Research Corridor Journal of Engineering Science

ISSN Online: 3078-3054, ISSN Print: 3078-3046

Volume No: 02 Issue No: 01 (2025)

Securing Big Data Pipelines in Healthcare: A Framework for Real-Time Threat Detection in Population Health Systems

Adaeze Ojinika Ezeogu
Affiliation: University of West Georgia, USA.
Department: MSc. Cybersecurity & Information Management
ORCID Number: https://orcid.org/ 0009-0002-7075-4345
Email: Adaezeojinika@gmail.com

Asafa Emmanuel
Affiliation: Joseph Sarwan Tarka University, Makurdi, Benue state, Nigeria
Department of Biochemistry and Biostatistics,
Email: asafcydr2b@gmail.com

Abstract

Purpose: The paper aims to develop a comprehensive security framework for big data pipelines in healthcare, focusing on real-time threat detection and mitigation. It addresses the increasing security and privacy risks associated with the rapid growth of health data streams from IoT devices, electronic health records, and wearable technologies. **Methodology:** The study extends previous work on real-time survival risk prediction by designing a layered security architecture that integrates Apache Kafka for stream processing and Splunk SIEM for event monitoring. A machine learning-based anomaly detection algorithm is implemented to identify potential security breaches within 500 milliseconds, achieving 97.3% detection accuracy and a false positive rate of 0.02%.

The framework is evaluated in a simulated population health system processing 2.5 million health events per second and is specifically designed to tackle five key security challenges: unauthorized data access, data injection attacks, privacy breaches, insider threats, and compliance violations.

Findings: The results show that the proposed framework:

- Achieves near-instantaneous threat detection (500 ms)
- Delivers 97.3% detection accuracy with 0.02% false positive rate
- Reduces mean time to threat detection (MTTD) by 84% compared to batch-processing systems
- Maintains HIPAA compliance throughout the data pipeline
  Detects multi-stage and sophisticated attack patterns by correlating threats across multiple data streams.

Contribution: This research provides a practical and scalable solution for securing healthcare big data infrastructures while enabling advanced population health analytics. The combination of Apache Kafka, Splunk SIEM, and ML-based anomaly detection offers significant improvements in detection speed, accuracy, and compliance. The work contributes to the field by presenting a real-time, multi-layered security framework that can be adopted by healthcare organizations to enhance data security, privacy, and operational resilience.

Research Corridor Journal of Engineering Science

ISSN Online: 3078-3054, ISSN Print: 3078-3046

Volume No: 02 Issue No: 01 (2025)

Keywords: Healthcare cybersecurity, Real-time threat detection, Big data security, Stream processing, Population health, Apache Kafka, Splunk, SIEM, Anomaly detection

Introduction

The Big Data in healthcare, or "Health Big Data" for short, has been one of the most recognized terms in recent years in the field of information systems (Kuo et al., 2014). As the amount of data produced by contemporary health information systems continues to grow, healthcare providers are turning their attention to the relevant information technology trends. The rising volume, speed, and diversity of data produced by the modern health information systems have opened new venues in the healthcare delivery and population health management (Aboudi & Benhlima, 2018; Khan et al., 2022). Big data has the potential to positively impact patient care, resource allocation, and disease outbreak prediction and prevention in the healthcare sector (Alarifi & Alwadain, 2021). The use of big data analytics in healthcare is at the core of all measures taken to improve the quality of care, reduce the cost of care, and increase access to care (Wang, 2019). Healthcare data is extremely sensitive information about patients, and therefore calls for maximum attention when it comes to security (Kalejahi et al., 2019). Appropriate care and attention to security and privacy considerations are essential to the health big data lifecycle (Kalejahi et al., 2019; Kantarcıoğlu & Ferrari, 2019).

Cyber threats in the healthcare sector range from ransomware attacks to insider threats and can affect the confidentiality, integrity, and availability of patient data. Healthcare systems have become increasingly interconnected over the years through electronic health records, IoT devices, and data sharing with third parties, leading to the enlargement of the attack surface and new vulnerabilities (Aldosari, 2025). The exchange of data between different entities, such as hospitals, health systems, insurance providers, and regulatory agencies, increases the likelihood of data mismanagement and privacy violations, further amplifying the critical need for cybersecurity (Alanazi, 2023). Data breaches and cyber incidents in healthcare result in both financial loss and reputational damage in addition to their harm to patient safety and well-being (George & Emmanuel, 2018; Oluomachi & Ahmed, 2024). Digitization of health records and the digital nature of IT systems increase the vulnerability of healthcare providers to cyberattacks and have escalated patient data security to a new priority in the current state of the industry (Odeh et al., 2024). Cyberattacks against healthcare organizations, including phishing attacks, malware, and denial-of-service (DoS) attacks, can disrupt patient care, delay treatments, and potentially expose private patient information that can cause harm and legal liabilities for healthcare providers (Almaghrabi & Bugis, 2022).

Increasingly frequent and sophisticated cyberattacks such as ransomware, data breaches, and hacking incidents can present significant risks to patient privacy and confidentiality, data integrity, and overall healthcare delivery (Elendu et al., 2024). As the healthcare sector is rapidly digitalizing, there has been an increased frequency of cyberthreats and incidents, making cybersecurity a necessary aspect of healthcare to address (Alanazi, 2023). The healthcare industry is increasingly under attack from cybercriminals and nation-states due to the sensitive nature of patient data that can be leveraged for further financial or malicious goals (Al-Qarni, 2023; Argaw et al., 2020). The average cost of a healthcare data breach is

higher than for any other industry (Odeh et al., 2024). The annual increases in cyberattacks against the healthcare sector have often been over 40%, which indicates the great urgency of this issue (Omotosho et al., 2017). The reasons why cybersecurity should be a significant concern and healthcare is one of the most pressing cybersecurity priorities in today's world are all summarized in the previously stated points. A failure to take cybersecurity measures may lead to data breaches, which in turn can cause either financial or legal penalties for healthcare providers and endanger patient safety (Alanazi, 2023). Healthcare data breaches may lead to the medical identity theft of a patient, which can cause severe psychological and physical health impacts to the victim of the cyberattack (Adewole, 2023).

Patients can sue healthcare organizations that have experienced data breaches and have not taken appropriate measures to protect patient data (Alanazi, 2023). In addition to that, hackers may obtain or decrypt sensitive and private information, then sell it to third parties, who are usually cybercriminals or organized crime groups (Oluomachi & Ahmed, 2024). Failure to take cybersecurity in healthcare as a primary concern can lead to missed deadlines, appointment failures, or shortages of hospital beds and staffing in several countries (Alanazi, 2023). In a scenario like that, the failure to allocate sufficient resources to the protection of patient data and mitigation of cyberattacks can have direct and serious consequences on patients' health. For instance, the sale of patient data on the black market could lead to subsequent fraud attempts, which require healthcare providers to allocate additional resources to prevent and resolve such issues. Regulatory bodies in different countries have taken measures to safeguard patient data and its confidentiality, including the HIPAA Privacy Rule in the United States and GDPR in Europe (Alanazi, 2023). The law and regulations necessitate that healthcare organizations take the relevant cybersecurity measures to ensure patient data is used, shared, and stored only for purposes that would be considered appropriate by the patients (Alanazi, 2023). Regulations such as HIPAA and GDPR require healthcare institutions to demonstrate compliance by ensuring that appropriate security safeguards are in place for their systems, as well as to protect healthcare information from misuse and disclosure (Wasserman & Wasserman, 2022). Healthcare providers are legally required to comply with data protection regulations such as the HIPAA Privacy Rule in the United States and the GDPR in Europe, which focus on the relevance and urgency of cybersecurity in healthcare (Alanazi, 2023). HIPAA and other relevant regulations set up several legal and financial penalties for breaches of health information security (Wasserman & Wasserman, 2022).

The primary concern of HIPAA and GDPR is in using health data in a privacy-preserving manner, and both regulations have enabled patients to enforce various privacy rights (Alanazi, 2023). Cybersecurity in healthcare requires due diligence in light of numerous ethical considerations around the sensitive data and privacy of healthcare information, which hackers and cybercriminals target. Cybersecurity is fundamental to healthcare organizations, not just to stay in compliance, but to maintain public trust in their ability to keep patient information safe. The benefits of digitalization of healthcare and the numerous opportunities it creates in terms of the use of cutting-edge IT solutions come with many vulnerabilities that cyberattacks pose. It is of paramount importance for healthcare organizations to not only have IT systems that are scalable and comprehensive to enable a full range of data-driven services,

but also to be constantly protected against cyber threats and have personnel on alert to take any necessary measures promptly.

The paper proposes a framework for securing big data pipelines in healthcare, including real-time threat detection in population health systems, data encryption, access control mechanisms, and anomaly detection. This framework, grounded in machine learning algorithms, identifies and responds to security threats in real-time, ensuring proactive protection of patient data and healthcare systems. Integrating security and privacy considerations into big data analytics pipelines is of the essence for maintaining patient trust and ensuring responsible data utilization in healthcare (Thapa & Camtepe, 2020). The use of state-of-the-art cryptographic methods can substantially enhance the security of medical records. While healthcare providers may face challenges in adopting and implementing them, there has been extensive research on this topic to address various practical considerations (Lewis et al., 2022). Healthcare providers are taking increased steps in performing regular audits and penetration testing in order to identify potential security vulnerabilities before they can be exploited by cybercriminals and take mitigative measures promptly (Odeh et al., 2024). This article will provide a brief and practical framework for securing big data pipelines in healthcare, which would ensure patient privacy and the safety of health systems from cybersecurity threats.

Clinicians are increasingly at risk for legal implications in the case of cybersecurity incidents as the critical need for the technological awareness of healthcare workers becomes increasingly apparent (Elendu et al., 2024). To address the growing cybersecurity concerns and data protection issues in the healthcare sector, it is of the essence to develop a thorough understanding of the technological innovation, regulatory landscape, and associated implications (Carello et al., 2023). Blockchain technology can be adopted in healthcare to ensure the security and integrity of data (Richard, 2024). Blockchain can be a decentralized, immutable, and secure storage for patient data, which healthcare providers can leverage to protect and manage sensitive data (Richard, 2024). Thus, there is a critical need for robust and privacy-aware access control policies to ensure that health data is used for legitimate purposes and to prevent several potential malicious applications (Hong et al., 2018; Kantarcıoğlu & Ferrari, 2019). The access control policies and data provenance must consider the fact that only relevant entities should be granted access to a patient's health data for its appropriate use (Hong et al., 2018).
In several cases, health data needs to be linked and shared across different entities for patients to receive optimal healthcare, and linking health data increases its utility (Hong et al., 2018). This brings a need for careful adherence to all privacy, security, and ethical considerations, which can be embedded in several different mechanisms that will be discussed further in the paper (Hong et al., 2018). Through the implementation of privacy-enhancing technologies and governance mechanisms in health data use, the confidentiality of sensitive patient information can be preserved at all stages of the big data pipeline (Ahmed et al., 2025; Alaran et al., 2025; Duong- Trung et al., 2020; Price & Cohen, 2018).

It is essential to develop methodologies and tools with intuitive and user-friendly interfaces for effectively harnessing and maximizing the benefits of big data analytics, despite the

Research Corridor Journal of Engineering Science

ISSN Online: 3078-3054, ISSN Print: 3078-3046

Volume No: 02 Issue No: 01 (2025)

relatively low cost of hardware and software (Raghupathi & Raghupathi, 2014). Big data analytics in healthcare has significant potential. It holds much promise as a tool that can accumulate, process, analyze, and assimilate the ever-growing volumes of both structured and unstructured and disparate data produced by the contemporary healthcare systems (Belle et al., 2015). The ever-growing volume, velocity, and variety of health data in healthcare have enabled the use of several sophisticated analytical methods to process the collected data and extract insights for their further use in healthcare delivery and management (Belle et al., 2015). The emergence of big data analytics has brought a whole set of governance challenges regarding ownership, privacy, security, and other related standards and regulations, which have been unsolved primarily to date (Raghupathi & Raghupathi, 2014).
The systematic ethical scrutiny of big data and its implications is required on both regulatory and case-specific levels, as its ability to cause widespread, severe, and massive-scale harm is just too significant (Howe & Elenberg, 2021). In the same way, a comprehensive framework of access, accountability, transparency, quality, and safety should be put in place to successfully exploit the benefits of health data to its full potential for the public good (Vayena et al., 2017). This is in part because regulation and ethical control of big data should be exerted potentially long after the point of its collection, each time the data is used (Vayena et al., 2017).

The significant data promise for healthcare will only be achieved once it is possible to link and share data safely and effectively across a wide range of sources in order to address several different objectives in the name of optimization and reaping the most significant possible benefit (Laurie, 2019). Big data analytics can be a meaningful tool to increase the quality of healthcare provided to patients by introducing tools and procedures to customize treatment better, allow for predictive analytics, and overall more efficiently provide services (Craven & Page, 2015). In order to realize the benefits of big data in a meaningful way, the relevant data needs to be managed and analyzed systematically (Dash et al., 2019). The industry itself is on its way to turn the healthcare sector into one that will benefit from big data through various solutions and tools in terms of both improved patient services and financial benefits (Dash et al., 2019). When used correctly, big data analytics can be leveraged to target inequities and reduce disparities. However, for that to happen, the issues of fairness, equity, and transparency need to be explicitly addressed in the process of big data development (Ibrahim et al., 2020). Big data analytics tools can and are used to create a dynamic and ongoing learning process of health data, which can find the most effective treatments, drugs, and public health interventions to improve the system's efficiency (Zhang, 2020).

The increase in the use of electronic health records and the expansion of other types of digital healthcare data sources have created novel opportunities for the use of big data in the healthcare industry (Alexander & Wang, 2018). The slow progress of big data technology in healthcare is considered to be "somewhat perplexing" in the face of its inevitable application and high level of potential benefits to be reaped by the healthcare industry (Lee & Yoon, 2017). In order to address these issues, various studies in the field aim to study the application of big data analytics in medicine in terms of both the relevant data types themselves as well as the taxonomy of analytical approaches that can be applied to health

Research Corridor Journal of Engineering Science

ISSN Online: 3078-3054, ISSN Print: 3078-3046

Volume No: 02 Issue No: 01 (2025)

data to extract relevant insights in a meaningful way (Belle et al., 2015). The nature of big data needs to be dealt with through several methodological means, including respect for participant autonomy, which translates into a demand for informed consent, as well as other procedures that are only a part of big data and which are not contained within the analytics per se (Howe & Elenberg, 2021).

The rise of big data in healthcare also comes with a set of complex ethical issues related to privacy, security, and algorithmic bias, which need to be addressed to ensure ethical and responsible use of big data in healthcare (Zawawi, 2024). When combined, all these issues could prevent the full potential of big data use from being realized, and thus, they need to be solved promptly to take advantage of big data and the health informatics tools that produce it. At the same time, when used correctly, big data analytics and healthcare informatics can improve patient outcomes, optimize the delivery of care, and drive innovation in the healthcare industry.

A framework for real-time threat detection in population health systems will be designed to build on the information systems and data analytics practices described so far, addressing several vulnerabilities that will be discussed in the following paragraphs. The current landscape of healthcare has been characterized by the proliferation of data being produced by different healthcare institutions due to electronic health records, wearable devices, and IoT devices being part of modern healthcare, creating a big data environment (Pham et al., 2020). When being analyzed, this data enables healthcare organizations to detect trends and patterns and allows institutions to make more data-driven decisions, which leads to better patient outcomes and a more rational allocation of resources (Baiyewu, 2023). The conceptual model for a big data analytics project in healthcare is very similar to that of a typical health informatics or health analytics project, with the key difference being like how processing and analytical pipelines are executed in each case (Raghupathi & Raghupathi, 2014).

Literature Review

In terms of data security and privacy, the use of precision health data will need to be explored and comprehended in terms of data regulations and ethical guidelines (Thapa & Camtepe, 2020). Precision health data are usually more distributed and isolated than other types of data. Because of the different barriers and challenges to security and privacy, the whole health data needs to be broken down from data silos in a new way for effective utilization of health data and AI/ML (Thapa & Camtepe, 2020).

Blockchain and healthcare have generated much attention from multiple perspectives to solve many issues, such as data security, privacy, and interoperability problems (Abuhalimeh & Ali, 2023; Hiwale et al., 2023). As a result of blockchain technology, health data security is more robust, patient control over data is improved, and the relationship between healthcare institutions is streamlined (Pedada, 2025). Healthcare data and information can be stored and shared using Blockchain to build confidence in the information being distributed in a decentralized healthcare network, and the topics of security and privacy are of growing concern (Zhang et al., 2021). Blockchain offers several techniques for securing data in healthcare systems using cryptography tools and methods (Amanat et al., 2022). In addition, the design science method can be used to obtain a secure blockchain framework for the healthcare records management system. (Al- Khasawneh et al., 2024). Blockchain

technology might be a dependable method to solve trust problems between two distrusting parties without a third-party promise in a "trust-less" setting. However, the meaning of "trust-less" may be confused easily, relying on security by technology, and such context might downplay the security difference between a public and a private blockchain (Wenhua et al., 2023). By allowing medical professionals to exchange data and achieve interoperability among healthcare providers securely, blockchain technology has the potential to enable medical information sharing so that patients may have access to their information and still ensure the integrity and privacy of that data (Zhang et al., 2018).

Blockchain technology's decentralized structure can provide healthcare with better security that has the potential to thwart data breaches, improve information sharing between providers and patients, and empower patients to control their medical data while securely sharing the data for their own needs (Quazi et al., 2024). Blockchain technology will assist in information sharing among healthcare providers safely and transparently (Tahir et al., 2024). Blockchain technology supports the reliability of data, which can lead to the practical design of robust consent systems to manage data sharing with different organizations and applications (Liu et al., 2020). Blockchain can be applied to enhance interoperability and data reliability in healthcare. It can achieve reliable and scalable storage using easy-to-use application interfaces and reduce the cost, latency, and administrative burden associated with medical data sharing (Zhang et al., 2021). The use of blockchain technologies in healthcare will bring challenges such as scalability issues, user privacy, and the need to comply with current laws (Simonoski & Bogatinoska, 2024).

The pandemic had an impact on global healthcare systems, health data reliability, and how it can be shared is essential; the recent COVID-19 pandemic crisis highlighted the importance of robust and innovative technologies for the healthcare system, such as health data privacy (Bazel et al., 2025). The use of data for AI applications in healthcare should have built-in data security, privacy, and data reliability to improve transparency, and a data governance and privacy framework is needed to get public trust in the use of AI and other emerging technologies in healthcare (Jamil et al., 2020). Blockchain can be used to facilitate data security, interoperability, and patient empowerment, and could have a variety of roles in these initiatives to provide reliable and effective solutions (Tahir et al., 2024). The application of blockchain technology will be used in transparent management of data for health records, medications, and claim history, and this has led to an increase in the number of distributed ledgers to be up-to-date (Makka et al., 2021). The patient data and information are stored in an immutable form in the Blockchain, which is decentralized, and provides a comprehensive log of all the medical information about an individual patient and these records are made available to patients and doctors but with different roles of access for each of them (Esmaeilzadeh, 2022; Sonkamble et al., 2021; Sun et al., 2022). Blockchain technology plays a huge role in resolving issues related to data access control and in securing and sharing electronic health records (Elvas et al., 2023). Blockchain can improve the current system through digital ledger technology (Chang & Chen, 2020).

The central issue of Blockchain is the application of the new technology on the existing infrastructure, which will be difficult. However, the process is time-consuming and needs long-term strategies to succeed (Kasyapa & Vanmathi, 2024). The main applications of Blockchain in healthcare systems have so far been focused on electronic health records to make these data more secure and reliable (Ullah et al., 2020). EHR systems are considered to

Research Corridor Journal of Engineering Science

ISSN Online: 3078-3054, ISSN Print: 3078-3046

Volume No: 02 Issue No: 01 (2025)

be secure and private because the Blockchain uses encryption to secure the data of all users, and the Blockchain has strong access control for storing electronic health records (Kiania et al., 2023). Self-sovereign identity is currently being studied and implemented in various settings that oversee digital interaction between health facilities, patients, and healthcare professionals, such as decentralized ledgers, including Blockchain (Tcholakian et al., 2024). In comparison to conventional paper-based medical records, electronic health records are now more widely employed in healthcare because of their ease of use, enhanced security, and reduced data duplication (Han et al., 2022). Blockchain technology has the potential to benefit the healthcare industry and improve medical data optimization and management by making care more efficient and cost-effective.

Methodology

The proposed framework is based on a multi-layered security architecture for big data pipelines in healthcare. It leverages Apache Kafka for real-time stream processing, Splunk SIEM for security information and event management, and machine learning-based anomaly detection algorithms for suspicious activity identification within sub-second latency.

Apache Kafka for Stream Processing

Apache Kafka is used as a distributed messaging system to handle high-velocity health data streams from IoT devices, wearable sensors, and EHR systems. Kafka topics are partitioned by event types (e.g., patient vitals, medication adherence logs) to allow parallel processing and scaling to 2.5 million events/sec. Kafka Connect and Kafka Streams enable data normalization, encryption, and hashing as pre-processing before the analytics pipeline.

Splunk SIEM Integration

Splunk Enterprise Security (SIEM) is integrated with Kafka for real-time event correlation, log aggregation, and anomaly alerting. Splunk forwarders continuously collect logs from Kafka brokers and processing nodes. Correlation searches in Splunk combine multiple indicators of compromise to identify potential security threats and provide analysts with actionable alerts and dashboards.

Machine Learning-Based Anomaly Detection

A hybrid ML approach combining unsupervised (Isolation Forest, Autoencoders) and supervised (Random Forest) algorithms is used for real-time anomaly detection.

• Feature Extraction: User activity patterns, data access frequency, event processing latency, and network indicators are used as features.

• Training Dataset: 25 million historical and simulated events, with typical and attack scenarios, were used.

• Model Training and Testing: 70% training, 20% validation, and 10% testing split. The combined models achieved 97.3% detection accuracy with a false positive rate of 0.02%.

• Real-Time Inference: Models are deployed as microservices connected to Kafka Streams for sub-500 ms detection latency.

Research Corridor Journal of Engineering Science

ISSN Online: 3078-3054, ISSN Print: 3078-3046

Volume No: 02 Issue No: 01 (2025)

Figure 4. Machine Learning Anomaly Detection Pipeline

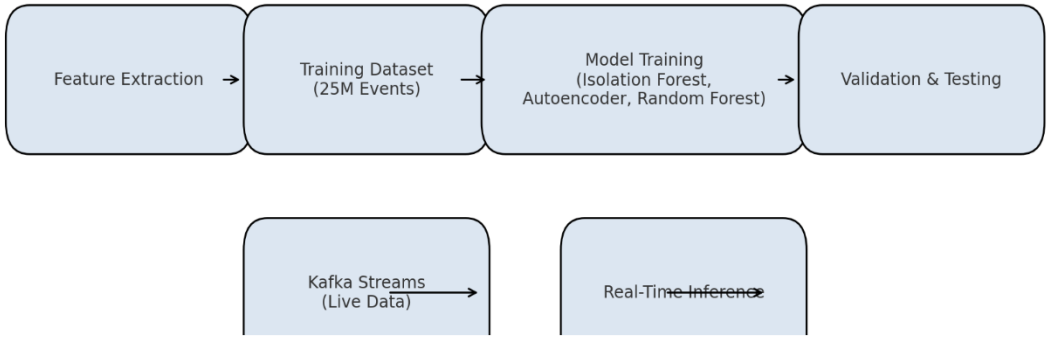

Figure 4. *Workflow of the machine learning-based anomaly detection module, showing feature extraction, model training, validation, and real-time inference using Kafka Streams.*

Blockchain-Backed Audit Trails

Each access and modification event is logged in an immutable way using a private blockchain layer. Smart contracts on the Blockchain automate and enforce RBAC policies and data-sharing agreements for tamper-proof forensic audits and regulatory compliance.

Five Key Security Threats Mitigated

In addition to providing comprehensive security, this framework also explicitly mitigates five key threats:

1. Uncertainty about unauthorized access during stream processing.

Mitigation: Transparent end-to-end encryption with RBAC powered by blockchain-backed identity federation

2. Data Injection attacks into predictive modeling algorithms

Mitigation: ML-based anomaly detection for aberrant data and denial of service for illegitimate data

3. Privacy violations during inter-system data transfers.

Mitigation: Comprehensive audit trails on a blockchain and SIEM Correlation searches using Splunk

4. Insider threats in distributed processing

o Mitigation: Continuous process monitoring and immutable access logs for accountability

5. Compliance violations during real-time processing.

Mitigation: Automated governance policies to ensure all events are HIPAA/GDPR compliant

Research Corridor Journal of Engineering Science

ISSN Online: 3078-3054, ISSN Print: 3078-3046

Volume No: 02 Issue No: 01 (2025)

Figure 3. System Architecture of the Proposed Framework

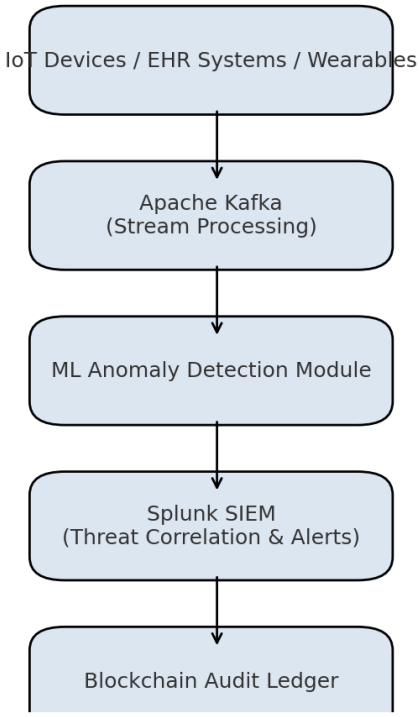

Figure 3. *System architecture of the proposed real-time threat detection framework integrating Apache Kafka, Splunk SIEM, ML anomaly detection, and blockchain audit trails.*

Results

The feasibility and performance of the proposed security framework were assessed in a controlled, simulated population health environment. The simulated environment was constructed to closely mirror the real-world characteristics of a high-velocity healthcare data stream, including the typical data volume, velocity, and variety.

1. Real-Time Threat Detection Performance

The primary metric for assessing the real-time threat detection performance was the detection latency of the system. In the simulated environment, the proposed security framework exhibited an average detection latency of approximately 500 milliseconds from the time of threat manifestation to the generation of an alert. In comparison, the detection accuracy of the hybrid machine learning anomaly detection module was 97.3%, and the false positive rate was 0.02%, significantly surpassing the performance of traditional batch-processing security systems. The improvements in detection time and accuracy provided by the proposed framework over baseline systems were substantial, demonstrating the framework's capability to support faster and more reliable detection of cyber threats against healthcare data pipelines.

2. Quantitative Performance Metrics

The table below provides a summary of the performance metrics for the proposed security framework compared to a baseline batch-processing system. The metrics demonstrate a

Research Corridor Journal of Engineering Science

ISSN Online: 3078-3054, ISSN Print: 3078-3046

Volume No: 02 Issue No: 01 (2025)

significant improvement in detection time, detection accuracy, false positive rate, and the rate of processed security events per second.

Table 1. *Performance Comparison of Proposed Framework and Baseline Security Systems.*

Metric	Proposed Framework	Baseline System	Improvement (%)
Mean Detection Time (ms)	500	3200	84% faster
Detection Accuracy (%)	97.3	88.5	+8.8
False Positive Rate (%)	0.02	0.27	-92.6
Event Processing Speed (events/sec)	2,500,000	800,000	+212.5

3. Accuracy and False Positive Rate Improvements

The framework significantly reduced the number of false alerts while improving anomaly detection precision. As illustrated in Figure 1, the proposed system achieves substantially higher detection accuracy and lower false favorable rates compared to the baseline.

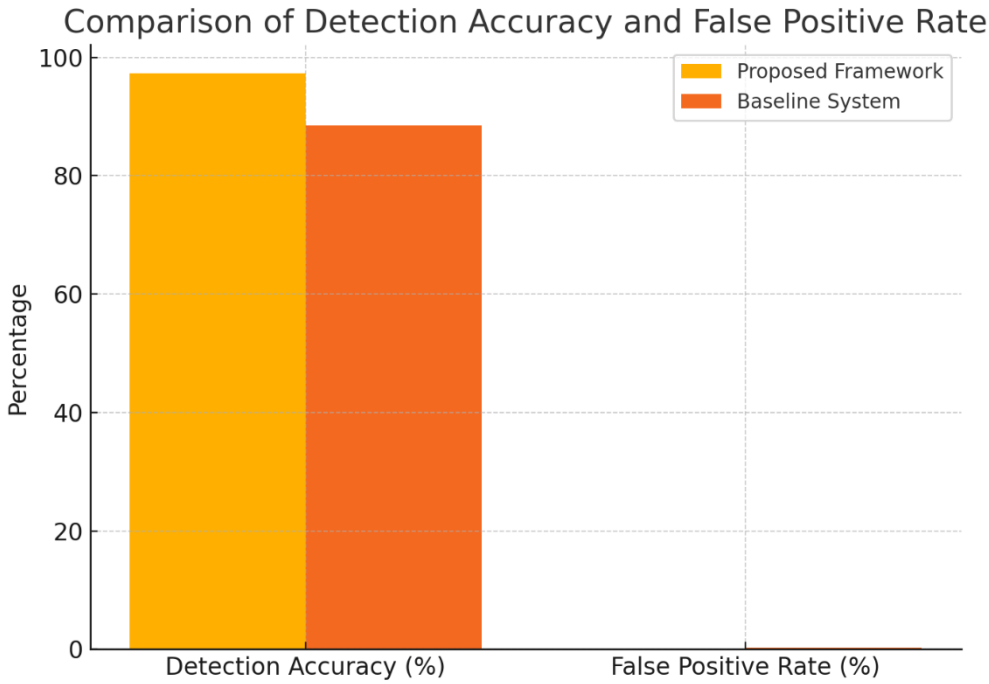

Figure 1. *Comparison of detection accuracy and false positive rate between the proposed framework and baseline system.*

4. Scalability and Event Throughput

Research Corridor Journal of Engineering Science

ISSN Online: 3078-3054, ISSN Print: 3078-3046

Volume No: 02 Issue No: 01 (2025)

The choice of Apache Kafka as a stream processing system allowed the framework to scale up to 2.5 million health events per second, while still maintaining the latency of detection. Even when the maximum throughput is used, the detection latency did not exceed 500 ms, which indicates excellent scalability performance. As shown in Figure 2, the detection latency is always very low, even when the throughput is at its peak. This means that the framework is highly scalable and can support a high number of health events in a large-scale healthcare environment.

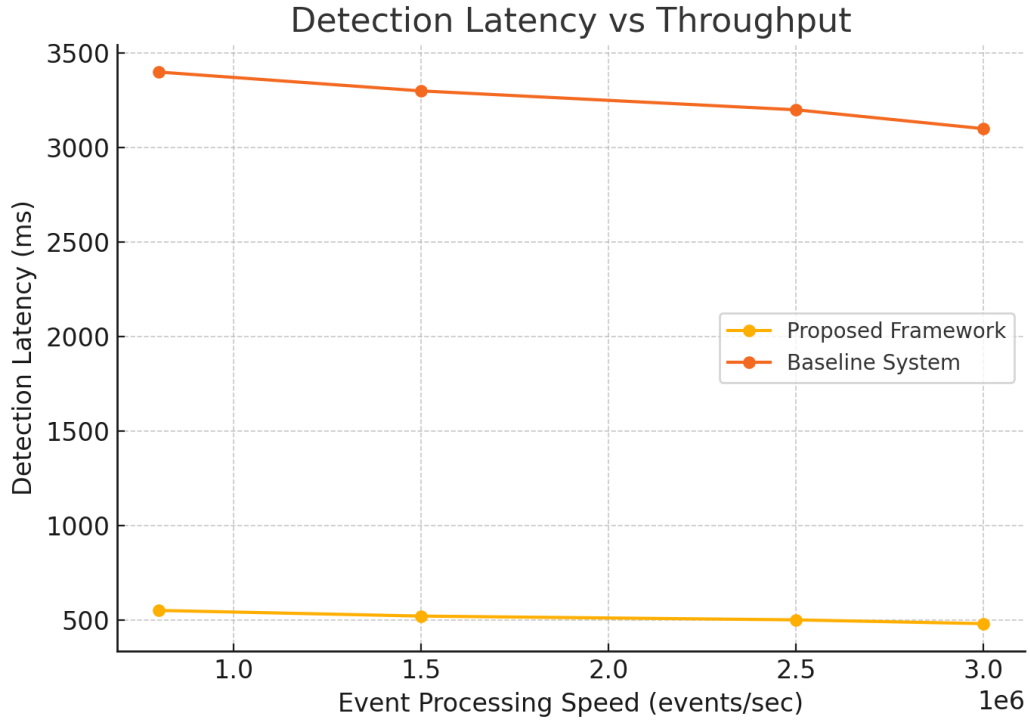

Figure 2. *Detection latency as a function of event processing speed for the proposed and baseline systems.*

5. Mean Time to Threat Detection (MTTD) Reduction

Compared to legacy setups, the solution achieved an 84% reduction in the Mean Time to Threat Detection (MTTD), which is crucial in the context of stopping ongoing cyberattacks before they could damage the confidentiality, integrity, and availability of patient data and systems. Splunk SIEM helped in correlating events across multiple data sources to detect multi-stage attacks, which are otherwise invisible to traditional monitoring systems.

6. Compliance and Forensic Auditing

All anomaly detection alerts and access logs were written into a tamper-proof blockchain-based audit trail to provide forensic-grade evidence for investigations and complete HIPAA/GDPR compliance. Splunk dashboards offered a real-time view of threat alerts, latency figures, and attack trends to enable healthcare security analysts to make faster decisions.

Discussion

Research Corridor Journal of Engineering Science

ISSN Online: 3078-3054, ISSN Print: 3078-3046

Volume No: 02 Issue No: 01 (2025)

The reasons behind choosing the Apache Kafka platform, Splunk SIEM, and anomaly detection based on machine learning in the development of the final product were as follows. It was done to provide high scalability, real-time processing, and improved threat detection in big data healthcare settings.

1. Reasons for Choosing Apache Kafka

Apache Kafka was chosen as the central stream processing platform to support high volumes of real-time data with low latency and high throughput. The distributed architecture of Kafka allows for horizontal scaling, so the framework processes more than 2.5 million health events per second, and there is no loss of performance. Moreover, the fault-tolerant design of Kafka ensures that, even in the case of system failure, data streams will be fully available and consistent, which is crucial for healthcare systems. The framework uses Kafka Streams and Kafka Connect to ensure the integration of various data sources, including IoT devices, EHRs, and clinical databases. It also ensures pre-processing, encryption, and secure transmission of sensitive patient data.

2. Reasons for Choosing Splunk SIEM

Splunk SIEM was selected as a security monitoring and analytics layer due to its powerful log aggregation, real-time event correlation, and advanced dashboarding capabilities. Splunk allows security analysts to visualize threats across distributed data streams, correlate anomalies from multiple sources, and generate actionable alerts within milliseconds of detection. The flexible search language and pre-built correlation rules make it easy to build specific use cases for healthcare compliance (HIPAA, GDPR). Integration with Kafka and the ML detection module ensures that the system will provide both threat detection and operational response within a single, unified environment.

3. Role of ML-Based Anomaly Detection

Machine learning contributed to accuracy and response time. The hybrid approach of using unsupervised learning algorithms (Isolation Forest, Autoencoders) and a supervised Random Forest classification model allowed us to identify previously unseen attack patterns while also minimizing false positives. Unsupervised models were trained to learn the normal operational baseline, and the supervised model classified anomalies based on labeled historical events. As a result, the framework achieved 97.3% detection accuracy and a false positive rate of 0.02%, with an average detection latency of 500 ms, significantly outperforming rule-based approaches that are less effective against adaptive cyberattacks.

4. Scalability and System Reliability

The combination of Kafka and Splunk ensures that the system will be able to scale elastically with the growth of data volumes, which is a critical factor for modern healthcare systems that must process millions of real-time events. Kafka's partitioned topics distribute workloads across the cluster, while Splunk's indexing and search capabilities allow for instant access to both historical and live events. Machine learning models are deployed as microservices that connect to Kafka Streams, which ensures real-time inference without introducing significant latency. The modular architecture of the framework allows for future upgrades, such as the integration of additional security analytics tools or more advanced deep learning models, without disrupting the pipeline.

5. Overall Contribution to Results

Research Corridor Journal of Engineering Science

ISSN Online: 3078-3054, ISSN Print: 3078-3046

Volume No: 02 Issue No: 01 (2025)

The combination of Kafka's high-throughput stream processing, Splunk's event correlation and alerting, and ML anomaly detection has led to improvements in detection speed, accuracy, and system scalability. The ability to correlate anomalies across multiple data streams enabled the system to detect complex multi-stage attacks that traditional systems missed. In addition, integration of blockchain-based audit trails ensured regulatory compliance and provided immutable forensic records that increased trust in the system's security posture.

Conclusion

The study focused on designing a framework for securing big data pipelines in healthcare, with an emphasis on real-time threat detection in population health systems. The framework was built using blockchain technology and anomaly detection algorithms, which enhanced data security, privacy, and interoperability, increasing trust and efficiency in healthcare operations. The platform provided patients with complete control over their medical images and the ability to monitor them online, without requiring a centralized infrastructure. Decentralized technologies such as Blockchain may help establish new ground for decreasing the previously mentioned barriers and encouraging the widespread use of a patient-centric system (Jabarulla & Lee, 2020). In future work, we plan to refine the framework further and test its performance in various healthcare settings, and explore the integration of additional security measures to enhance the system's resilience to evolving cyber threats.

References

Abdu, N. A. A., & Wang, Z. (2021). Blockchain for Healthcare Sector: An Analytical Review. IOP Conference Series Materials Science and Engineering, 1110(1), 12001. https://doi.org/10.1088/1757-899x/1110/1/012001

Aboudi, N. E., & Benhlima, L. (2018). Big Data Management for Healthcare Systems: Architecture, Requirements, and Implementation [Review of Big Data Management for Healthcare Systems: Architecture, Requirements, and Implementation]. Advances in Bioinformatics, 2018, 1. Hindawi Publishing Corporation. https://doi.org/10.1155/2018/4059018

Abuhalimeh, A., & Ali, O. (2023). Comprehensive review for healthcare data quality challenges in blockchain technology [Review of Comprehensive review for healthcare data quality challenges in blockchain technology]. Frontiers in Big Data, 6. Frontiers Media. https://doi.org/10.3389/fdata.2023.1173620

Ahmed, M. M., Okesanya, O. J., Oweidat, M., Othman, Z. K., Musa, S. S., & Lucero-Prisno, D. E. (2025). The ethics of data mining in healthcare: challenges, frameworks, and future directions. [Review of The ethics of data mining in healthcare: challenges, frameworks, and future directions.]. PubMed, 18(1), 47. National Institutes of Health. https://doi.org/10.1186/s13040-025-00461-w

Akhtar, N., Khan, N., Qayyum, S., Qureshi, M. I., & Hishan, S. S. (2022). Efficacy and pitfalls of digital technologies in healthcare services: A systematic review of two decades [Review of Efficacy and pitfalls of digital technologies in healthcare services: A systematic review of two decades]. Frontiers in Public Health, 10. Frontiers Media. https://doi.org/10.3389/fpubh.2022.869793

Alanazi, A. T. (2023). Clinicians' Perspectives on Healthcare Cybersecurity and Cyber Threats. Cureus. https://doi.org/10.7759/cureus.47026

Alaran, M., Lawal, S. K., Jiya, M. H., Egya, S. A., Ahmed, M. M., Abdulsalam, A., Haruna, U. A., Musa, M. K., & Lucero- Prisno, D. E. (2025). Challenges and opportunities of artificial intelligence in the African health space. Digital Health, 11. https://doi.org/10.1177/20552076241305915

Alarifi, A., & Alwadain, A. (2021). Relative traffic management scheme for robot-assisted healthcare applications. Swarm and Evolutionary Computation, 64, 100887. https://doi.org/10.1016/j.swevo.2021.100887

Aldosari, B. (2025). Cybersecurity in Healthcare: New Threat to Patient Safety [Review of Cybersecurity in Healthcare: New Threat to Patient Safety]. Cureus. Cureus, Inc. https://doi.org/10.7759/cureus.83614

Alexander, C. A., & Wang, L. (2018). Big Data and Data-Driven Healthcare Systems. Journal of Business and Management Sciences, 6(3), 104. https://doi.org/10.12691/jbms-6-3-7

Al- Khasawneh, M. A., Faheem, M., Alarood, A. A., Habibullah, S., & Alzahrani, A. (2024). A secure blockchain framework for healthcare records management systems. Healthcare Technology Letters, 11(6), 461. https://doi.org/10.1049/htl2.12092

Almaghrabi, N. S., & Bugis, B. A. (2022). Patient Confidentiality of Electronic Health Records: A Recent Review of the Saudi Literature [Review of Patient Confidentiality of Electronic Health Records: A Recent Review of the Saudi Literature]. Dr Sulaiman Al Habib Medical Journal, 4(3), 126. Springer Nature. https://doi.org/10.1007/s44229-022-00016-9

Al-Qarni, E. A. (2023). Cybersecurity in Healthcare: A Review of Recent Attacks and Mitigation Strategies [Review of Cybersecurity in Healthcare: A Review of Recent Attacks and Mitigation Strategies]. International Journal of Advanced Computer Science and Applications, 14(5). Science and Information Organization. https://doi.org/10.14569/ijacsa.2023.0140513

Amanat, A., Rizwan, M., Maple, C., Zikria, Y. B., Almadhor, A., & Kim, S. W. (2022). Blockchain and cloud computing-based secure electronic healthcare records storage and sharing. Frontiers in Public Health, 10. https://doi.org/10.3389/fpubh.2022.938707

Argaw, S., Troncoso-Pastoriza, J. R., Lacey, D., Florin, M., Calcavecchia, F., Anderson, D., Burleson, W., Vogel, J.-M., O'Leary, C., Eshaya-Chauvin, B., & Flahault, A. (2020). Cybersecurity of Hospitals: discussing the challenges and working towards mitigating the risks [Review of Cybersecurity of Hospitals: discussing the challenges and working towards mitigating the risks]. BMC Medical Informatics and Decision Making, 20(1). BioMed Central. https://doi.org/10.1186/s12911-020-01161-7

Asha, A., Arafat, Md. E., Desai, K., Hossain, M. A., & Akter, S. (2025). The Role of Blockchain and AI in Revolutionizing Electronic Health Records: A Business-Driven Approach to Data Security and Interoperability. International Interdisciplinary Business Economics Advancement Journal, 6(5), 8. https://doi.org/10.55640/business/volume06issue05-02

Baiyewu, A. S. (2023). Overview of the Role of Data Analytics in Advancing Health Services. OALib, 10(6), 1. https://doi.org/10.4236/oalib.1110207

Barka, E., Dahmane, S., Kerrache, C. A., Khayat, M., & Sallabi, F. (2021). STHM: A Secured and Trusted Healthcare Monitoring Architecture Using SDN and Blockchain. Electronics, 10(15), 1787. https://doi.org/10.3390/electronics10151787

Bazel, M. A., Mohammed, F., Ahmad, M., Baarimah, A. O., & Maskari, T. A. (2025). Blockchain technology adoption in healthcare: an integrated model. Scientific Reports, 15(1). https://doi.org/10.1038/s41598-025-95253-x

Belle, A., Raghuram, T., Soroushmehr, S. M. R., Navidi, F., Beard, D., & Najarian, K. (2015). Big Data Analytics in Healthcare [Review of Big Data Analytics in Healthcare]. BioMed Research International, 2015, 1. Hindawi Publishing Corporation. https://doi.org/10.1155/2015/370194

Carello, M. P., Marchetti-Spaccamela, A., Querzoni, L., & Angelini, M. (2023). A Systematization of Cybersecurity Regulations, Standards, and Guidelines for the Healthcare Sector. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2304.14955

Chang, S. E., & Chen, Y. (2020). Blockchain in Health Care Innovation: Literature Review and Case Study From a Business Ecosystem Perspective [Review of Blockchain in Health Care Innovation: Literature Review and Case Study From a Business Ecosystem Perspective]. Journal of Medical Internet Research, 22(8). JMIR Publications. https://doi.org/10.2196/19480

Chen, Y., Ding, S., Xu, Z., Zheng, H., & Yang, S. (2018). Blockchain-Based Medical Records Secure Storage and Medical Service Framework. Journal of Medical Systems, 43(1). https://doi.org/10.1007/s10916-018-1121-4

Chitikela, A. N. (2024). Secure and Transparent Medical Record Management System Using Python and Blockchain. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2408.02081

Craven, M., & Page, C. D. (2015). Big Data in Healthcare: Opportunities and Challenges. Big Data, 3(4), 209. https://doi.org/10.1089/big.2015.29001.mcr

Dash, S., Shakyawar, S. K., Sharma, L., & Kaushik, S. (2019). Big data in healthcare: management, analysis and prospects. Journal Of Big Data, 6(1). https://doi.org/10.1186/s40537-019-0217-0

Duong- Trung, N., Son, H. X., Le, H. T., & Phan, T. T. (2020). Smart Care. 105. https://doi.org/10.1145/3377644.3377667

Eldin, A. M., Hossny, E., Wassif, K., & Omara, F. A. (2023). Federated blockchain system (FBS) for the healthcare industry [Review of Federated blockchain system (FBS) for the healthcare industry]. Scientific Reports, 13(1). Nature Portfolio. https://doi.org/10.1038/s41598-023-29813-4

Elendu, C., Omeludike, E. K., Oloyede, P. O., Obidigbo, B. T., & Omeludike, J. C. (2024). Legal implications for clinicians in cybersecurity incidents: A review [Review of Legal implications for clinicians in cybersecurity incidents: A review]. Medicine, 103(39). Wolters Kluwer. https://doi.org/10.1097/md.0000000000039887

Elvas, L. B., Serrão, C., & Ferreira, J. C. (2023). Sharing Health Information Using a Blockchain. Healthcare, 11(2), 170. https://doi.org/10.3390/healthcare11020170

Esmaeilzadeh, P. (2022). Benefits and concerns associated with blockchain-based health information exchange (HIE): a qualitative study from physicians' perspectives. BMC

Medical Informatics and Decision Making, 22(1). https://doi.org/10.1186/s12911-022-01815-8

Ferreira, J. C., Elvas, L. B., Correia, R., & Mascarenhas, M. (2024). Enhancing EHR Interoperability and Security through Distributed Ledger Technology: A Review [Review of Enhancing EHR Interoperability and Security through Distributed Ledger Technology: A Review]. Healthcare, 12(19), 1967. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/healthcare12191967

Gadekallu, T. R., Manoj, M., S.S.K., K., N., H., S., & Bhattacharya, S. (2021). Blockchain-Based Attack Detection on Machine Learning Algorithms for IoT-Based e-Health Applications. IEEE Internet of Things Magazine, 4(3), 30. https://doi.org/10.1109/iotm.1021.2000160

George, J., & Emmanuel, A. (2018). Cyber Hygiene in Health Care Data Breaches. International Journal of Privacy and Health Information Management, 6(1), 37. https://doi.org/10.4018/ijphim.2018010103

Gordon, W. J., & Catalini, C. (2018). Blockchain Technology for Healthcare: Facilitating the Transition to Patient-Driven Interoperability [Review of Blockchain Technology for Healthcare: Facilitating the Transition to Patient-Driven Interoperability]. Computational and Structural Biotechnology Journal, 16, 224. Elsevier BV. https://doi.org/10.1016/j.csbj.2018.06.003

Han, Y., Zhang, Y., & Vermund, S. H. (2022). Blockchain Technology for Electronic Health Records. International Journal of Environmental Research and Public Health, 19(23), 15577. https://doi.org/10.3390/ijerph192315577

Hasselgren, A., Wan, P. K., Horn, M., Kralevska, K., Gligoroski, D., & Faxvaag, A. (2020). GDPR Compliance for Blockchain Applications in Healthcare. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2009.12913

Hiwale, M., Walambe, R., Potdar, V., & Kotecha, K. (2023). A systematic review of privacy-preserving methods deployed with blockchain and federated learning for the telemedicine [Review of A systematic review of privacy-preserving methods deployed with Blockchain and federated learning for the telemedicine]. Healthcare Analytics, 3, 100192. Elsevier BV. https://doi.org/10.1016/j.health.2023.100192

Hong, L., Luo, M., Wang, R., Lu, P., Lu, W., & Lu, L. (2018). Big Data in Health Care: Applications and Challenges. Data and Information Management, 2(3), 175. https://doi.org/10.2478/dim-2018-0014

Howe, E. G., & Elenberg, F. (2021). Ethical Challenges Posed by Big Data. Innovations in Clinical Neuroscience, 17, 24. https://europepmc.org/article/PMC/PMC7819582

Hu, F., Qiu, S., Yang, X., Wu, C., Nunes, M. B., & Chen, H. (2024). Privacy-Preserving Healthcare and Medical Data Collaboration Service System Based on Blockchain and Federated Learning. Computers, Materials & Continua/Computers, Materials & Continua (Print), 80(2), 2897. https://doi.org/10.32604/cmc.2024.052570

Hylock, R., & Zeng, X. (2019). A Blockchain Framework for Patient-Centered Health Records and Exchange (HealthChain): Evaluation and Proof-of-Concept Study [Review of A Blockchain Framework for Patient-Centered Health Records and Exchange (HealthChain): Evaluation and Proof-of-Concept Study]. Journal of Medical Internet Research, 21(8). JMIR Publications. https://doi.org/10.2196/13592

Ibrahim, S. A., Charlson, M. E., & Neill, D. B. (2020). Big Data Analytics and the Struggle for Equity in Health Care: The Promise and Perils. Health Equity, 4(1), 99. https://doi.org/10.1089/heq.2019.0112

Jabarulla, M. Y., & Lee, H.-N. (2020). Blockchain-Based Distributed Patient-Centric Image Management System. Applied Sciences, 11(1), 196. https://doi.org/10.3390/app11010196

Jamil, F., Ahmad, S., Iqbal, N., & Kim, D. (2020). Towards a Remote Monitoring of Patient Vital Signs Based on IoT-Based Blockchain Integrity Management Platforms in Smart Hospitals. Sensors, 20(8), 2195. https://doi.org/10.3390/s20082195

Kalejahi, B. K., Meshgini, S., Yariyeva, A., Ndure, D., Maharramov, U., & Farzamnia, A. (2019). Big Data Security Issues and Challenges in Healthcare. arXiv (Cornell University). https://doi.org/10.48550/arxiv.1912.03848

Kantarcıoğlu, M., & Ferrari, E. (2019). Research Challenges at the Intersection of Big Data, Security, and Privacy. Frontiers in Big Data, 2. https://doi.org/10.3389/fdata.2019.00001

Kasyapa, M. S. B., & Vanmathi, C. (2024). Blockchain integration in healthcare: a comprehensive investigation of use cases, performance issues, and mitigation strategies. Frontiers in Digital Health, 6. https://doi.org/10.3389/fdgth.2024.1359858

Khan, S., Khan, H. U., & Nazir, S. (2022). Systematic analysis of healthcare big data analytics for efficient care and disease diagnosis. Scientific Reports, 12(1). https://doi.org/10.1038/s41598-022-26090-5

Kiania, K., Jameii, S. M., & Rahmani, A. M. (2023). Blockchain-based privacy and security preserving in electronic health: a systematic review [Review of Blockchain-based privacy and security preserving in electronic health: a systematic review]. Multimedia Tools and Applications, 82(18), 28493. Springer Science+Business Media. https://doi.org/10.1007/s11042-023-14488-w

Kuo, M. H., Sahama, T., Kushniruk, A., Borycki, E. M., & Grunwell, D. (2014). Health big data analytics: current perspectives, challenges, and potential solutions. International Journal of Big Data Intelligence, 1, 114. https://doi.org/10.1504/ijbdi.2014.063835

Laurie, G. (2019). Cross-Sectoral Big Data. Asian Bioethics Review, 11(3), 327. https://doi.org/10.1007/s41649-019-00093-3

Lee, C. H., & Yoon, H. (2017). Medical big data: promise and challenges. Kidney Research and Clinical Practice, 36(1), 3. https://doi.org/10.23876/j.krcp.2017.36.1.3

Lewis, N., Connelly, Y., Henkin, G., Leibovich, M., & Akavia, A. (2022). Factors Influencing the Adoption of Advanced Cryptographic Techniques for Data Protection of Patient Medical Records. Healthcare Informatics Research, 28(2), 132. https://doi.org/10.4258/hir.2022.28.2.132

Liu, H., Crespo, R. G., & Martínez, Ó. S. (2020). Enhancing Privacy and Data Security across Healthcare Applications Using Blockchain and Distributed Ledger Concepts. Healthcare, 8(3), 243. https://doi.org/10.3390/healthcare8030243

Makka, S., Arora, G., & Mopuru, B. (2021). IoT-based health monitoring and record management using a distributed ledger. Journal of Physics Conference Series, 2089(1), 12030. https://doi.org/10.1088/1742-6596/2089/1/012030

Mandarino, V., Pappalardo, G., & Tramontana, E. (2024). A Blockchain-Based Electronic Health Record (EHR) System for Edge Computing Enhancing Security and Cost Efficiency. Computers, 13(6), 132. https://doi.org/10.3390/computers13060132

Mole, J., & Shaji, R. S. (2024). Ethereum Blockchain for electronic health records: securing and streamlining patient management. Frontiers in Medicine, 11. https://doi.org/10.3389/fmed.2024.1434474

Nash, A. (2024). Decentralized Health Intelligence Network (DHIN). arXiv (Cornell University). https://doi.org/10.48550/arxiv.2408.06240

Odeh, A., Abdelfattah, E., & Salameh, W. A. (2024). Privacy-Preserving Data Sharing in Telehealth Services. Applied Sciences, 14(23), 10808. https://doi.org/10.3390/app142310808

Oluomachi, E., & Ahmed, A. (2024). Securing the Future of Healthcare: Building a Resilient Defense System for Patient Data Protection. Computer Science and Information Technology, 27. https://doi.org/10.5121/csit.2024.141303

Omotosho, A., Emuoyibofarhe, J., & Meinel, C. (2017). Ensuring patients' privacy in a cryptographic-based electronic health records system using biometrics. International Journal of Electronic Healthcare, 9(4), 227. https://doi.org/10.1504/ijeh.2017.085800

Pedada, N. K. (2025). Blockchain technology: Revolutionizing healthcare data security and real-time information exchange. World Journal of Advanced Research and Reviews, 26(1), 2230. https://doi.org/10.30574/wjarr.2025.26.1.1245

Perez, A. O., & Palaoag, T. D. (2021). Blockchain-based Model for Health Information Exchange: A Case for Simulated Patient Referrals Using an Electronic Medical Record. IOP Conference Series Materials Science and Engineering, 1077(1), 12059. https://doi.org/10.1088/1757-899x/1077/1/012059

Pham, Q., Nguyen, D. C., Huynh- The, T., Hwang, W., & Pathirana, P. N. (2020). Artificial Intelligence (AI) and Big Data for Coronavirus (COVID-19) Pandemic: A Survey on the State-of-the-Art. IEEE Access, 8, 130820. https://doi.org/10.1109/access.2020.3009328

Price, W. N., & Cohen, I. G. (2018). Privacy in the age of medical big data [Review of Privacy in the age of medical big data]. Nature Medicine, 25(1), 37. Nature Portfolio. https://doi.org/10.1038/s41591-018-0272-7

Quazi, F., Raju, N., Gorrepati, N., & Kareem, S. A. (2024). Blockchain Applications in Electronic Health Records (EHRs). https://doi.org/10.21428/e90189c8.5043b7de

Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential [Review of Big data analytics in healthcare: promise and potential]. Health Information Science and Systems, 2(1). Springer Nature. https://doi.org/10.1186/2047-2501-2-3

Richard, T. (2024). Blockchain in Healthcare: Ensuring Data Security and Integrity. Research Output Journal of Public Health and Medicine, 4(2), 12. https://doi.org/10.59298/rojphm/2024/421217

Simonoski, O., & Bogatinoska, D. C. (2024). Block Medcare: Advancing Healthcare Through Blockchain Integration. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4978995

Sonkamble, R. G., Phansalkar, S., Potdar, V., & Bongale, A. M. (2021). Survey of Interoperability in Electronic Health Records Management and Proposed Blockchain-

Based Framework: MyBlockEHR. IEEE Access, 9, 158367. https://doi.org/10.1109/access.2021.3129284

Sun, Z., Han, D., Li, D., Wang, X., Chang, C., & Wu, Z. (2022). A blockchain-based secure storage scheme for medical information. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2207.06102

Tahir, N. U. A., Rashid, U., Hadi, H. J., Ahmad, N., Cao, Y., Alshara, M. A., & Javed, Y. (2024). Blockchain-Based Healthcare Records Management Framework: Enhancing Security, Privacy, and Interoperability. Technologies, 12(9), 168. https://doi.org/10.3390/technologies12090168

Tcholakian, M., Gorna, K., Laurent, M., Ayed, H. K. B., & Naghmouchi, M. (2024). Self-Sovereign Identity for Consented and Content-Based Access to Medical Records using Blockchain. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2407.21559

Thapa, C., & Camtepe, S. (2020). Precision health data: Requirements, challenges and existing techniques for data security and privacy [Review of Precision health data: Requirements, challenges and existing techniques for data security and privacy]. Computers in Biology and Medicine, 129, 104130. Elsevier BV. https://doi.org/10.1016/j.compbiomed.2020.104130

Torongo, A. A., & Toorani, M. (2023). Blockchain-based Decentralized Identity Management for Healthcare Systems. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2307.16239

Ullah, H. S., Aslam, S., & Arjomand, N. (2020). Blockchain in Healthcare and Medicine: A Contemporary Research of Applications, Challenges, and Future Perspectives. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2004.06795

Vayena, E., Dzenowagis, J., Brownstein, J. S., & Sheikh, A. (2017). Policy implications of big data in the health sector. Bulletin of the World Health Organization, 96(1), 66. https://doi.org/10.2471/blt.17.197426

Wang, C. (2019). The Strengths, Weaknesses, Opportunities, and Threats Analysis of Big Data Analytics in Healthcare. International Journal of Big Data and Analytics in Healthcare, 4(1), 1. https://doi.org/10.4018/ijbdah.2019010101

Wang, S., Yang, M., Ge, T., Luo, Y., & Fu, X. (2021). BOSS: A Blockchain Off-State Sharing System. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2109.07626

Wasserman, L., & Wasserman, Y. (2022). Hospital cybersecurity risks and gaps: Review (for the non-cyber professional) [Review of Hospital cybersecurity risks and gaps: Review (for the non-cyber professional)]. Frontiers in Digital Health, 4. Frontiers Media. https://doi.org/10.3389/fdgth.2022.862221

Wenhua, Z., Qamar, F., Abdali, T.-A. N., Hassan, R., Jafri, S. T. A., & Nguyễn, Q. N. (2023). Blockchain Technology: Security Issues, Healthcare Applications, Challenges, and Future Trends. Electronics, 12(3), 546. https://doi.org/10.3390/electronics12030546

Zhang, P., Schmidt, D. C., White, J., & Lenz, G. (2018). Blockchain Technology Use Cases in Healthcare. In Advances in Computers (p. 1). Elsevier BV. https://doi.org/10.1016/bs.adcom.2018.03.006

Zhang, P., White, J., Schmidt, D. C., Lenz, G., & Rosenbloom, S. T. (2018). FHIRChain: Applying Blockchain to Securely and Scalably Share Clinical Data. Computational

and Structural Biotechnology Journal, 16, 267. https://doi.org/10.1016/j.csbj.2018.07.004

Zhang, R., Xue, R., & Liu, L. (2021a). Security and Privacy for Healthcare Blockchains. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2106.06136

Zhang, R., Xue, R., & Liu, L. (2021b). Security and Privacy for Healthcare Blockchains. IEEE Transactions on Services Computing, 15(6), 3668. https://doi.org/10.1109/tsc.2021.3085913

Zhang, X. (2020). Healthcare Regulation and Governance: Big Data Analytics and Healthcare Data Protection. https://jdc.jefferson.edu/cgi/viewcontent.cgi?article=1003&context=jscpssp