Invisible Threats: Rethinking Privacy in Digital Healthcare

Parinaz Tabari(✉), Mattia De Rosa, Gennaro Costagliola, and Vittorio Fuccella

Department of Informatics, University of Salerno, Fisciano, Italy
{ptabari,matderosa,gencos,vfuccella}@unisa.it

Abstract. The concept of privacy is essential, as it simultaneously considers other core values such as respect, individuality, dignity, and personal autonomy. Safeguarding data in health research is also crucial since the process generally involves gathering, storing, and utilizing a vast amount of Personally Identifiable Information (PII) or Protected Health Information (PHI), which can be sensitive. This study aims to examine privacy measures in healthcare, providing a comparative analysis of privacy problems and technical solutions. To extract articles, we searched the PubMed database. Manual searches were also conducted in Google Scholar and the reference lists of relevant papers to ensure comprehensiveness. The results show, among others, that the pseudonymization of healthcare data is a widely valued approach as it addresses the challenges posed by other methods. Users and patients can also be actively involved in this process. In some cases, AI models are not only vulnerable to breaches but can also be exploited to amplify privacy attacks. Additionally, while standards like FHIR provide guidance on privacy and security, full compliance with regulations such as HIPAA or GDPR may not always be feasible. Therefore, implementers must carefully consider relevant legal requirements to mitigate privacy risks.

Keywords: Privacy · user concerns · healthcare · telehealth · mHealth · EHR

1 Introduction

Over the past decades, a large number of healthcare applications have been created to enhance medical practices [1]. Digital health systems such as Electronic Health Records (EHRs) significantly improve information and healthcare delivery and can vastly transform patient treatment and disease diagnosis [2]. Clinical Decision Support Systems (CDSSs) are beneficial systems that can be embedded in EHRs to assist clinicians in the decision-making processes. These systems can provide disease-specific and personalized recommendations based on patient data in multiple forms, such as expert systems and diagnostic tools [3,4]. Digital health has introduced several possibilities, such as enhanced information retrieval, telemedicine, and Machine Learning (ML)-based applications.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
O. Gervasi et al. (Eds.): ICCSA 2025 Workshops, LNCS 15893, pp. 267–284, 2026.
https://doi.org/10.1007/978-3-031-97645-2_18

However, due to the nature of e-health, a wide range of individuals and users can access health-related data, raising privacy concerns. This issue becomes even more critical when these applications involve AI, as they require access to larger amounts of data [5]. Privacy and security of patient data are critical issues in healthcare systems. Privacy considerations may be even more critical when dealing with information about drug use or certain special diseases, as information leakage could potentially cause psychological distress to patients [6]. The system designers and developers need to ensure the confidentiality and security of this information to ultimately maintain trust between patients and healthcare providers. To mitigate these concerns, some solutions have been proposed, such as utilizing advanced encryption methods, enforcing stringent access control measures, and trying to comply with applicable legal privacy standards and frameworks, including the Health Insurance Portability and Accountability Act (HIPAA) [4]. The concept of privacy is essential, as it simultaneously considers other core values such as respect, individuality, dignity, and personal autonomy. Safeguarding data in health research is also crucial since the process generally involves gathering, storing, and utilizing a vast amount of Personally Identifiable Information (PII) or Protected Health Information (PHI), which can be sensitive [7]. Security breaches expose individuals whose healthcare information was improperly accessed to various risks and harms. This revelation of personal details may result in inherent harm since private and identifiable information has now been disclosed to others. Apart from these harms, one might be affected economically or psychologically by private information disclosure [7,8]. Therefore, privacy remains a critical concern when managing and evaluating online trust [9]. To address this, it is essential to protect this information by implementing privacy and security measures from the data origin [10]. Accordingly, these considerations will make consumers—*specifically, patients in this context*—more comfortable sharing their health-related information once comprehensive privacy regulations are established [11]. Additionally, these protective measures, such as *strict authorized access*, positively influence public trust in healthcare systems and the willingness to utilize data stored in EHRs and other electronic systems for various purposes, *such as medical research and big data analysis*. These possibilities will ultimately benefit a wide range of patients in the future with more personalized care [12].

1.1 Research Questions and Objectives

This study aims to review privacy measures, concerns, and solutions in healthcare settings, providing a basis for future researchers and implementers of electronic medical systems. Additionally, we provide a comparative analysis of privacy concerns and technical solutions. While previous reviews have examined the privacy and security aspects of healthcare systems [2,13], our study seeks to provide a more comprehensive analysis. We focus on core concepts, technical approaches, risks, and regulations related to privacy, incorporating recent publications to offer an up-to-date perspective. The research questions are as follows:

- **RQ1:** How do privacy, security, confidentiality, other core concepts, and related approaches for their protection emerge in digital health systems?
- **RQ2:** What are the privacy risks and challenges in healthcare systems and platforms?
- **RQ3:** What mechanisms (*from the methodological point of view*) ensure secure data sharing between healthcare providers while preserving patient privacy?
- **RQ4:** Which privacy regulations must be considered when implementing digital health applications?

1.2 Paper Structure

In Sect. 2, we present the survey methodology. Then, in Sect. 3, we give the results about core privacy concepts (Sect. 3.1 -> RQ1), privacy-preserving concepts in healthcare (Sect. 3.2 -> RQ1), privacy risks and challenges in electronic health (Sect. 3.3 -> RQ2), more technical privacy approaches (Sect. 3.4 -> RQ3), and privacy regulations for e-health (Sect. 3.5 -> RQ4). Section 4 provides the conclusion of the research.

2 Methods

To extract articles, we searched PubMed using related keywords, Medical Subject Headings (MeSH) terms, and Boolean operators. The full search strategy is presented in the following text box. *For the PubMed search*, we restricted the results to the papers published within the last 10 years restricting to original studies. Additionally, manual searches in Google Scholar and paper reference lists were conducted to also include extra aspects needed to be declared. We excluded the papers that did not address the privacy considerations in health-related systems or platforms.

((((((((((((((telemedicine[MeSH Terms]) OR (data processing, electronic[MeSH Terms])) OR (electronic health record[MeSH Terms])) OR (clinical decision support systems[MeSH Terms])) OR (clinical decision support systems[MeSH Terms])) OR (health information technologies[MeSH Terms])) OR (telemedicine[Title/Abstract])) OR (telehealth[Title/Abstract])) OR (tele-health[Title/Abstract])) OR ("electronic health"[Title/Abstract])) OR (ehealth[Title/Abstract])) OR (e-health[Title/Abstract])) OR (m-health[Title/Abstract])) OR (mhealth[Title/Abstract])) OR ("mobile health"[Title/Abstract])) AND ((((((privacy[Title/Abstract]) OR ("trust"[Title/Abstract])) OR ("user privacy"[Title/Abstract])) OR (data privacy, patient[MeSH Terms])) OR (patient data privacy[MeSH Terms])) OR (privacy[MeSH Terms]))

This research focused on papers that discuss privacy or security concepts, technical methodologies, or privacy standards and regulations within the healthcare sector. Consequently, studies that did not address these aspects were excluded from this research. In the first phase of article screening, the title and abstract of each paper were assessed, and the papers not conforming to the inclusion criteria were discarded. In the second step, the same approach was performed for the full texts of the remaining articles. *Zotero* web library was used for data collection and organization of papers.

3 Results

Based on the initial search on PubMed, 88 papers were retrieved. Of those, 23 papers were selected for the initial round by reading titles/abstracts and full texts. Subsequently, a manual search on Google Scholar and papers' reference lists was performed, retrieving an additional 36 papers. In total, 59 papers plus nine webpages were analyzed to structure the content of the results section. First, concepts and techniques related to privacy are discussed. Next, privacy challenges and risks in healthcare applications are examined. This is followed by a section focusing on technical methodologies for privacy preservation in electronic health systems. Finally, relevant regulations and standards concerning e-health privacy and security are presented.

3.1 Core Privacy Concepts in Healthcare

With the rapid expansion of health-related data and the subsequent need for data sharing among individuals, patient privacy is a critical aspect that should be protected. Patient attitudes towards data and information privacy should be understood correctly to design privacy protection strategies and methods. As health data becomes more accessible through multiple heterogeneous sources, health informaticians should consider and balance data sharing along with privacy safeguards [14].

Privacy and Security. The concepts of privacy and security are closely related in the field of software engineering and design. However, there are subtle differences in these concepts. While privacy addresses personal data protection, security considers the processes and methods to safeguard assets essential for a system. Sensitive data inherent in electronic health apps demands stringent privacy and information security considerations. Privacy attacks and breaches incur financial and health-related costs that endanger patients' lives due to decisions based on faulty information. Therefore, it is important to consider protecting this information in designing and implementing e-health apps [15]. Security concerns and challenges to accessing EHR data include data ownership, confidentiality and integrity, access control, and user authentication [16]. Various access control models have been implemented in EHR, including Role-Based Access Control (RBAC), Attribute-Based Access Control (ABAC), Mandatory Access Control (MAC), and Discretionary Access Control (DAC), each employing unique authorization mechanisms to manage data access [17].

Confidentiality. As described before, privacy refers to patients' right to control access to their health data. The concept of confidentiality, however, extends the privacy domain, focusing on the protection of sensitive clinical information. It emphasizes the agreement between patients and healthcare providers (or settings) regarding confidential data handling and preservation. Health professionals with access to healthcare data (specifically sensitive information) are

ethically and legally obliged to maintain the confidentiality of this information. Security measures within health information systems are fundamental to protecting the privacy and confidentiality of health data [18]. Additionally, in some critical areas such as adolescent care, preserving the privacy and confidentiality of EHR data may pose several challenges. One issue is the parents' or guardians' access to sensitive health information, such as sexual health or mental health records, through billing systems or patient portals. One solution is developing adolescent-specific modules inherent in EHRs with a customized confidentiality consideration. Healthcare providers should ultimately address the balance between patient privacy, legal requirements, and the accessibility of medical records in an evolving digital domain [19].

Informed Consent. The rise of health information exchange for secondary purposes (such as research, policy development, or public health programs) has brought about various discussions among health professionals, patient advocacy groups, and privacy organizations. The most critical issue is whether patients (*and in some cases the general public*) are appropriately informed of these uses and whether proper consent procedures are in place to ensure responsible, secure, and informed data exchange [20]. The idea of informed consent is a critical field in medicine. This concept entails patients "voluntarily" making decisions regarding associated risks, research procedures, protective measures, and possible benefits. This information should be based on accurate and relevant information delivered to patients. Therefore, patients' participation in research should stem from their trust in healthcare settings and clinicians. Providing clear and fact-based information about privacy protections in research can help improve participants' long-term trust, satisfaction, and voluntariness in research processes. Essentially, when researchers and institutions emphasize *privacy measures and ethical protections*, participants may feel more secure and informed, which leads to a greater willingness to participate in research, as they trust the institutions conducting it [21]. Additionally, the process of obtaining consent from individuals should include systems that make it easy for them to access their data. When users can request the correction or removal of improperly stored or inaccurate data, they have a sense of control over their data. These rights are currently protected under the General Data Protection Regulation (GDPR), which ensures that individuals can manage and safeguard their personal information [22]. Research on consumer willingness to share electronic data showed that a significant number of users were unwilling to share their data for health-related purposes. This reluctance suggests that current privacy-preserving measures may not be sufficient, and introducing stronger and stricter regulations could help build consumer trust to consent to data sharing [23].

Ethical Approval. Obtaining research ethics approval is a critical and necessary step that should be performed before clinical research, even for low-risk studies. Most academic journals mandate that each research, including human participants, should report and confirm the acquisition of ethical approval by

the Human Research Ethics Committee (HREC) in the manuscript body. This confirmation guarantees the compliance of the research with relevant ethical standards and strengthens the credibility of health promotion research processes [24]. Without appropriate informed consent and ethical approval, participants' information may be at risk of privacy and confidentiality violations [25].

Audit Trails and Access Control. The "audit" method focuses on the assessment of e-health software security and is defined by usability, agility, objectivity, repeatability, and systematicity [15]. Access control measures can be used for credentials identification and patient registration when using EHRs or patient portals. A **shared** or proxy access control can also be implemented in special cases, such as in elderly profiles. This is a usable approach, however, it is important to recognize its technical, workflow, and informational barriers [26].

3.2 Privacy-Preserving Concepts in Healthcare

Anonymization. This method involves an irreversible de-identification process (*removing identifiable data elements*) designed to transform personal data into anonymous data before sharing. Anonymizing personal data reduces the risk of identifying an individual [27]. For data to be truly *anonymous* under the General Data Protection Regulation (GDPR), it must be processed in a way that re-identification is impossible by reasonable means. While GDPR provides overarching principles for anonymization, its practical application is not quite straightforward, considering the technological constraints and evolving data processing abilities. Anonymization can be performed using noise addition, microaggregation, and rank swapping. Different organizations may, however, interpret the requirements differently, making the anonymization method selection subject to factors like data type, scope, and intended use [28]. In the healthcare domain, data anonymization preserves patient privacy and anonymity during the secondary use of health data maintained in EHR. One suggested approach to overcome the limitations and challenges of dealing with heterogeneous data sources is converting source data to Common Data Model (CDM) standards to obtain suitable metadata for sharing and research purposes. Anonymization of this converted data guarantees secure and private sharing of harmonized data. Privacy models include l-diversity, t-closeness, differential privacy, and k-anonymization. These approaches can be performed on CDM databases such as PCORnet, Clinical Data Interchange Standards Consortium (CDISC), Informatics for Integrating Biology & the Bedside (i2b2), Observational Medical Outcomes Partnership (OMOP), or Fast Healthcare Interoperability Resources (FHIR)-formatted data [29].

De-identification. Some measures to preserve patient confidentiality include data *de-identification* that can follow the Health Insurance Portability and Accountability Act of 1996 (HIPAA) Privacy Rule guidelines. HIPAA suggests the *safe harbor* method or *expert determination* to ensure the correct

de-identification. The former involves removing PHI elements, while the latter relies on statistical analysis conducted by experts to ensure the minimal risk of re-identification [30]. In some cases, both methods can be used together to ensure privacy. In one study, to ensure the compliance of the de-identified clinical notes with HIPAA rules, the "safe harbor" criterion was applied to all PHI, except for dates. The expert determination was then used for de-identifying date-type fields [31]. The de-identification process can sometimes be rather insufficient, as the de-identified data has the potential to be re-identified, allowing the participants to be traced back. However, some data types, such as genomics, are not always classified as "identifiable data," which can make privacy protection even more complicated [32]. Unlike anonymization, the de-identification of healthcare data can be reversible, re-linking the data to external datasets to re-identify patients. Therefore, the decision to select each privacy approach depends on the scope and purpose of data usage [33]. De-identification technologies based on "*encryption*" may become susceptible to future computing advancements. To mitigate the massive leakage of de-identified healthcare data, data protection practices should be implemented among healthcare providers. Additionally, regulatory and legal safeguards should be enhanced [34]. Patient privacy and safety precautions become even more critical when applying ML and Artificial Intelligence (AI) methods, as these approaches often require extensive data access and transfer. The de-identification process of patient notes was previously performed using Artificial Neural Networks (ANN) with high precision and recall, without the need for manual feature engineering (i.e., a human-performed process for PHI removal) [35]. However, de-identification is not always an optimal technique for maintaining privacy since it may be susceptible to a Membership Inference Attack (MIA) (*i.e., revealing the inclusion of a data point in a dataset or a training set of an ML model*) when ML methods are utilized. Therefore, using the Large Language Model (LLM)-generated *"synthetic"* data has been considered an alternative technique for preserving patient privacy [36].

Pseudonymization. While both anonymization and encryption of medical data offer privacy protection, they each encompass limitations:

- Anonymization:

- *Irreversible* - limiting the ability of healthcare providers to use this data for primary care purposes when the corresponding patient should be known.
- *Identity removal* - patients cannot be informed about new findings from clinical studies [37]. This poses a bigger challenge in research when the data relates to rare diseases, as clinicians need to identify participants/patients [38].

- Encryption:

- *Access restriction* - prevents clinical research progress without explicit patient permission (as decryption reveals the patient's identity).
- *Time-consuming* - encryption of large medical records [37].

Therefore, the "*pseudonymization*" approach, *a specific type of anonymization*, is presented to overcome these challenges. EHR extracts can be pseudonymized either partially or totally; the former retains some demographic data (such as date of birth, sex, and place of birth) at a user-defined precision level, while the latter removes all references to demographic data [39]. In this method, the identifiable information is modified and substituted with a unique identifier (i.e., pseudonym) that remains unconnected with the original data without access to a "secret" key. Pseudonymization (also called linked-anonymization or pseudo-anonymization) enables the link between data and the corresponding patient only under controlled and specified conditions. This process requires two tables: one to store personal data and another to encompass pseudonyms and pseudonymized data. The process of isolating personal information from other data is known as "*depersonalization*". In some cases, re-identification is also possible. PIPE *(Pseudonymization of Information for Privacy in e-Health)* is a protocol for health data pseudonymization that can securely integrate both the primary and secondary use of healthcare data. It offers "*traceable anonymity*" to safeguard the confidentiality of health records [37]. In the pseudonymization process, certain issues need to be identified and resolved, including the generation of pseudonyms *without the possibility of being traced back* to a patient's real identity, ensuring *verifiability of multiple different pseudonyms* generated for one patient, ensuring verification of the *genuineness of the pseudonyms* at patient registration and administration settings, and *enabling authorized access* to all the encrypted EHRs associated with a patient's different pseudonyms and IDs to enhance update or secondary use (*in the patient's absence*) [40]. There are several pseudonymization tools for biomedical research, including but not limited to OpenPseudonymiser, SPIDER, and Mainzelliste, each with different specifications and contexts of use. Other tools like the "*FHIR Pseudonymizer*" are also available to maintain clinical data in various interoperability formats [41]. Some approaches are also presented in order to enable patients to be involved in the pseudonymization process. EHR systems can ensure privacy by pseudonymizing healthcare records before storage. Identifiers and QIs can be removed to prevent patient data (identity) disclosure. Each patient generates a unique and locally-created pseudonym (a long and random number) without any information exchange with the system. These pseudonyms are encrypted and then appended to the healthcare records, requiring subsequent decryption by the same patient at the time of new data addition [16]. Another suggestion is using different/separate information centers to manage EHR metadata in hospitals. Clinicians can submit diagnoses and request medical records from these centers without being necessarily involved in the processes of EHR sharing, searching, encryption, and decryption. Patients use different pseudonyms for each visit and control the encryption and decryption keys of their EHRs. This approach is mainly useful when dealing with the medical information of *Very Important Patients (VIPs)* like political leaders [40]. Multicentric pseudonymization was also considered in another study. It was suggested to use an identical pseudonym regardless of the data collection center/setting. In

this approach, two steps were recommended to link data across different institutions; first, the study center pseudonymizes the patient's identity locally. Then, a trusted service takes the pseudonymized data from the previous step and generates a consistent and unique pseudonym that can be shared across different settings without exposing the patient's identity [42]. Although pseudonymization has several merits, it is crucial to perform this procedure without compromising data quality for research.

3.3 Privacy Risks and Challenges in Healthcare Platforms

PII leak is a serious threat to patient privacy. However, *with recent advancements in AI (such as supervised learning paradigms), online repositories, and analytics methods*, sensitive patient data can be learned without access to PII. Advanced data analytics and ML techniques can be used to link seemingly anonymous data back to specific individuals [43,44]. Even in the absence of PII, privacy can be compromised either through identity disclosure or attribute/membership disclosure. Patient data privacy can be at risk by AI inference, linkage attack, attribute disclosure, differential attack, and Membership inference attack. *Training data extraction* attacks and *model invasion* attacks pose privacy risks to AI models. Therefore, these models are not only vulnerable to breaches but can also be used to amplify privacy attacks [43]. These privacy issues can be exacerbated when EHR data (*even de-identified*) is used and shared for AI research projects. There have been cases where PHI was sold to companies "*without written authorization*" from patients, constituting a HIPAA violation. Nonetheless, in these cases, the *data use agreement (DUA)* may explicitly hinder AI companies from EHR data re-identification [45]. Due to these concerns and the lack of strict legal and ethical considerations for privacy preservation, a limited number of AI-based apps are utilized in real healthcare settings. In this respect, *federated learning* and some *hybrid approaches* can be utilized to enhance privacy [46]. Digital radiologic images can also be vulnerable to threats such as data breaches, hacking, and ransomware (*where hospital data is held hostage*), which can lead to fraud and identity theft [47]. Telemedicine can introduce some privacy challenges as well, during online meetings and videoconferencing, which may include intruder participation and eavesdropping due to weak encryption techniques [48]. Furthermore, in some areas, such as **genomics**, there are specific privacy threats like phenotype inference and identification, each with relevant attack techniques (*e.g., inference from markers, data linkage, genotype imputation, or DNA phenotyping*) [49]. In the field of big data, guidelines and recommendations are needed to enhance ethics and collaboration between partners. There are some challenges in big data and AI, including data source and collection (*gathering and storing data while ensuring ethics and privacy*), interpreting complex data models (*understanding and interpreting the results from advanced data analyses*), and clinical implications (*determining how to translate big data insights into practical applications for healthcare decision making*) [50].

3.4 More Technical Aspects of Privacy Preservation

Privacy-enhancing technologies (PET) can be generally categorized into *architectural*, *algorithmic*, and *augmentation* categories. Federated Learning (FL) and multi-party computation are among the techniques considered as architectural PETS; homomorphic encryption, Differential Privacy (DP), and zero-knowledge proofs are categorized as algorithmic PETS; and synthetic data and digital twining are categorized as augmentation PETS [51]. According to the definition provided by NIST, DP is described as "a mathematical definition of what it means to have privacy" [52]. In other words, the inability to differentiate between two distributions of output probability is considered DP. Figure 1 illustrates this concept, where $\mathcal{M}$ represents the *mechanism*, which is the differentially private analysis. As the Large Language Models (LLMs) are now very popular in every field, such as medicine, DP can be used to mitigate privacy concerns before utilizing the EHR data as LLM input. In addition to DP, other strategies - such as de-identification, FL, surrogate generation, and synthetic data generation - are recommended as context-aware approaches to preserving healthcare data privacy when using LLMs [53].

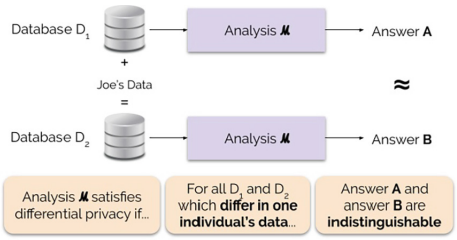

Fig. 1. Definition of Differential Privacy (DP). Credit: Nist.gov [52]

Healthcare data interoperability specifications, like OpenEHR, can be tested for their conformance with broader privacy regulations like the GDPR. Souza et al. [54] declared that it is challenging to fulfill some of GDPR's organizational requirements with any EHR specification standards like OpenEHR. In a study, the researchers identified 50 General Data Protection Regulation (GDPR) requirements and mapped them against 8 openEHR design principles. Each openEHR principle was compatible with GDPR; however, as a set, they addressed only 30% of the GDPR requirements (15 out of 50). This implies that openEHR is principally "privacy by design" but its intrinsic mechanisms are not sufficient to achieve **complete** GDPR compliance [55]. Like OpenEHR, the full conformance of the FHIR standard to HIPAA or GDPR may not be possible. Therefore, implementers should take the relevant requirements and regulations into consideration. FHIR offers some guidance to preserve privacy and security according to the specific risks and nature of each data type. Most resources need access control measures to *Create, Update, or Delete*. However, for READ

and QUERY access, some general guidance is needed based on the resource specifications such as **anonymous READ access resources** (not containing sensitive data, privacy approach: *server authenticated https (TLS)*), **business sensitive resources** (contain business or service-sensitive data, authentication approaches: *"mutual-authenticated-TLS, APIKey, App signed JWT or App OAuth client-id JWT"*), **individual sensitive** (not contain patient data but other participants', access control: *role-specific Role specific methods such as ABAC or RBAC*), **patient sensitive** (with very sensitive health information, access control: *privacy consent, security labels, and declared purpose of use*), **no dominant category** (a broad domain containing non-sensitive to very sensitive, privacy measures: *various above-mentioned approaches*) [56]. As mentioned earlier, anonymization is defined as removing PHI using the HIPAA safe harbor guideline. Nevertheless, before all else, it is important to correctly distinguish this data before eliminating it. The most common method for assessing detection quality is by comparing it to a human expert-curated model solution. One such solution is CensorCheck, designed to assess PHI detection in text specifically for anonymization and to define what qualifies as a match [57]. Considering security, the **audit** approach can utilize the combination of several methodologies, such as *manual evaluation*, *SAST* (Static Application Security Testing), and *DAST* (Dynamic Application Security Testing) to assess the security quality of an application. The security requirements of mHealth applications can be evaluated using a **continuous audit** method, which aligns with DevSecOps (Development, Security, and Operations). In this ongoing approach, the list of security requirements, referred to as the *catalog*, is implemented to ensure that measures are in place. Furthermore, it is continuously monitored and tested to verify that these measures function as intended. It is also verified for compliance with relevant security regulations and standards [15]. For multi-centric medical data protection, cryptographic methodologies have been recommended, such as public key and symmetric key encryption systems. The symmetric key approach uses an identical key for the encryption and decryption processes, while the public key method utilizes a pair of keys for encryption and decryption separately. Cryptographic approaches can be utilized in CDM-based research within multi-centric healthcare settings simultaneously enables coherent analysis while ensuring data protection throughout the process [58]. The multi-institutional or collaborative studies (*such as clinical trials*) can also be performed using privacy-preserving approaches to ensure compliance with strict data protection regulations, such as GDPR, by not sharing patient data. These ideas include local analysis at a single center and sharing only summary statistics. Further, meta-analyses can be performed at a multi-center level to synthesize results from different centers or studies. This method prevents privacy breaches and complies with GDPR by safeguarding individual patient data within each local center [59]. Anonymization techniques can be used to anonymize personal records by clustering similar Quasi-Identifier (QI) attributes (*such as sex, weight, and age* of the patients), ensuring they become indistinguishable from other record sets within the same table. In the absence of direct identifiers (*e.g., Social Security*

Number (SSN) or name), QI can be considered an identifier, too. Therefore, the disclosure of identity is a major privacy breach in EHRs that arises internally (within the organization) or externally (between the data collector and the owner). In this respect, multiple data owners can anonymize their records separately in a collaborative approach. Then, *with k-anonymity*, the identity disclosure is prevented while allowing dynamic participation of data owners using a greedy heuristic technique. The researchers acknowledged the trade-off between privacy and data utility in the anonymization procedure. However, they did not quantitatively measure the impact of anonymization on data utility or report information loss metrics [60]. Generally, the core challenge in sharing EHR sensitive data for AI- and ML-based studies is performing valuable and meaningful research while ensuring patient confidentiality and data privacy. Additionally, there are limitations and risks in traditional privacy-preserving methods (*such as de-identifications*) and a lack of clear recommendations and guidance on proper anonymization of PII [61]. For k-anonymity, the challenge lies in identifying the best approach for data suppression and generalization that preserves as much data utility as possible while satisfying the requirements of k-anonymity and ensuring the lowest information loss. The anonymized dataset's utility is evaluated by the degree to which it can still represent the statistical characteristics of the original data, while ensuring privacy [62]. It has been reported that using the ARX de-identification tool to preserve data privacy also compromised dataset utility since certain records were eliminated and predictor variables were concealed [63]. To overcome these challenges, an approach called the Anonymization through Data Synthesis using Generative Adversarial Networks (ADS-GAN) was proposed by Yoon et al. [61]. This framework generates synthetic EHR data that preserves patient confidentiality. By defining and minimizing the "identifiability" of private and identifiable data (*using the generation of de-identified synthetic versions*), ADS-GAN ensures secure and legally compliant anonymized data sharing to enhance the development of ML tools. Therefore, this approach models a stronger privacy-utility tradeoff compared to other methods. Other researchers proposed ML approaches to automatically de-identify healthcare narratives. The *two-pass tagging, priority sorting, and multi-class models* method was proposed and validated in the study by Dehghan et al. [64] to enhance entity recognition in psychiatric evaluation notes. They proved that this approach could successfully identify HIPAA-defined PHI. However, some identities, such as"organization and profession", were challenging due to expression variability in these classes. In the cases of digital health images, it is advisable to maintain only the minimum required information for experiments and de-identified data. The de-identification of medical images can be performed by either using *"Digital Imaging and Communications in Medicine (DICOM) header de-identification*" or "*file conversion de-identification*". Before the de-identification process, data should be stored in a secure and isolated location with no internet connection ("*air-gapped*") and protected through data encryption and obfuscation. This minimizes the risk of unauthorized access to personal data and data tampering [65]. Additionally, the de-identification pro-

cess of medical images was performed previously, involving the anonymization of metadata and pixel data while allowing authorized access and search of the original data. This method was integrated into an open-source *Picture Archiving and Communication System* (PACS) archive and validated in a collaborative setting [66]. Mobile health (mHealth) apps can be evaluated for their compliance with user privacy policies. In a study, researchers employed supervised ML techniques to analyze these policies and predict whether personal data disclosures were included in the policy text. A dataset of 350 expert-annotated privacy policies (APP-350) from popular mHealth apps was used to train the ML-based system. The results showed that about 94.4% of the most popular mHealth apps available on Google Play included a privacy policy [67].

3.5 E-health Data Privacy Standards and Regulations

Despite various privacy regulations, the dynamic nature of healthcare technologies like cloud storage, AI, and telemedicine introduces new challenges in maintaining EHR-based data privacy [5,68]. The understanding and application of

Table 1. Electronic health data communication standards

Privacy regulation/standard	Description
ISO 25237:2017(en) [69]	*"Health informatics - Pseudonymization"*, facilities the clinical information exchange for secondary use [39]
ISO 27799:2016 [70]	*"Health informatics—Information security management in health using ISO/IEC 27002"*, provides further guidance to address some specific requirements of health information security [71]
ISO 22857:2013(en) [72]	*"Health informatics—Guidelines on data protection to facilitate trans-border flows of personal health data"*, the training and predictions should remain consistent with the intended purposes and permissions granted at the time of data collection [73]
ISO 81001-1:2021(en) [74]	*"Health software and health IT systems safety, effectiveness and security"*, it states that adopting health IT systems affects clinical and business workflows; therefore, these changes should be carefully managed [75]
ISO/TR 21332:2021 [76]	*"Health informatics—Cloud computing considerations for the security and privacy of health information systems"* and EHRs [77]
ISO/TS 14441:2013 [78]	*"Health informatics—Security and privacy requirements of EHR systems for use in conformity assessment"*

these privacy standards are critical to preserving confidentiality, integrity, and trust in EHR systems in an increasingly interconnected healthcare environment. Table 1 summarizes some of these standards and regulations related to electronic healthcare data.

The Health Information Technology for Economic and Clinical Health - (HITECH) Act aims to promote patient engagement in their care and coordinated healthcare and considers information safety, privacy, and security. This act strengthened HIPAA privacy and security rules [79]. More specifically, the National Institute of Standards and Technology (NIST) SP 800-53 rev. 5 introduces a broader guideline for "Security and Privacy Controls for Information Systems and Organizations" [80]. It is worth noting that NIST has also introduced a more specialized practice guide on *"Securing Electronic Health Records on Mobile Devices"*. Best practices and related standards in the NIST cybersecurity framework and the HIPAA security rule can be implemented using this guide [81]. Other concerns *namely confidentiality and personal data integrity* can be addressed with audit trail, versioning, and access control [54].

4 Conclusion

This study aimed to provide a comprehensive overview of the privacy concepts, techniques, regulations, and challenges in the healthcare domain. We categorized and assessed the papers dealing with these items in developing healthcare systems and applications such as EHRs. The privacy-preserving techniques in healthcare contain approaches such as automatic de-identification, pseudonymization, and anonymization. Some approaches utilized AI and ML methodologies to ensure privacy in healthcare. However, these approaches themselves may pose some extra privacy threats. Therefore, the developers of healthcare systems need to consider compliance with privacy regulations and standards from the beginning of the development phase to prevent data leakage and other privacy hazards.

References

1. Ali, T.E., Ali, F.I., Dakić, P., Zoltan, A.D.: Trends, prospects, challenges, and security in the healthcare internet of things. Computing **107**(1), 28 (2025)
2. Gariépy-Saper, K., Decarie, N.: Privacy of electronic health records: a review of the literature. J. Can. Health Libr. Assoc. **42**(1), 74 (2021)
3. Spini, G., et al.: New approach to privacy-preserving clinical decision support systems for HIV treatment. J. Med. Syst. **46**(12), 84 (2022)
4. Chen, Z., et al.: Harnessing the power of clinical decision support systems: challenges and opportunities. Open Heart. **10**(2), e002432 (2023)
5. Yadav, N., Pandey, S., Gupta, A., Dudani, P., Gupta, S., Rangarajan, K.: Data privacy in healthcare: in the era of artificial intelligence. Indian Dermatol. Online J. **14**(6), 788–92 (2023)
6. Gong, M., et al.: Privacy protection of sexually transmitted infections information from Chinese electronic medical records. Sci. Rep. **15**(1), 1296 (2025)

7. Nass, S.J., Levit, L.A., Gostin, L.O., et al.: The value and importance of health information privacy. In: Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research. National Academies Press (US) (2009)
8. Nowrozy, R., Ahmed, K., Kayes, A., Wang, H., McIntosh, T.R.: Privacy preservation of electronic health records in the modern era: a systematic survey. ACM Comput. Surv. **56**(8), 1–37 (2024)
9. Costagliola, G., Fuccella, V., Pascuccio, F.A.: Towards a trust, reputation and recommendation meta model. J. Vis. Lang. Comput. **25**(6), 850–7 (2014)
10. Filkins, B.L., et al.: Privacy and security in the era of digital health: what should translational researchers know and do about it? Am. J. Transl. Res. **8**(3), 1560 (2016)
11. Gupta, R., et al.: Consumer views on privacy protections and sharing of personal digital health information. JAMA Netw. Open **6**(3), e231305-5 (2023)
12. Belfrage, S., Helgesson, G., Lynøe, N.: Trust and digital privacy in healthcare: a cross-sectional descriptive study of trust and attitudes towards uses of electronic health data among the general public in Sweden. BMC Med. Ethics **23**(1), 19 (2022)
13. Basil, N.N., Ambe, S., Ekhator, C., Fonkem, E., Nduma, B.N.: Health records database and inherent security concerns: a review of the literature. Cureus **14**(10) (2022)
14. Holmes, J.H., Soualmia, L.F., Séroussi, B.: A 21st century embarrassment of riches: the balance between health data access, usage, and sharing. Yearb. Med. Inform. **27**(01), 005–006 (2018)
15. Mejía-Granda, C.M., Fernández-Alemán, J.L., de Gea, J., García-Berná, J.A.: A method and validation for auditing e-Health applications based on reusable software security requirements specifications. Int. J. Med. Inform. **194**, 105699 (2025)
16. Rai, B.K.: Patient-controlled mechanism using pseudonymization technique for ensuring the security and privacy of electronic health records. Int. J. Reliable Qual. E-Healthcare (IJRQEH) **11**(1), 1–15 (2022)
17. Al-Zubaidie, M., Zhang, Z., Zhang, J.: PAX: using pseudonymization and anonymization to protect patients' identities and data in the healthcare system. Int. J. Environ. Res. Public Health **16**(9), 1490 (2019)
18. Bani Issa, W., et al.: Privacy, confidentiality, security and patient safety concerns about electronic health records. Int. Nurs. Rev. **67**(2), 218–30 (2020)
19. ACOG-Committee, et al.: Confidentiality in adolescent health care: ACOG committee opinion, number 803. Obstet Gynecol. **135**(4), e171-7 (2020)
20. Riordan, F., Papoutsi, C., Reed, J.E., Marston, C., Bell, D., Majeed, A.: Patient and public attitudes towards informed consent models and levels of awareness of electronic health records in the UK. Int. J. Med. Inform. **84**(4), 237–47 (2015)
21. Golembiewski, E.H., et al.: An electronic tool to support patient-centered broad consent: a multi-arm randomized clinical trial in family medicine. Ann. Fam. Med. **19**(1), 16–23 (2021)
22. Heslop, P.A., Davies, K., Sayer, A., Witham, M.: Making consent for electronic health and social care data research fit for purpose in the 21st century. BMJ Health Care Inform. **27**(1), e100128 (2020)
23. Grande, D., et al.: Consumer willingness to share personal digital information for health-related uses. JAMA Netw. Open **5**(1), e2144787-7 (2022)
24. Newson, A.J., Lipworth, W.: Why should ethics approval be required prior to publication of health promotion research? Health Promot. J. Austr. **26**(3), 170–5 (2015)

25. Harnett, J.D.: Research ethics for clinical researchers. In: Clinical Epidemiology: Practice and Methods, pp. 53–64 (2021)
26. Gleason, K.T., et al.: A multisite demonstration of shared access to older adults' patient portals. JAMA Netw. Open **8**(2), e2461803-3 (2025)
27. Saporito, A., Tabari, P., De Rosa, M., Fuccella, V., Costagliola, G.: Exploring the privacy horizons: a survey on HCI & HRI. In: International Conference on Computational Science and Its Applications, pp. 113–125. Springer (2024)
28. Mehtälä, J., et al.: Utilization of anonymization techniques to create an external control arm for clinical trial data. BMC Med. Res. Methodol. **23**(1), 258 (2023)
29. Wabo, G.K., Prasser, F., Gierend, K., Siegel, F., Ganslandt, T., et al.: Data quality- and utility-compliant anonymization of common data model-harmonized electronic health record data: protocol for a scoping review. JMIR Res. Protoc. **12**(1), e46471 (2023)
30. Appelbaum, L., et al.: Development and experience with cancer risk prediction models using federated databases and electronic health records. Exon Publ., 17–31 (2022)
31. Radhakrishnan, L., et al.: A certified de-identification system for all clinical text documents for information extraction at scale. JAMIA Open. **6**(3), ooad045 (2023)
32. Goodman, D., Johnson, C.O., Bowen, D., Smith, M., Wenzel, L., Edwards, K.L.: A comparison of views regarding the use of de-identified data. Transl. Behav. Med. **8**(1), 113–8 (2018)
33. Differences Between De-Identification And Anonymization. https://www.protecto.ai/blog/differences-between-de-identification-and-anonymization/. Accessed 14 Mar 2025
34. Mandl, K.D., Perakslis, E.D.: HIPAA and the leak of "deidentified" EHR data. N. Engl. J. Med. **384**(23), 2171–2173 (2021)
35. Dernoncourt, F., Lee, J.Y., Uzuner, O., Szolovits, P.: De-identification of patient notes with recurrent neural networks. J. Am. Med. Inform. Assoc. **24**(3), 596–606 (2017)
36. Sarkar, A.R., Chuang, Y.S., Mohammed, N., Jiang, X.: De-identification is not enough: a comparison between de-identified and synthetic clinical notes. Sci. Rep. **14**(1), 29669 (2024)
37. Neubauer, T., Heurix, J.: A methodology for the pseudonymization of medical data. Int. J. Med. Inform. **80**(3), 190–204 (2011)
38. Shin, S.Y., Kim, H.S.: Data pseudonymization in a range that does not affect data quality: correlation with the degree of participation of clinicians. J. Korean Med. Sci. **36**(44) (2021)
39. Somolinos, R., et al.: Service for the pseudonymization of electronic healthcare records based on ISO/EN 13606 for the secondary use of information. IEEE J. Biomed. Health Inform. **19**(6), 1937–44 (2014)
40. Zhu, H., Keong, N.W.: Transitive pseudonyms mediated EHRs sharing for very important patients. In: Web Services–ICWS 2019: 26th International Conference, Held as Part of the Services Conference Federation, SCF 2019, San Diego, CA, USA, 25–30 June 2019, Proceedings, vol. 26, . pp. 80–94. Springer (2019)
41. Abu Attieh, H., Müller, A., Wirth, F.N., Prasser, F.: Pseudonymization tools for medical research: a systematic review. BMC Med. Inform. Decis. Mak. **25**(1), 1–10 (2025)
42. Iacono, L.L.: Multi-centric universal pseudonymisation for secondary use of the EHR. Stud. Health Technol. Inf. **126**, 239 (2007)

43. Narayan, S.M., Kohli, N., Martin, M.M.: Addressing contemporary threats in anonymised healthcare data using privacy engineering. npj Digit. Med. **8**(1), 145 (2025)
44. Murdoch, B.: Privacy and artificial intelligence: challenges for protecting health information in a new era. BMC Med. Ethics **22**, 1–5 (2021)
45. Duffourc, M.N., Gerke, S.: Health care AI and patient privacy–Dinerstein v Google. JAMA **331**(11), 909–10 (2024)
46. Khalid, N., Qayyum, A., Bilal, M., Al-Fuqaha, A., Qadir, J.: Privacy-preserving artificial intelligence in healthcare: techniques and applications. Comput. Biol. Med. **158**, 106848 (2023)
47. Desjardins, B., et al.: DICOM images have been hacked! Now what? Am. J. Roentgenol. **214**(4), 727–35 (2020)
48. Jalali, M.S., Landman, A., Gordon, W.J.: Telemedicine, privacy, and information security in the age of COVID-19. J. Am. Med. Inform. Assoc. **28**(3), 671–672 (2021)
49. Bonomi, L., Huang, Y., Ohno-Machado, L.: Privacy challenges and research opportunities for genomic data sharing. Nat. Genet. **52**(7), 646–654 (2020)
50. Gossec, L., et al.: EULAR points to consider for the use of big data in rheumatic and musculoskeletal diseases. Ann. Rheum. Dis. **79**(1), 69–76 (2020)
51. Jordan, S., Fontaine, C., Hendricks-Sturrup, R.: Selecting privacy-enhancing technologies for managing health data use. Front. Public Health **10**, 814163 (2022)
52. Differential Privacy for Privacy-Preserving Data Analysis: An Introduction to our Blog Series. https://www.nist.gov/blogs/cybersecurity-insights/differential-privacy-privacy-preserving-data-analysis-introduction-our. Accessed 24 Mar 2025
53. Jonnagaddala, J., Wong, Z.S.Y.: Privacy preserving strategies for electronic health records in the era of large language models. npj Digit. Med. **8**(1), 34 (2025)
54. Sousa, M., et al.: openEHR based systems and the general data protection regulation (GDPR). In: Building Continents of Knowledge in Oceans of Data: The Future of Co-Created eHealth, pp. 91–95. IOS Press (2018)
55. Gonçalves-Ferreira, D., et al.: OpenEHR and general data protection regulation: evaluation of principles and requirements. JMIR Med. Inform. **7**(1), e9845 (2019)
56. FHIR security. https://build.fhir.org/security.html. Accessed 24 Mar 2025
57. Klapaukh, R., Boyle, D.: CensorCheck: a tool for evaluating protected health information detection systems. In: Health Innovation Community: It Starts With US, pp. 174–175. IOS Press (2024)
58. Jeon, S., Shin, C., Ko, E., Moon, J.: A security scheme for distributing analysis codes supporting CDM-based research in a multi-center environment. Comput. Methods Programs Biomed. **226**, 107159 (2022)
59. Duraku, L.S., et al.: Collaborative hand surgery clinical research without sharing individual patient data; proof of principle study. J. Plast. Reconstr. Aesthetic Surgery **75**(7), 2242–50 (2022)
60. Andrew, J., Eunice, R.J., Karthikeyan, J.: An anonymization-based privacy-preserving data collection protocol for digital health data. Front. Public Health **11**, 1125011 (2023)
61. Yoon, J., Drumright, L.N., Van Der Schaar, M.: Anonymization through data synthesis using generative adversarial networks (ADS-GAN). IEEE J. Biomed. Health Inform. **24**(8), 2378–88 (2020)
62. Karagiannis, S., Ntantogian, C., Magkos, E., Tsohou, A., Ribeiro, L.L.: Mastering data privacy: leveraging K-anonymity for robust health data sharing. Int. J. Inf. Secur. **23**(3), 2189–201 (2024)

63. Im, E., Kim, H., Lee, H., Jiang, X., Kim, J.H.: Exploring the tradeoff between data privacy and utility with a clinical data analysis use case. BMC Med. Inform. Decis. Mak. **24**(1), 147 (2024)
64. Dehghan, A., Kovacevic, A., Karystianis, G., Keane, J.A., Nenadic, G.: Learning to identify protected health information by integrating knowledge-and data-driven algorithms: a case study on psychiatric evaluation notes. J. Biomed. Inform. **75**, S28-33 (2017)
65. Parker, W., et al.: Canadian association of radiologists white paper on de-identification of medical imaging: Part 2, practical considerations. Can. Assoc. Radiol. J. **72**(1), 25–34 (2021)
66. Silva, J.M., Pinho, E., Monteiro, E., Silva, J.F., Costa, C.: Controlled searching in reversibly de-identified medical imaging archives. J. Biomed. Inform. **77**, 81–90 (2018)
67. Tangari, G., Ikram, M., Ijaz, K., Kaafar, M.A., Berkovsky, S.: Mobile health and privacy: cross sectional study. BMJ, 373 (2021)
68. Sharma, S., Rawal, R., Shah, D.: Addressing the challenges of AI-based telemedicine: best practices and lessons learned. J. Educ. Health Promot. **1**, 338 (2023)
69. ISO 25237:2017 Health informatics—Pseudonymization. https://www.iso.org/standard/63553.html. Accessed 21 Mar 2025
70. ISO 27799:2016 - Health informatics—Information security management in health using ISO/IEC 27002. https://www.iso.org/standard/62777.html. Accessed 21 Mar 2025
71. Box, D., Pottas, D.: Trust–Can it be controlled? In: MEDINFO 2010, pp. 651–655. IOS Press (2010)
72. ISO 22857:2013-Health informatics—Guidelines on data protection to facilitate trans-border flows of personal health data. https://www.iso.org/standard/52955.html. Accessed 21 Mar 2025
73. Dai, L., Zhou, M., Liu, H.: Recent applications of convolutional neural networks in medical data analysis. In: Federated Learning and AI for Healthcare 5.0, pp. 119–131. IGI Global Scientific Publishing (2024)
74. ISO 81001-1:2021 Health software and health IT systems safety, effectiveness and security. https://www.iso.org/standard/52955.html. Accessed 21 Mar 2025
75. MacMahon, S.T., Richardson, I.: Pathways, technology and the patient–connected health through the lifecycle. Front. Digit. Health **5**, 1057518 (2023)
76. ISO/TR 21332:2021 Health informatics—Cloud computing considerations for the security and privacy of health information systems. https://www.iso.org/standard/70568.html. Accessed 21 Mar 2025
77. Carello, M.P., Spaccamela, A.M., Querzoni, L., Angelini, M.: A systematization of cybersecurity regulations, standards and guidelines for the healthcare sector (2023). arXiv preprint arXiv:2304.14955
78. ISO/TS 14441:2013 Health informatics—Security and privacy requirements of EHR systems for use in conformity assessment. https://www.iso.org/standard/61347.html. Accessed 21 Mar 2025
79. Rosenbloom, S.T., Smith, J.R., Bowen, R., Burns, J., Riplinger, L., Payne, T.H.: Updating HIPAA for the electronic medical record era. J. Am. Med. Inform. Assoc. **26**(10), 1115–9 (2019)
80. Force, J.T.: Security and privacy controls for information systems and organizations. NIST Spec. Publ. **1**(5), 465 (2020)
81. O'Brien, G., et al.: Securing electronic health records on mobile devices, 1b. NIST Special Publication (2018)