior Conference on Lighting (Lighting), Gabrovo, Bulgaria. 2021: 1-6.

[7] Hood C. Telehealth cybersecurity. A Practical Guide to Emergency Telehealth, Oxford University Press, New York, NY. 2021: 81-92.

[8] Nifakos S, Chandramouli K, Nikolaou CK, Papachristou P, Koch S, et al. Influence of human factors on cyber security within healthcare organisations: A systematic review. Sensors. 2021; 21: 5119.

[9] Reid GA. Improving HIPAA Compliance Efforts with Modern Cloud Technologies (Doctoral dissertation, Capitol Technology University) 2021.

[10] KB Adedeji, NI Nwulu, C Aigbavboa, SL Gbadamosi. Assessment of Encryption and Decryption Schemes for Secure Data Transmission in Healthcare Systems. 2019 IEEE AFRICON, Accra, Ghana. 2019: 1-6.

[11] Rahmad C, Syulistyo AR, Sumari ADW. Securing the electronic medical record by implementing Advanced Encryption Standard (AES) on the information system of a health service place. In IOP Conference Series: Materials Science and Engineering IOP Publishing. 2021; 1073: 012057.

[12] Kumar U, Pathak RK, Kumar A. Handling secure healthcare data streaming using R2E algorithm. In 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC). 2020: 732-737.

[13] Shakil KA, Zareen FJ, Alam M, Jabin S. BAMHealthCloud: A biometric authentication and data management system for healthcare data in cloud. Journal of King Saud University-Computer and Information Sciences. 2020; 32: 57-64.

[14] Konstantinidis G. Identity and access management for e-government services in the European Union–state of the art review. 2021; https://hellanicus.lib.aegean.gr/handle/11610/23968.

[15] Kaul SD, Murty VK, Hatzinakos D. Secure and privacy preserving biometric based user authentication with data access control system in the healthcare environment. In 2020 International Conference on Cyberworlds (CW). 2020: 249-256.

[16] Adusumilli SBK, Damancharla H, Metta AR. AI-Powered Cybersecurity Solutions for Threat Detection and Prevention. International Journal of Creative Research In Computer Technology and Design. 2021; 3.

[17] C Feng, S Wu, N Liu. A user-centric machine learning framework for cyber security operations center. 2017 IEEE International Conference on Intelligence and Security Informatics (ISI), Beijing, China. 2017: 173-175.

[18] HM Farooq, NM Otaibi. Optimal Machine Learning Algorithms for Cyber Threat Detection. 2018 UKSim-AMSS 20th International Conference on Computer Modelling and Simulation (UKSim), Cambridge, UK. 2018: 32-37.

[19] S Kumar, BP Singh, V Kumar. A Semantic Machine Learning Algorithm for Cyber Threat Detection and Monitoring Security. 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India teristics of malware. Instead of relying solely on signature-based detection, ML can analyze file behavior and network traffic for suspicious activity. 2021: 1963-1967.

[20] S Baskar, K Ramar, H Shanmugasundaram. Data Security in Healthcare Using Blockchain Technology," 2021 International Conference on Decision Aid Sciences and Application (DASA), Sakheer, Bahrain. 2021: 354-359.

[21] F Xiaohua, C Marc, E Elias, H Khalid. Artificial Intelligence and Blockchain for Future Cyber Security Application. 2021 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), AB, Canada. 2021: 802-805.

[22] SM Tadaka, L Tawalbeh. Applications of Blockchain in Healthcare, Industry 4, and Cyber-Physical Systems. 2020 7th International Conference on Internet of Things: Systems, Management and Security (IOTSMS), Paris, France. 2020: 1-8.

[23] JK Cheng, EM Lim, YY Krikorian, DJ Sklar, VJ Kong. A Survey of Encryption Standard and Potential Impact Due to Quantum Computing. 2021 IEEE Aerospace Conference (50100), Big Sky, MT, USA. 2021: 1-10.

[24] C Yang, T Liu, L Zuo, Z Hao. An Empirical Study on the Data Security and Privacy Awareness to Use Health Care Wearable

Devices. 2019 16th International Conference on Service Systems and Service Management (ICSSSM), Shenzhen, China. 2019: 1-6.

[25] Barlow P, Stuckler D. Globalization and health policy space: Introducing the WTOhealth dataset of trade challenges to national health regulations at World Trade Organization, 1995–2016. Social Science & Medicine. 2021; 275: 113807.