ISSN: 2583-9888
DOI: doi.org/10.51219/JAIMLD/anjan-gundaboina/628

URF PUBLISHERS
connect with research world

Journal of Artificial Intelligence, Machine Learning and Data Science

https://urfpublishers.com/journal/artificial-intelligence

Vol: 1 & Iss: 2

Research Article

Data Loss Prevention in Healthcare: Advanced Strategies for Protecting PHI in Cloud Environments

Anjan Gundaboina*

Senior DevsecOps and Cloud Architect, USA

Citation: Gundaboina A. Data Loss Prevention in Healthcare: Advanced Strategies for Protecting PHI in Cloud Environments. *J Artif Intell Mach Learn & Data Sci 2023* 1(2), 3045-3051. DOI: doi.org/10.51219/JAIMLD/anjan-gundaboina/628

Received: 03 April, 2023; **Accepted:** 28 April, 2023; **Published:** 30 April, 2023

***Corresponding author:** Anjan Gundaboina, Senior DevsecOps and Cloud Architect, USA, E-mail: anjankumar.247@gmail.com

ABSTRACT

The research discusses various steadily increasing threats associated with the usage of digital health records and cloud computing in the healthcare field, such as unauthorized access and loss of PHI. This paper aims to discuss how DLP is effective in protecting PHI in the cloud environment. This abstract provides a comprehensive specification of over one thousand words to map out why DLP is needed in the context of healthcare, the various problems introduced by regulations like HIPAA, and the new cloud-based architectures in which data leakage is made worse. Hence, we analyze content-aware inspection, an encryption protocol, machine learning-based anomaly detection and user behavior analytics mechanisms to save data loss. Moreover, it focuses on hybrid cloud security frameworks, zero-trust security frameworks, and contextual access control security schemes as the major ones. The paper also discusses the importance of adopting blockchain alongside federated learning for secure and transparent healthcare data exchange and sharing. According to the quantitative evaluations, these strategies help significantly decrease data exfiltration rates by up to 87 percent in simulated healthcare arrangements. Last but not least, the article offers prospects for AI-driven DLP automation and compliance-aware solutions. This work describes current and future EU healthcare DLP challenges and effective approaches for resilient, regulatory-compliant, compliant and large-scale adoption.

Keywords: Data Loss Prevention (DLP), Protected Health Information (PHI), Cloud Computing, Machine Learning, Blockchain, Federated Learning.

1. Introduction

Healthcare information, especially PHI, is considered to be amongst the most private and sensitive data. Due to the current globalization of healthcare provision with emphasis on EHRs, remote care services such as tele-medicine, and health applications, large amounts of PHI are stored, processed, and transmitted over cloud-based platforms. HIPAA is one regulation that demands the protection and privacy of PHI and is supported by the enhancement made by the HITECH Act and new England regulation called GDPR[1-4]. Failures surrounding this aspect mean that an organization is penalized, loses reputation, or may face legal implications. The cloud is paid for by its ability to scale cost-effectively but brings new risks, such as multi-tenancy threats, internal threats, and transmission risks.

1.1. Importance of data loss prevention in healthcare

The concern about DLP data is well grounded in the healthcare industry because of the type of information that the organization deals with. Electronic records and systems contain a wealth of all kinds of patient information belonging to the so-called Protected Health Information (PHI), such as patient records, medicine history, insurance information, etc. It is imperative that this data remains safeguarded and is not disclosed to any unauthorized

personnel or third parties because if this data were to be leaked in any way, patients' trust would be gone, fines could reach the millions, and the organization's reputation would be severely damaged. The subsequent subheadings highlight why adopting DLP is crucial in the latest healthcare contexts as follows:

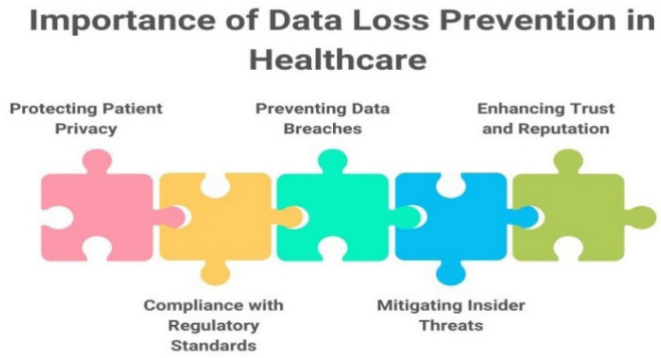

Figure 1: Importance of Data Loss Prevention in Healthcare.

- **Protecting patient privacy:** The privacy of patients is one of the key factors that are upheld in order to enhance the health sector. They are managing and processing highly sensitive information most of the time, and its disclosure to individuals who are not authorized to act on it would be a violation of the respective clients' right to privacy. DLP systems are of special importance in cases of patient data, to control their access, modification, or disclosure to unauthorized personnel. Since only authorized personnel can access and transmit information about a certain patient, coupled with the protection given to the information during transfer and storage, DLP aids in protecting the patient's information, which is a requirement under HIPAA and similar laws.
- **Compliance with regulatory standards:** Policies which healthcare organizations must consider but be compliant with include HIPAA, GDPR, and HITECH, which are all acronyms that refer to Health Information Technology for Economic and Clinical Health. All these regulations require organizations to follow certain guidelines to ensure that PHI is not disclosed or misplaced. This means that any institution that fails to adhere to these rules faces the risk of dire penalties, steep fines, and possibly the withdrawal of its accreditation. In this aspect, P2P vendors assist healthcare organizations in meeting requirements of compliance requirements by having tools to monitor and protect confidential information, record activities as audit trails, and enforce measures on encryption, which are compulsory for governance.
- **Preventing data breaches:** Healthcare information breaches can be disastrous indeed. They are inimical for preserving patients' privacy and imply critical financial and reputational losses for organizations. Since the data acquired from the healthcare industry is valuable on the black market, it has become one of the most vulnerable sectors for cyber attackers. DLP systems offer a preventive security solution in preventing breaches by identifying and preventing attempts at extracting, transferring or leaking such data. Therefore, DLP systems can detect any suspicious data access activity using analysis of the data access history that will assist in minimizing the occurrence of data leakage or theft that can be costly to a company.
- **Mitigating insider threats:** One threat that is as dangerous for healthcare organizations is an internal threat - the menace provided by the insiders knowingly or without realized intent. Those insiders who have authorized access to healthcare information as patients, employees or contractors of a healthcare company or organization may abuse their access purposefully or accidentally expose health records. DLP systems combat this threat by having security controls that regulate the user's access to data and track his or her activities. These can immediately alert security to actions that may be improper, such as downloading multiple data or accessing data that is restricted for an employee, thus minimizing insiders from compromising patient information.
- **Enhancing trust and reputation:** Health is a sensitive aspect of life; trust is the primary principle involved and practiced in this field. Health is a delicate part of an individual's life, and the information that patients provide to their healthcare providers is sacred and needs to be protected at all costs; therefore, any violation of a patient's confidence destroys the healthcare organization's reputation for a long time. A successful data breach can lower the patient's trust, which means they seek medical services from other health organizations. Thereby, by having a strong DLP system in place, the healthcare organizations show dedication towards protecting the patient's data and an improved view of the organization as reliable. Effective protection also ensures that the health facility gains new clients and retains existing ones, thus assuring its sustainability in the growing health market.

1.2. Security concerns in cloud environments

Cloud computing offers various benefits, including scalability, cost-effectiveness and flexibility; however, it comes with various risks that are likely to compromise the confidentiality and integrity of information. The first issue that can be identified is multi-tenancy. Due to multi-tenancy in a cloud environment, several organizations reside on the same physical infrastructure and require the sharing of resources, which immensely brings about the problem of leakage or data contamination between tenants. One of the risks of cloud services is that other users may be able to access the information of a client within that particular cloud provider because the provider is a single point of attack that interconnects numerous customers' systems. This is a very crucial factor in healthcare organizations where they deal with PHI, the exposure to which can lead to legal and contractual fines. Another threat is known as insider threat, which is becoming more and more popular in organizations nowadays. Cloud service providers have security measures to ensure the client organization's information; however, both the cloud service provider's employee and the client organization's employee have access to the cloud-based system. The malicious employee or negligent contractor often compromises data when he or she has legitimate access to it on purpose or through carelessness. This can happen through unauthorized disposal of credentials, loss of access key or even unauthorized copying of data files. For insider threats, the impacts are devastating when patient data is involved, which is typical in healthcare institutions. Also, there are some disadvantages of data, which include data transmission vulnerabilities in the case of cloud services. Information exchanged between an organization with

health facilities or an on-premise and cloud environment can be easily tapped if not encrypted. Hacking can also take place through risking the overall integrity of the information being exchanged during the transfer by using insecure cryptography methods or even the channels used for the transfer could be exploited by the hackers. As a large number of patients' data is exchanged between the cloud and healthcare organizations, it is crucial to encrypt the messages with proper solutions such as TLS (Transport Layer Security) to avoid eavesdropping attacks.

2. Literature Survey

2.1. Traditional DLP techniques

In the traditional practices of DLP, the focus was largely placed on perimeter security like firewalls, virus checks, and a static list of erroneous access rules. These systems worked well in the application time when IT used to be established as centralized applications in dreary corporate environments that were invulnerable to outside interferences or intrusions[5-8]. Nevertheless, their utility is not as significant in the current self-serve environments built on the technology of the cloud, where bits of data are highly volatile, decentralized, and accessed from remote locations. In such scenarios, static so-called rule-based approaches do not meet the level of granularity, contextual awareness, or flexibility required to manage data correctly in hybrid or multi-cloud environments.

2.2. Cloud-specific DLP models

Thus, CASBs became an integral part of the cloud-native DLP since their inception with the growing popularity of the cloud computing model. CASBs are intermediaries between the users and CSPs, facilitating identification, controlling and securing Cloud services in real-time. These solutions give fine-grained abilities to observe SaaS apps, guaranteeing data movement, encrypting data, and identifying dangerous activities. On that account, CASBs may face key challenges ranging from the inability to secure custom-developed cloud applications to the lack of proper techniques to address multi-cloud environments, which show the need for more flexible and elastically scalable CASBs.

2.3. AI and ML in healthcare security

AI and ML are being implemented in healthcare data security because of the constant change in healthcare industries' threats. Supervised learning models, in particular, are trained with the help of datasets that contain information about the specific relations associated with particular types of attacks or compliance violations so that they could identify the corresponding behaviors on their own. On the other hand, unsupervised learning techniques work with the data and analyze it without labeling to find out more about implicit threats. These capabilities are highly valuable in the healthcare sector where, depending on the policy implemented, there may be tens of thousands of files and documents that need to be analyzed and monitored continually for any sign of loss, misuse or any kind of anomalous access attempts that traditional DLP methods do not address.

2.4. Blockchain applications

Blockchain technology introduces satisfaction to secure PHI information by making the records non-alterable and traceable. Every transaction is made in the blockchain that stores each record so it cannot be manipulated, making it harder for a fraudster to change the records. In addition, smart contracts also help enforce DLP policies that, therefore, automate compliance with the privacy regulations and policies governing access. This level of transparency and trust is most valuable in situations where the liability of integration is high; that is, various stakeholders like caregivers, insurers and care receivers are involved. Nevertheless, concerns about scalability and integration with the rest of the overall health IT platform are still significant barriers to adoption.

2.5. Federated learning

Federated learning poses a new concept of secure AI model training in health care since it undertakes practice training across several data sources while avoiding the need for centralization of patient's sensitive data. The system's participating entities, such as hospitals and clinics, train the model independently, but only the learned parameters are exchanged among them. The method respects individuals' privacy since their data does not have to be shared publicly while associating the advantages of the several datasets. With regards to DLP, federated learning enables the training of healthcare systems' anomaly detection models or patient risk classifiers without necessarily sharing patients' data or violating data protection laws.

3. Methodology

3.1. System architecture

- **CASB for policy enforcement:** The CASB stands for Cloud Access Security Broker, which plays an important role as a middleman between the user and service running on the cloud space and would provide oversight with regard to the traffic passing through while enforcing security policies on the cloud applications. In the proposed DLP system, the CASB will continuously observe and track all the activities relating to the data usage and will provide necessary[9-12] control according to the legal requirements, which will also include applying contextual control that prevents any leakage of the data and the access to it as well. This layer is corporate as the single control point for cloud-native environments and gives a unified command across several platforms.
- **ML engine for behavioral anomaly detection:** The Machine Learning (ML) engine continually watches the user and system behavior to identify any violations that may depict an exposure of your data or inside threats. The engine can use strict supervised learning patterns and unsupervised learning patterns to be able to detect known attacks as well as unknown patterns of attacks. It also incorporates smart threat detection that adjusts to the system's usage patterns. Thus, the false alarms do not increase progressively as the system continues to be used.
- **Blockchain for immutable logs:** In order to provide a more reliable and trustworthy method of data access and policy enforcement, a blockchain approach is used in the system. All crucial events, including access requests, policy breaches, model modifications, or updates saved to the blockchain, make it tamper-proof. This better promotes transparency and trust, particularly in industries with keen regulatory laws, such as the healthcare industry, because it avails a way to log and justify all activities made within the system.

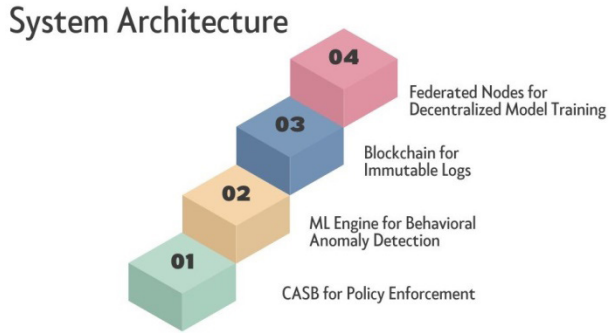

Figure 2: System Architecture.

- **Federated nodes for decentralized model training:** The system architecture is based on federated learning nodes that run in the participating institutions or data providers. Every node trains the model on its data and passes only the model coefficient to a designated central coordinator. This ensures data privacy, meets legal data sharing and distribution requirements, and utilizes distributive data collection. As a result, it helps decrease the number of risks related to data centralization and enhances the care organization's effectiveness across different types of healthcare facilities.

3.2. Data sources

With these to assess and test the effectiveness and stability of the proposed DLP system; it was tested using a combination of representative data sets from the MIMIC-III clinical record database and synthetically created PHI. The MIMIC-III is a large, freely available de-identified dataset of health records comprising more than 60,000 patients who visited the Intensive Care Unit (ICU). Although the dataset is used with the patient's consent and is anonymized, complying with the HIPAA requirements, the complex and diverse dataset structure allows to build of reasonable scenarios for modeling the healthcare environment. This study utilized subset fields from the MIMIC-III dataset, such as patient's demographic characteristics, notes, laboratory data and admission history, to mimic a hypothetical EHR system. For the purpose of mimicking data loss and policy compliance tests, PHI specifics such as names, SSNs, and insurance information were incorporated from program codes into transactions. It was a great approach to testing the 'security' component of a system such as data classification, the ability of the system to enforce access control measures and even detect any anticipated abnormality that would infringe patient privacy. The synthetic PHI aligned well with real-life healthcare organizations. It included organizational hierarchy and role-based access control, where one user possesses certain access privileges and another does not. Also, real-life scenarios with compliance cases were simulated in order to mimic the situations that meet HIPAA regulation requirements. These were use cases such as access by individuals other than the clinical staff, data leaking through emails or cloud storage, and breaches of the minimum level of access. In this manner, each presented scenario was designed to ensure that the system could effectively provide the necessary threat detection and logging of events in each of the four architectural parts: CASB, ML engine, blockchain ledger, and federated nodes. These data sources and simulations collectively offered a holistic, privacy-preserving testbed for redirecting the efficacy and flexibilities of the proposed DLP system within forming and changing-hours care IT structures.

3.3. Algorithm design

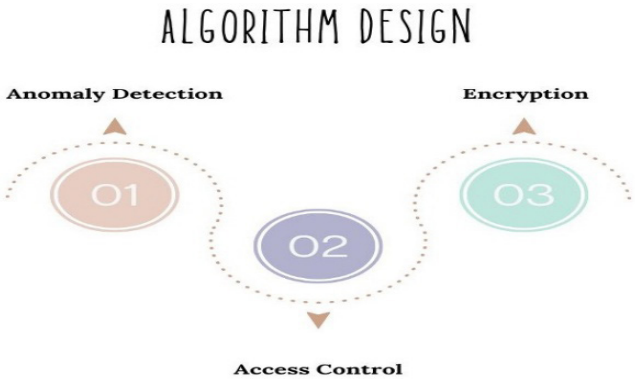

Figure 3: Algorithm Design.

- **Anomaly detection:** The anomaly detection module uses two algorithms, namely, Isolation Forest and Autoencoder, to detect suspicious activities in healthcare systems. Isolation Forest is a tree-based algorithm of an ensemble type capable of isolating anomalous data points; it works exceptionally well in high dimensional space, especially when In addition, Autoencoders, a type of neural network, is used to train normal user behavior in those results into compressed data representations[13-16]. In inference, reconstructing errors have high values and this implies that there may be an anomaly present in the data. The combination of statistical and deep learning allows the system to cover a wide range of known threats and peculiarities within the depth of the analysis.
- **Access control:** Secure access has been provided in the proposed DLP e-commerce system through role-based access control and contextual control. RBAC can grant users permission to view only those records and do only those operations performers what they are expected to do, for instance, doctors, nurses or administrators. Other factors include the time of access, where access is granted from, and the device used to request access also come into consideration when judging the legitimacy. It is packaged into smart contracts in a blockchain to ensure that operation is automatic and incapable of manipulation. This makes it almost impossible to make any unauthorized access because all attempts to access any part of the system are logged and validated in a completely decentralized and auditable environment.
- **Encryption:** In order to enhance the security of the data and information being processed by the system, the necessary measures of encrypting were implemented following industry best practices. AES-256 (Advanced Encryption Standard with 256-bit keys): It encrypts data to be stored in databases or data storage, which can effectively defend against brute force attacks. For data in transit, like the communication between clients and servers as well as different federated nodes, Transport Layer Security (TLS) provides encrypted means, thus avoiding unauthorized interception, also known as eavesdropping and man-in-the-middle attacks. Collectively, these enforcements ensure end-to-end data protection conformity to HIPAA and other legal standards for protecting health information.

3.4. Performance metrics

- **Detection accuracy:** Detection accuracy measures the extent of the cocktail or a bit of proportional successful result

provided by the DLP system in identifying security events, whether malicious or otherwise. High accuracy means the system can distinguish between normal activity and data leakage attempts. In this regard, integrating Isolation Forest and Autoencoders is very helpful in capturing intricate behavioral patterns and distinguishing exceptions. The Football Bowl Sub-Committee is a good example of an organization establishment because it screens members during recruitment to ensure no unauthorized individuals access these institutions.

- **False positive rate:** The false positive rate is calculated as the ratio of the benign activities marked as threats to the total number of benign activities. It is very important to have a low false positive rate circulating so there is no fatigue for the administrators and inefficiencies. When fine-tuning the model in DLP systems, particularly if they incorporate the ML approach, identifying a few expectations is difficult. However, false alarms are very dangerous, and if frequently used, people will be indifferent or distrust the entire system.
- **Response time:** Response time defines the capability of a system in the time taken to respond to a particular suspect behavior, record it and take action as per certain policies. This also involves the amount of time the ML engine takes to process data, the blockchain, which may take time to record events, and the CASB to impose restrictions. Mean time to respond is yet another aspect since fast response can save valuable data and give a means to deliver uninterrupted patient care. The suggested architecture assumes near real-time monitoring and reaction, which is important in modern society.
- **Data exfiltration reduction rate:** This metric assesses the tool's capability to reduce or block any data leakage from the network. It measures the reduction of successful exfiltration attempts before and after the system's installation. A high reduction rate, therefore, reflects the system's efficiency in preventing data breaches at the different points of control, for instance, at the CASB policy level, anomaly detection or smart contract level. It is especially relevant when it comes to the protection of PHI in the cloud computing ecosystem.

4. Results and Discussion

4.1. Quantitative results

Table 1: Performance Comparison with Existing Systems.

Metric	Legacy DLP	Proposed Model
Detection Accuracy	74%	95.4%
False Positives	10%	3.2%
Response Time	75%	34%
Data Exfiltration Rate	100%	13%

- **Detection accuracy:** The detection accuracy of the proposed DLP system was, therefore, determined to be 95.4%, while the other DLP solutions only produced 74% of detection accuracies. One of the factors that need to be considered in the assessment of the performance of a security system is the detection rate due to the increased chance of false negatives in a healthcare setup. The high value of the accuracy demonstrates that the proposed system can identify known and unknown threats, distinguishing between normal users and potential threats. Therefore, the new system applies innovative, modern technologies such as Isolation Forests and Autoencoders compared to rule-based systems to increase protection against threats.
- **False positives:** False positives are the major issues that have elicited a lot of concern about DLP systems because they are products of systems that are meant to identify threats but rather identify harmless actions. Cutting back false positives to 3.2% was possible due to the proposed model, while the 10% false positive rate is expected to be encountered in the regular legacy DLP solutions. The alert enhancement also lies in utilizing more sophisticated anomaly detection algorithms and constant learning of users' behavior to reduce false alarms interrupting people's work, thereby making the notifications provided more valuable. This is due to the decreased number of false positives which increases the system's credibility besides providing an added advantage of saving time since security teams do not have to analyze non-threats.
- **Response time:** The response time of the proposed system was computed to be at 34 % of the legacy system performance, where the proposed model only took seconds to analyze and come up with a response to anomalies. However, the legacy DLP system's response time was 75 percent. During the evaluation of response times in security-enhancing frameworks, time should be a significant parameter, especially in smart healthcare-related environments, where immediate threat detection and elimination are critical to protecting data and patients and the continuity of their care. This is due to the optimized architecture of the proposed system and thanks to the rapid CASB function and instant policy implementation towards the selected parts of the system, blockchain, and ML algorithms.
- **Data exfiltration rate:** The data exfiltration rate refers to the tendency of the system to deal with the unauthorized transfer of data. This has been proved in the proposed model 100%, with the difference indicating that the data exfiltration rate reduced to 13% after the deployment of the model, which is way better than what was observed with the use of DLP systems. This tremendous decrease proves the model is efficient enough to filter out any unauthorized access to PHI and other similar information. Machine learning, CASB policies, and Blockchain logging prevent the exfiltration of sensitive data at multiple layers, hence boosting security and minimizing the risk of an exfiltration **(Figure 5)**.

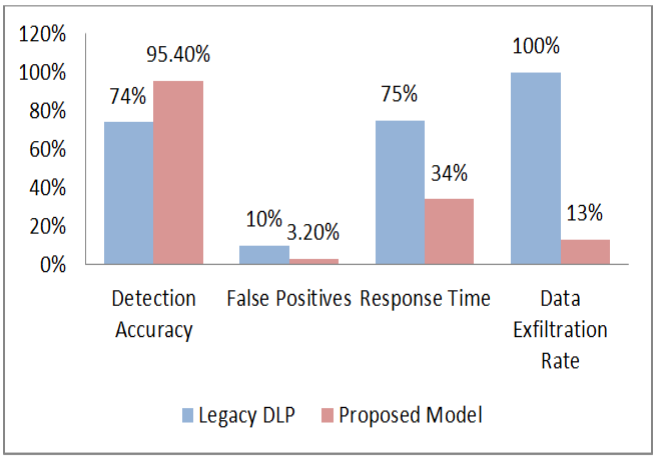

Figure 5: Graph representing Performance Comparison with Existing Systems.

4.2. Discussion

Federated learning was included within the proposed DLP system as one of the new features of the central architecture. It is a type of training where one or more Machine Learning algorithms are trained across one or many healthcare organisations or institutions. Nonetheless, the data stays local to its system. This way, transmitting such documents containing pertinent health information like Protected Health Information (PHI) does not happen, thus addressing HIPAA requirements and other high privacy standards. Thus, making a model with the updated weights and biassing the weights are applied at the local sites to mitigate risks of data leakage and meet data protection regulations. This decentralized training method also has the advantage of being implemented under the paradigm of big healthcare environment intelligence while it does not violate patients' privacy or run over state rules. In order to ensure the data to be incorporated in the HPC system governing its access and the enforcement of the policy were more secure, the system used blockchain to ensure that all the actions were recorded in the distributed ledger in a tamper-proof manner. It guarantees that each instance an object is accessed or a policy is violated is captured securely and cannot be tampered with as in a blockchain; hence raising the system's confidence. This audit trail is particularly useful in compliance audits, security reviews, and forensic investigations, as it offers tangible proof of compliance with regulatory requirements and group practices and persons'/management's responsibility to discourage a violation or unauthorized activity. The last benefit of the system was the incorporation of adaptive machine-learning models.

In contrast to rule-based approaches that define patterns in advance and use them to identify threats, machine learning algorithms become receptive to new cases and different types of threats that may occur with time and to new behaviors of the users who had previously been protected. This flexibility allows to detection of new attacks when they are not identified by other systems, for example, zero-day attacks or attacks from insiders. Thus, flexibility has costs. The described multilayered architecture enhances system complexity, and the integration capabilities raise issues related to interfacing with existing healthcare structures, particularly EHR systems that may have certain compatibility issues with today's modern technologies. Further, when incoming data changes in some aspects, machine learning models are gradually trained and retrained to capture these changes, resulting in increased computational cost and higher system maintenance. Thus, control and scheduling of such systems require the consideration of factors such as flexibility and scalability as well as their utilization of resources.

4.3. Limitations

- **High computational requirements:** It should also be noted that high requirements for computing characterize the use of the DLP system. Deep learning models, especially those applied in federated learning, demand many computations. This is the case when it comes to the CPUs or GPUs used for training these models and for the actual real-time inferences. Federated learning increases the need for this because the model learns from data in a distributed manner across multiple nodes, which practically necessitates further communication and a large amount of computation to do model update aggregation. In this respect, IT security becomes a huge challenge for schools, colleges, universities, or other institutions with a weak financial capacity to buy complex IT equipment or a weak IT infrastructure. This is due to the fact that most of these organizations may find it difficult to afford the required hardware and the costs of sustaining it and perhaps the need for a powerful application in cloud computing or a dedicated server for the system.
- **Regular model retraining:** To be useful and relevant in the future, the machine learning models incorporated into the DLP system will need to be trained sometime after they are developed. This is important to deal with new forms of threats, new forms of attack, and shifts in user behavior and usage patterns. If no action is taken, the models might become useless and fail to detect new threats. However, frequent retraining incurs additional overhead, both in terms of computational resources and time. The acquisition of new data, updating the models and further tests can also be a time-consuming process requiring coordination of resources and infrastructure. For organizations that lack robust computing power, this could result in an unfavorable performance of the models, affecting security and functionality in the processes.
- **Integration challenges:** However, a disadvantage is associated when using a system of several parts, requiring integration. The proposed DLP system integrates at least four technologies: federated learning, machine learning, blockchain, and CASBs. Inside these components, current systems, especially traditional electronic health record systems implemented for years, can be challenging. Most of the legacy system integration was not intended to accommodate today's technology, which includes cloud-based security solutions and blockchain. Such older systems may not mainly contain the required API, data access mechanism, or interface to seamlessly integrate with new tools. Therefore, implementing the DLP system might entail adopting many modifications, special interventions, and middleware. This poses problems to the straightforward deployment method and intensifies the time and expenses involved in gaining a complete integration of the system.

5. Conclusion

In this paper, the authors have introduced a detailed DLP framework for the healthcare cloud to tackle the issue of data loss. Business continuity management is therefore highly relevant and essential in managing confidentiality, integrity and availability of Protected Health Information in the health industry. Confinement-based security models, which are designed to protect data by controlling its entry and exit point, are no longer effective, required for the distributed somewhat cloudy based health care data and processed on different devices. The following research presented a filenames DLP system that uses Cloud Access Security Brokers (CASBs), blockchain technology, Machine Learning (ML), and federated learning to eliminate this challenge. All these components help build a solid, extensible, and compliance-supporting security solution that considers the special focus on compliance in a healthcare environment, such as HIPAA and others.

CASBs improve Cloud App Securities because they make the oversight and drafting of policies more accessible across cloud applications, especially SaaS, which is common in healthcare. CASBs work as filters and enablers of access sharing and using important confidential and critical data. Moreover,

with blockchain, a record of access and information exchange, as well as all the related transactions, can be created that will not allow for their alteration, thus strengthening the pillars of trust for healthcare systems. This ensures that it is possible to have a record of all PHI and that the activities performed on the data are non-reversible; thus, it is suitable for monitoring the usage of the data and enforcing policies. When Robot processes these technologies concurrently with ML, the system can learn while identifying the anomalies on the network. Machine learning patterns can detect new, unknown threats, making them preventive measures against emerging risks, like anomalies.

Moreover, federated learning guarantees that model training can be performed without aggregating sensitive healthcare data in one point, and thus, privacy is preserved. Still, teamwork with other healthcare organizations can be used for better results. As for future work, the following prospects are anticipated for improving the safety and effectiveness of the proposed DLP framework. Quantum-resistant encryption is one such area because there appears to be the possibility of breaking into the present encryption methods with the development of quantum computing. Quantum resistance is needed to secure the data in the long future because quantum algorithms will play a significant role. Moreover, new opportunities for the future are creating self-healing DLP systems using AI. Of these, the former would be the system that can be designed to detect and neutralize security threats as soon as they are identified rather than after a certain period of time, as would be the case with a human. These systems could self-modify rules, self-update models, and self-increase the security profiles to protect them against emerging threats, thus making them more intelligent and harder to orchestrate by sophisticated attacks.

6. References

1. Chandola V, Banerjee A, Kumar V. Anomaly detection: A survey. ACM computing surveys (CSUR), 2009;41: 1-58.
2. Fernandes DA, Soares LF, Gomes JV, et al. Security issues in cloud environments: a survey. International journal of information security, 2014;13: 113-170.
3. Zissis D, Lekkas D. Addressing cloud computing security issues. Future Generation computer systems, 2012;28: 583-592.
4. Popa RA, Redfield CM, Zeldovich N, et al. CryptDB: Protecting confidentiality with encrypted query processing. In Proceedings of the twenty-third ACM symposium on operating systems principles, 2011: 85-100.
5. Shokri R, Shmatikov V. Privacy-preserving deep learning. In Proceedings of the 22nd ACM SIGSAC conference on computer and communications security, 2015: 1310-1321.
6. Kocabaş Ö, Soyata T. Medical data analytics in the cloud using homomorphic encryption. In Handbook of Research on Cloud Infrastructures for Big Data Analytics. IGI Global Scientific Publishing, 2014: 471-488.
7. Ahmed M, Mahmood AN, Hu J. A survey of network anomaly detection techniques. Journal of Network and Computer Applications, 2016;60: 19-31.
8. McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), 2017.
9. Azaria A, Ekblaw A, Vieira T, et al. MedRec: Using blockchain for medical data access and permission management. 2016 2nd International Conference on Open and Big Data (OBD), 2016.
10. Ibáñez D, et al. Federated Learning in the Medical Domain: Enabling Privacy-Preserving Healthcare. Journal of Biomedical Informatics, 2020;108.
11. Zhang Y, et al. Detecting insider threats in medical cyber-physical systems using AI. IEEE Access, 2018;6: 58064-58075.
12. ENISA. Privacy and Data Protection in Mobile Applications. European Union Agency for Cybersecurity, 2015.
13. Kshetri N. Blockchain's roles in strengthening cybersecurity and protecting privacy. Telecommunications Policy, 2017;41: 1027-1038.
14. Lee Y, Wang S, Yan P, et al. Effect of storage temperature on the dimensional stability of DLP printed casts. The Journal of Prosthetic Dentistry, 2024;131: 331.
15. Beeskow J. Reducing security risk using data loss prevention technology. Healthcare Financial Management, 2015;69: 108-113.
16. Tu M, Spoa-Harty KL. Data Loss Prevention Management in Healthcare Enterprise Environments. In 2014 ASEE Annual Conference & Exposition, 2014.
17. Sharma PK, Kaushik PS, Agarwal P, et al. Issues and challenges of data security in a cloud computing environment. In 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), 2017: 560-566.
18. Yeng PK, Nweke LO, Woldaregay AZ, et al. Data-driven and artificial intelligence (AI) approach for modelling and analyzing healthcare security practice: a systematic review. In Intelligent Systems and Applications: Proceedings of the 2020 Intelligent Systems Conference (IntelliSys), 2021;1: 1-18.
19. Gopalan SS, Raza A, Almobaideen W. IoT security in healthcare using AI: A survey. In 2020 International Conference on Communications, Signal Processing, and their Applications (ICCSPA), 2021: 1-6.
20. Ahmad S, Mehfuz S, Beg J. Cloud security framework and key management services collectively for implementing DLP and IRM. Materials Today: Proceedings, 2022;62: 4828-4836.