Review Article

DIGITAL HEALTH

Application of privacy protection technology to healthcare big data

DIGITAL HEALTH
Volume 10: 1–22
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076241282242
journals.sagepub.com/home/dhj
Sage

Hyunah Shin[1], Kyeongmin Ryu[1], Jong-Yeup Kim[1,2,3] and Suehyun Lee[4]

Abstract

With the advent of the big data era, data security issues are becoming more common. Healthcare organizations have more data to use for analysis, but they lose money every year due to their inability to prevent data leakage. To overcome these challenges, research on the use of data protection technologies in healthcare is actively underway, particularly research on state-of-the-art technologies, such as federated learning announced by Google and blockchain technology, which has recently attracted attention. To learn about these research efforts, we explored the research, methods, and limitations of the most widely used privacy technologies. After investigating related papers published between 2017 and 2023 and identifying the latest technology trends, we selected related papers and reviewed related technologies. In the process, four technologies were the focus of this study: blockchain, federated learning, isomorphic encryption, and differential privacy. Overall, our analysis provides researchers with insight into privacy technology research by suggesting the limitations of current privacy technologies and suggesting future research directions.

Keywords

Blockchain, differential privacy, federated learning, homomorphic encryption privacy protection, healthcare

Submission date: 26 September 2023; Acceptance date: 23 August 2024

Introduction

A paradigm of the fourth industrial revolution, the demand for technologies that process and analyze big data in the healthcare sector has led to active research in the field. However, as the use of medical data increases, issues regarding the protection of personal information in the data are being raised. The most prominent issue is the area of personal information protection, as electronic records and billing data used in research contain personal information, such as patient gender, age, and address. This has led to many medical data breaches, and cases of personal information breaches caused by this can be found in many papers.[1] Many countries have enacted laws to prevent such information from leaking, such as the Personal Information Protection Act (PIPA) in Korea[2] and the General Data Protection Regulation (GDPR) in Europe.[3]

Due to such laws, conducting epidemiological research using patients' personal information contradicts protecting their personal information.[4] Therefore, various attempts are being made to analyze medical data while protecting patients' personal information, specifically using privacy protection technologies. Examples of such technologies are differential privacy (DP) and homomorphic encryption. More recently, concepts, such as blockchain and federated learning, have also been utilized to ensure the privacy of personal information.

When blockchain is decentralized, it shows excellent performance in the security sector, indicating that patients'

[1]Department of Healthcare Data Science Center, Konyang University Hospital, Daejeon, Republic of Korea
[2]Department of Otorhinolaryngology–Head and Neck Surgery, Konyang University College of Medicine, Daejeon, Republic of Korea
[3]Department of Biomedical Informatics, Konyang University College of Medicine, Daejeon, Republic of Korea
[4]College of IT Convergence, Gachon University, Seongnam, Republic of Korea

Corresponding author:
Suehyun Lee, College of IT Convergence, Gachon University, Seongnam, 13120, Republic of Korea.
Email: leesh@gachon.ac.kr

data can be used safely.[5] Blockchain has also been applied in the field of drug supply network management. However, since the risk of counterfeiting is high, it is applied only for certification, sales, and distribution monitoring. For the initial application, drug transactions were monitored by storing all the information on drug transactions in a blockchain.[6] Another technology, federated learning, enables data diversity when applied in the medical field. Federated learning improves artificial intelligence (AI) learning for patients with similar patient symptoms by connecting data from multiple medical institutions. Furthermore, the more data collected for similar patients, the greater the predictive accuracy of AI for the patient group. This can provide valuable insight to clinicians along with individual physician findings.[7]

Which methods were used to store and protect medical information before these developments emerged? Personal information protection technologies can be divided into two categories: firewall and encryption. A firewall is a type of network security that manages data entry and exit by monitoring received and transmitted information.[8] Although building a firewall on a network is expensive, it is generally one of the most effective data protection methods. Types of firewall include packet filter, stateful packet inspection, and application gateways. Encryption is a method of protecting an attack by encrypting an object, and data are protected by using decryption as a concrete example or by using a user-specified identify document (ID)/password (PW) method.[9]

Despite the use of such technologies, there have been instances where personal information was compromised. In the United States, medical data leaks occur every year through attacks on specific organizations, suspension of services, and data leakage. For example, the 2021 Accellion FTA hacking incident was a large-scale data leak that affected more than 100 companies. It is estimated that the health information of more than 3.51 million people was stolen. In the same year, 20/20 Eye Care Network, a Florida-based eye and ear management service provider, exposed the protected health information of 3,253,822 individuals because of an incorrect configuration of Amazon Web Services S3 cloud storage buckets.[10] Such vulnerabilities have revealed limitations in applying personal information protection technologies and the need to investigate solutions.

According to a report released on 23 March by the OECD,[11] more and more institutions are considering applying privacy technology. However, he mentions that the application of these technologies has been delayed due to many limitations in terms of their completeness. In doing so, we introduce four areas that can be divided into representative categories: data obfuscation, encrypted data processing, fed, distributed analytics, and data accountability tools. With this in mind, we wrote this article with the aim of getting researchers a guide to privacy technology in the future by selecting and reviewing technologies that are of high interest (many studies, popularity, etc.).

With this aim, the current study examines the methods, technology, research, and limitations of the most widely used personal information protection technologies for storing and protecting medical big data in recent years. The results of this study aim to provide robust insight into personal information protection technology for researchers to utilize medical big data safely for future research efforts.

What sets this paper apart from other papers is that it does not focus solely on just one technology, but provides insights on various privacy technologies. This has the advantage of allowing researchers hoping to gain various insights into privacy technologies to quickly learn about the current overall flow of related technology in a short period of time, which is different from other papers.

Methods

There are many privacy protection technologies. Data are protected in various ways, such as by adding noise to the data to protect the data or by using methods, such as protecting the data during its movement. Among them, Automated Validation of Internet Security Protocols and Applications (AVISPA)[12] is a tool that enables the analysis of privacy protection methods. It was announced in 2005. It is a push-button tool for automatic verification of protocols and applications sensitive to internet security, and it greatly helps the technology to be converted into a protocol in a systematic way. However, we will not cover this technology in this paper and will proceed with examining the trend by focusing on the technology.

A specific period was selected to explore recent trends in personal information protection technologies. This period was selected as 7 years up to December 2023. Thus, papers published between 2017 and December 2023 were examined. Next, we selected two databases for obtaining relevant research papers: IEEE and PubMed. We used the search terms "privacy" and "healthcare" to have maximum search results. Then, we reviewed the search results and selected appropriate works.

Inclusion and exclusion criteria

Papers that improve the technology by conducting analysis or research on each technology, or papers that improved the limitations of each technology were selected and reviewed. In the area of improving technology limitations, papers were selected and reviewed, focusing on the fact that the limitations of the technology can be understood, and new trends can be seen. Review papers with similar topics to this paper or papers that have to pay for access are excluded. The final selection was made after considering the duplication and quality of the studies.

Data analysis

To investigate recent research trends and specific privacy technologies, the title and abstract of finalized research papers were extracted and organized in Figure 1. Among the keywords derived, we performed keyword grouping in the "Healthcare data" and "Privacy protection technology" sections. Electronic medical records (EMR), claim data, omics data, and patient generated health data (PGHD) were the most common keywords concerning healthcare data. We explored privacy protection and technology separately in terms of privacy protection technology and obtained four keywords—blockchain, federated learning, homomorphic encryption, and DP —that this study focuses on.

The databases that were used to extract the data are IEEE and PubMed. IEEE is the Institute of Electrical and Electronics Engineers, which publishes papers on research related to engineering and technology. PubMed is a free website where you can search academic literature on biomedical and healthcare. To cover the search range that encompasses the keywords of science and technology and medical care, research papers were searched and organized through two databases.

Literature review

The IEEE and PubMed platforms have compiled a huge database of research papers published between January 2013 and December 2023 that comprise the keywords "healthcare" and "privacy." We found nearly 12,000(11,742) research papers published from 2017 to 2023. We chose this timeframe to observe relatively recent research trends. Papers that were not freely accessible, unrelated to data privacy, or did not utilize medical big data were excluded from the analysis. Finally, a total of 922 research papers were selected in the first search. Then, four technical keywords—blockchain, federated learning, homomorphic encryption, and DP —were derived through word counting in the 922 research papers. Related papers were selected from the 922 papers through four technical keywords, and 15 papers were selected for each technology through manual reviews. Figure 2 summarizes the process of selecting the research papers.

Results

In addition, Figure 3 shows the research trends from 2013 to 2023 regarding the application of medical big data for each personal information protection technology. It is to examine the flow of research trends. The period of data confirmed only in PubMed is represented. Research papers on four technology have been published only since 2015. But, papers began to be published in all four technologies from 2017 when the research began actively. Blockchain, which was introduced in 2008, has been the most actively studied topic until recently, and federated learning has also seen a sharp increase in research interest.

Privacy protection technologies

Blockchain. Blockchain is derived from electronic money technology developed to overcome the weaknesses of the

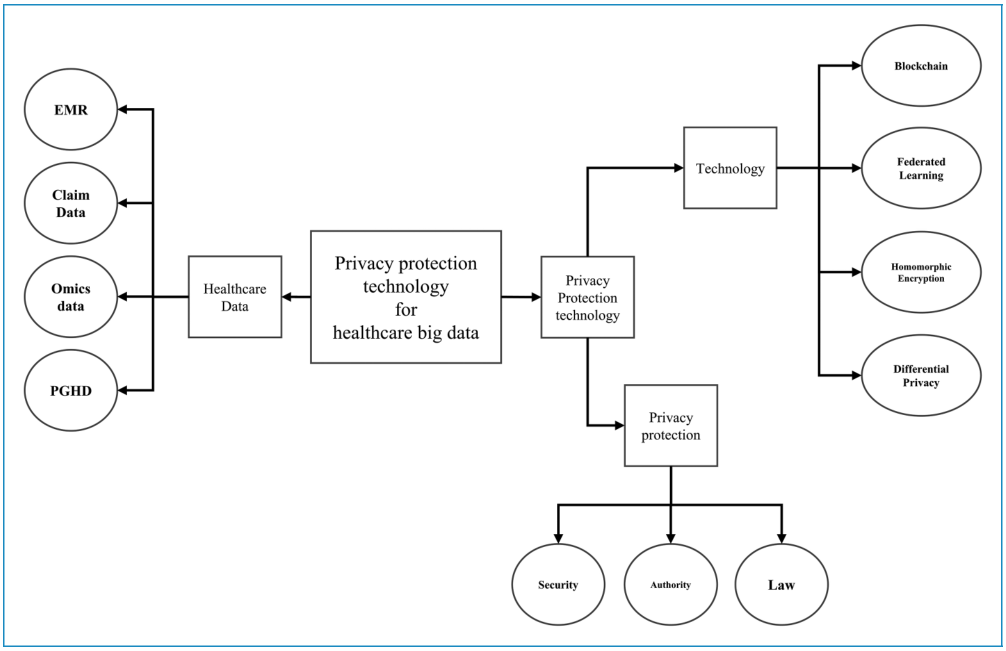

Figure 1. A schematic diagram of keyword calculation through word counting.

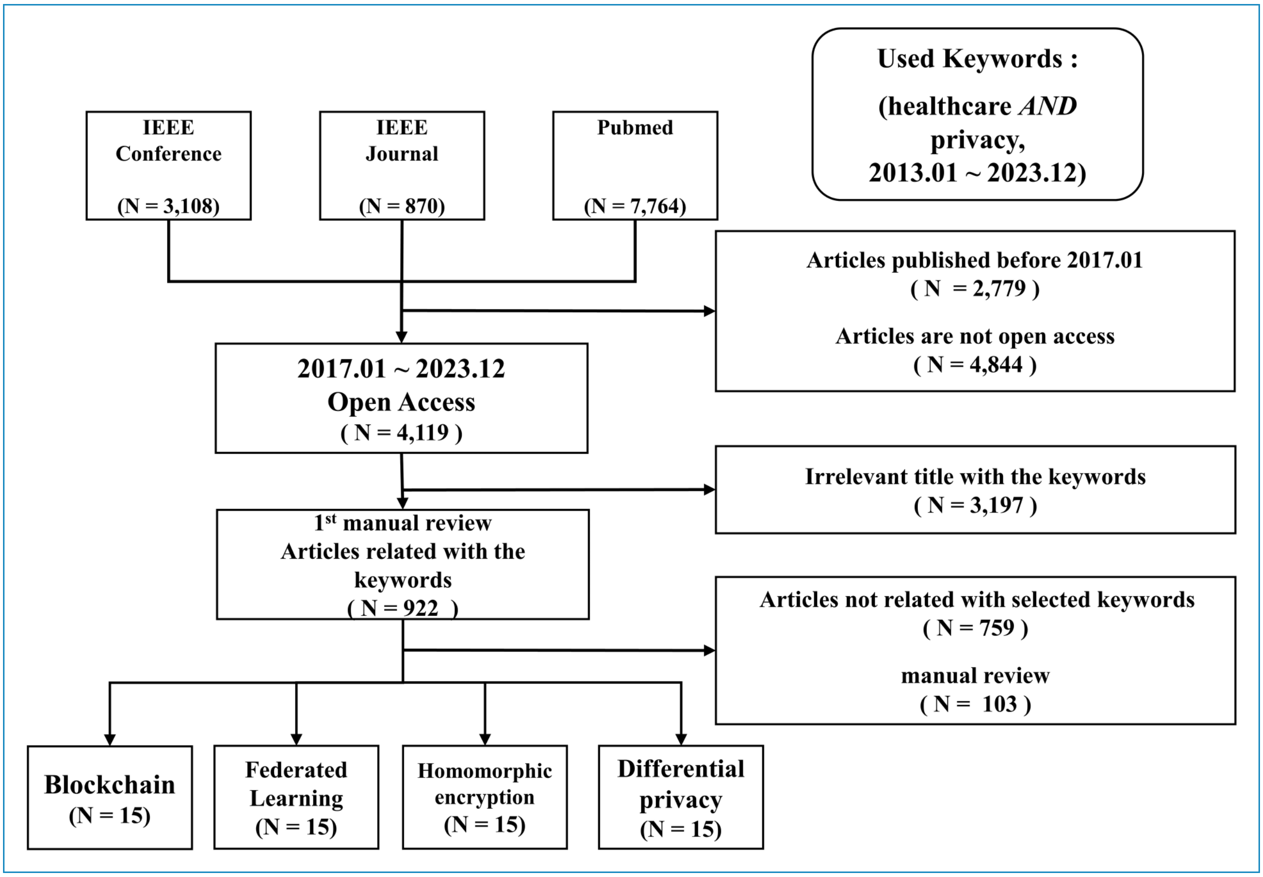
Used Keywords :
(healthcare AND privacy, 2013.01 ~ 2023.12)
IEEE Conference (N = 3,108)
IEEE Journal (N = 870)
Pubmed (N = 7,764)
Articles published before 2017.01 (N = 2,779)
Articles are not open access (N = 4,844)
2017.01 ~ 2023.12 Open Access (N = 4,119)
Irrelevant title with the keywords (N = 3,197)
1st manual review Articles related with the keywords (N = 922)
Articles not related with selected keywords (N = 759)
manual review (N = 103)
Blockchain (N = 15)
Federated Learning (N = 15)
Homomorphic encryption (N = 15)
Differential privacy (N = 15)

Figure 2. Flow of research for systematic literature review.

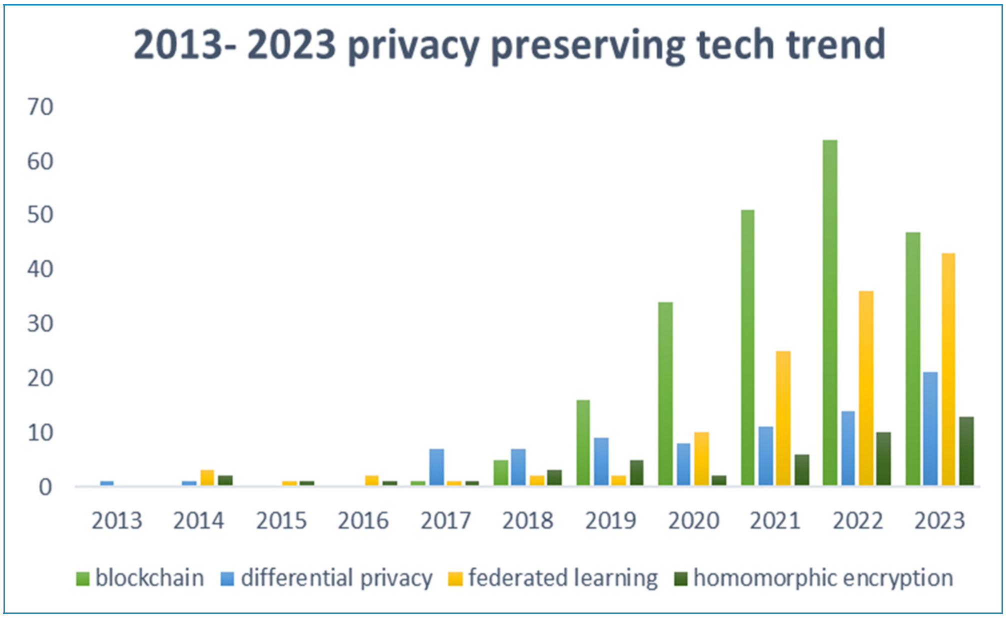

Figure 3. Personal information protection technology trends (2013–2023 in PubMed).

money transaction system through authorized third parties (banks, countries).[13] Blockchain can be interpreted as a series of data with timestamps, which form a block by combining the original data with a function called the hash. Each block forms a chain with the previous hash included in the next block, and hence, the term ‘blockchain’ is used. Blockchain authenticates the use of blocks constructed using timestamps to prevent duplicate payments, which is the same as replacing the role of an authorized third party, such as an existing country or bank. If a Block 1 user has malicious intent and wants to contact an external block immediately after contacting Block 2, the time stamp history of Chain 1 that has already been processed blocks the malicious attempt of Block 1.[14] Figure 4 presents a conceptual schematic design of blockchain.

Blockchain has been largely utilized for preserving, managing, and exchanging electronic health records (EHRs) (Table 1). EHRs were developed to overcome the difficulties in tracing and managing existing medical documentation methods and providing better medical services to patients. EHRs can help patients improve their health conditions by providing accurate information about previous diagnoses and treatments to their doctors.[23] This electronically stored data are also very useful in situations where a patient is rushed to another hospital. If electronically stored patient data can be safely transferred to another hospital, better medical treatment can be provided by quickly referring to the patient's previous medical records, especially in emergencies. However, the most important concern regarding EHRs is ensuring safe storage and movement of patients' personal and medical information because the data comprise very sensitive personal information. This study explains how blockchain can provide a solution for this concern.

All the research papers examined in this study have suggested ways to utilize blockchain to protect and utilize EHR data. Two papers have particularly focused on securely exchanging EHR data and cited leakage of personal information as the biggest barrier to health information exchange.[20,25] One of them has focused on patient identification through blockchain transaction information, an issue pointed out in conventional blockchain-based HIE research.[20] Data identification was introduced as a way to analyze transaction recipients and callers, even if the data are contained in blocks and encrypted in case a patient's personal information is leaked. It was also introduced to infer the patient through the treatment they have received. Therefore, the study provided a personal information protection solution through blockchain by concealing the personal information of the caller and receiver.[20] Both papers developed the framework using Ethereum, a private blockchain.[20,25]

Other studies have explained how blockchain has been applied to all aspects of storage, management, and exchange of medical data without focusing solely on EHR exchanges.[21–24,26–29] Some studies have focused solely on the security of health information derived from wearable devices and Internet of Things (IoT) devices.[21,22,26,27] The IoT technology refers to a network of physical objects with built-in internet-enabled software and sensor devices, and data collected from such devices are managed and exchanged without being directly operated by a person.[21] Such technology is being extensively applied in the medical field. Devices, such as the electrocardiogram (ECG), electroencephalogram (EEG), and blood pressure monitor, can easily collect and monitor medical data.[22] In addition to blockchain, research on IoT technologies has introduced several key privacy technologies, such as fog computing, Public Health Information Management System (PHIMS), and Interplanetary File System (IPFS). Fog computing is a service structure that creates a new relay layer between

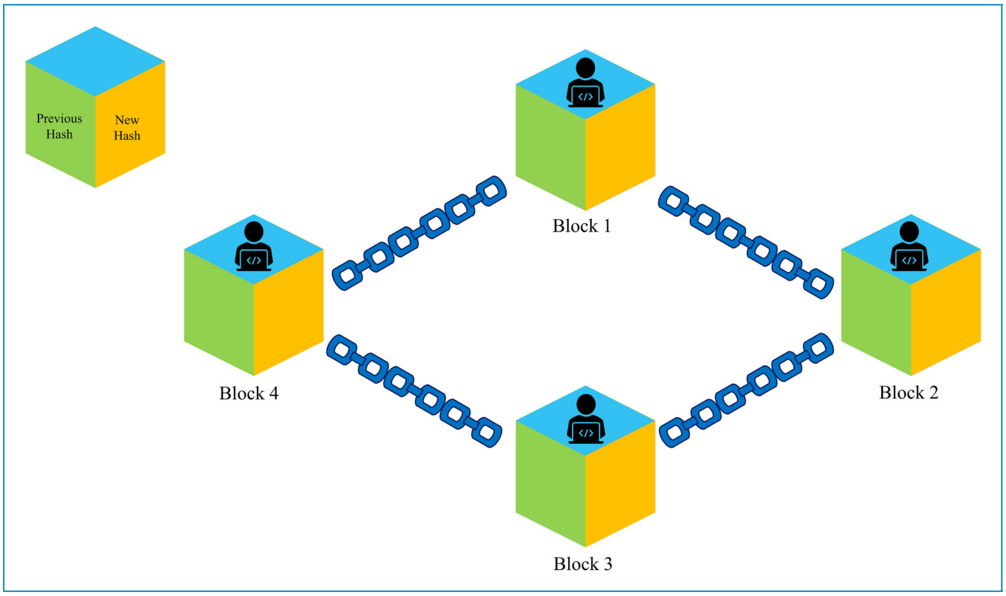

Figure 4. Conceptual schematic diagram of blockchain.[13]

Table 1. Blockchain (Bc) technology paper list.

Research paper	Year	Technology	Used/target data	Limitations	Key findings
15	2023	Bc			Gives better response time Improves the overall performance of the smart health supply chain management system
16	2023	Bc	Not publicly available	Consumes much energy	Achieved the optimal delay transactions
17	2022	Bc		Storing a large amount of data may create inefficiency and lead to more expensive issues	Improved security against known threats A slower rate of traffic growth More transparency Instantaneous traceability Robustness
18	2022	Bc			Ensures security properties, particularly data integrity, forge, binding, uniqueness, peer-indistinguishability, and revocation
19	2021	Bc			A throughput higher than 100 TPS could be achieved
20	2021	Bc	Health information exchange	Processing time for high access requests Trade-off between feature and time-consuming Buck passing responsibility to patient information	Solved problems in personal information inference Enhanced the security of personal information
21	2021	Bc	University of Queensland Vital Signs Dataset	Single-server infrastructure Not a real-time data-based result Need experiments on more diverse situations	Used the Hyperledger Fabric blockchain model Built a model that integrates blockchain with fog computing
22	2021	Bc	Medical Internet of Things (IoT) device data and Google fit application data	Buck passing responsibility to patient information	Applied the blockchain system to IoT using the Hyperledger Fabric blockchain model Proposed the blockchain-based Public Health Information Management System (PHIMS) for managing health data
23	2021	Bc	Electronic health records (EHR) data	Existence of time differences based on the size of data	Used a bilinear map function for improved security Proposed the Blockchain Security Framework (BSF) for effective preservation of EHR
24	2020	Bc	Radiation, medical, and surgical information for oncology information system	Risk of having a point of system failure Lack of provisions for emergencies System incompatible with	Used AWS storage, Hyperledger Fabric blockchain models, and HIPAA-compliant cloud storage Developed ACTION-EHR, a system for

(continued)

Table 1. Continued.

Research paper	Year	Technology	Used/target data	Limitations	Key findings
				patients' "Right to be Forgotten"	the radiation treatment of cancer patients
[25]	2020	Bc	SEER (Surveillance, Epidemiology, and End Results) dataset	Need to convert the server to a hospital environment Need to agree on an interoperability standard Scalability constraints	Leveraged Ethereum blockchain systems Developed a secure Health Information Exchange (HIE) technology that allows one to selectively control EHR records
[26]	2019	Bc	Data collected using wearable sensors	Lack of scalability with multiple data providers Need to reduce latency Potential data leakage during mining Service quality assurance not guaranteed	Combined the Ethereum blockchain model with decentralized storage Interplanetary File System (IPFS) Proposed a cloud blockchain framework for sharing EHRs
[27]	2019	Bc	Medical data collected from wearable devices	Potential security attacks due to resource constraints in IoT	Combined the advantages of private key, public key, blockchain, and many other lightweight encryption fundamental elements Developed a patient-centric access control for electronic medical records
[28]	2018	Bc	Radiation oncology data	Restrictions in the metadata structure Limited size of exchangeable data No actual application	Used blockchain to store, manage, and share oncology data Developed secure and reliable EHR data management and sharing systems
[29]	2018	Bc	EHR and data exchange	Increased costs based on institutions and the number of patients	Multiple rights for attribute-based signatures (ABSs) Proposed an MA-ABS system of blockchain structure to ensure the anonymity and invariance of information

cloud servers and IoT devices using a simpler concept than cloud computing. One study reported that using a combination of fog computing and blockchain in a medical environment would benefit both patients and hospitals.[21] In terms of relatively recent papers, rather than introducing new or additional technologies, many of the studies aimed to improve throughput or processing speed in existing technologies, with the goal of making them more practical and applicable to real life while preserving privacy.[15–19] The advancement of these technologies can help overcome the current scalability problem of blockchain. Currently, blockchain faces the problem of scalability due to the principle of the technology, and this problem is a prerequisite for improving the problems faced by this technology (accommodating a large amount of data or protecting various data). By solving these processing speeds or processing volumes, it will be possible to further grow into a technology closely related to real life.[30]

Federated learning. Deployed through the cloud infrastructure, the federated learning technology provides an additional approach to models trained through an interaction between mobile devices and users. Google has applied this technology to Gboard, a virtual keyboard app for Android.[31] Expanding into multi-agency research, first, each institution downloads a general global model. The downloaded model uses data from each institution to create a unique model. The difference from the existing model is encrypted by mixing noise that both cloud servers and institutions are aware of and then transmitted to the cloud. The transmitted information is used for

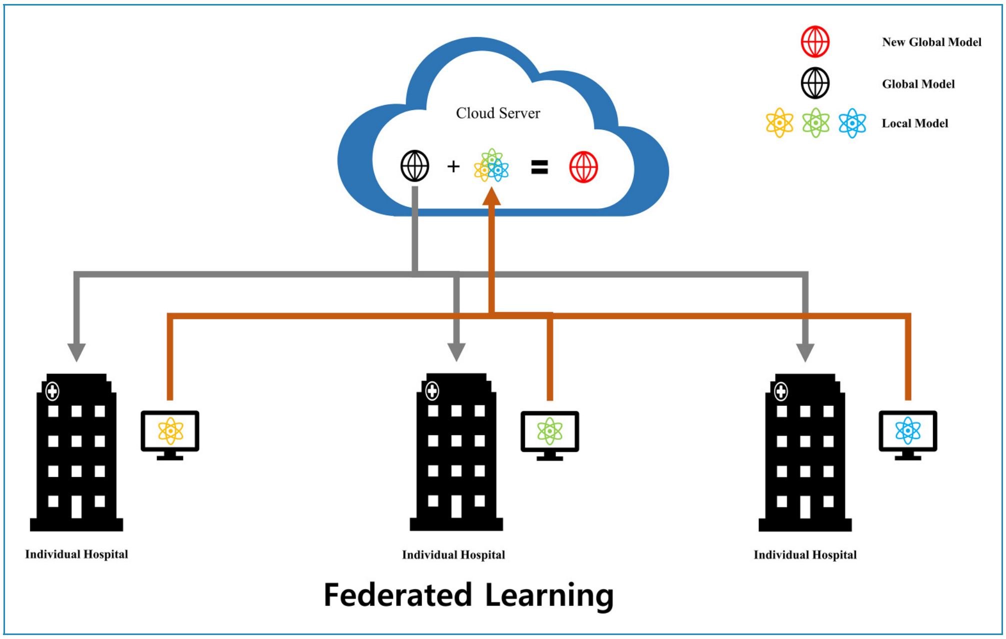

Figure 5. Conceptual diagram of the federated learning technology.[32]

model learning after decoding. These steps are repeated continuously until the model achieves the desired performance or the final deadline is achieved.[32] Figure 5 presents a conceptual diagram of the federated learning technology.

Most studies on federated learning have been conducted to prove its validity by first using open-source datasets to create a similar environment (Table 2).[38,40,43,45–47] A typical dataset, called MIMIC-III, has been primarily used in these studies. The MIMIC-III is an open-source database containing data of patients aged 16 or more admitted to the Beth Israel Deaconess Medical Center (BIDMC), a teaching hospital of Harvard Medical School.[46] Other studies have been conducted by extracting actual data from specific hospitals and making actual predictions, such as predicting abnormalities based on CT images of COVID-19 patients[41] and hospital EHR data.[42]

Research on joint learning is divided into two themes. The first research theme compares the federated learning model with the conventional machine learning model to prove the validity of the former's use.[40–43,47] The second research theme addresses issues related to federated learning, such as communication costs.[38,40,44] Some of these studies have applied federated learning from textual data,[43,47] mobile data,[40] and image data.[41] All of these study have examined whether federated learning can be applied to the selected data type, and if so, whether there are existing methods and advantages of doing it. Opinions on the study results were divided into (a) not different from the existing methods[43] or (b) presents an advantage in terms of performance.[40–42,47] One study explained that although there was a time advantage, it was a slight one in terms of the area under curve.[46]

Another study was conducted to predict a person's emotional state using data collected through wearable mobile devices.[40] Wearable devices are a way to obtain patient data in real time. Various types of patient data, such as body temperature, heart rate, and electrical skin activity, can be extracted through a patient's worn device, which allows data related to an individual's cognitive, behavioral, and emotional states to be collected in real time. Data collected from these wearable devices are used to analyze AI models created by a central server, and instead of learning predictive models directly, a brief training is conducted on each user's device. A copy of the parameters calculated through the training conducted on each device is sent to the central server and used to enhance the model's accuracy. Through this process, a model for predicting patient's emotional state is created.[40] Based on the relatively recent papers, similar to blockchain, there have been many efforts to apply this technology to real-world applications, such as the application of non-IID data,[33] which is similar to the environment of real-world datasets,[34,35] or the study of communication costs.[36] These efforts can be said to be aimed at finding answers to the problems currently facing federated learning. Federated learning was designed as a way to protect personal information from its origins, but it also has limitations. These problems range from problems that can generally infer members or data to attacks that can maliciously tamper with the model.[48] Future papers urgently need to develop models that can block these malicious attacks and can be used in realistic situations.

Differential privacy. DP is a technology developed to combine two conflicting concepts: privacy and data learning. DP aims to utilize useful information from data.

Table 2. Federated learning (Fe) technology paper list.

Research paper	Year	Technology	Used/target data	Limitations	Key findings
33	2023	Fe	MNIST, Fashion-MNIST (FaMNIST), CIFAR-10		Demonstrating superior overall performance Robustness to non-IID dataset skew
34	2022	Fe	Parkinson's disease dataset, NSL-KDD dataset	Computational complexity is limited by the growing number of hidden layers	Basic method is less expensive and faster.
35	2022	Fe	Some public and private local hospital datasets	The number of patients was small	Shows the promise of AI providing low-cost and scalable tools for lesion burden
36	2022	Fe	The MIND dataset	The lack of local data can lead to poor learning ability The reliability of the server is not guaranteed The computational cost of the local client increases slightly (in some cases)	Reduce 94.89% of communication cost Achieve competitive results with centralized model learning.
37	2021	Fe	Publicly available data	Limitations by design constraints : arbitrary data exploration	Application of federated techniques to modeling health data introduces new open questions and challenges
38	2021	Fe	CMS datasets, Medical Information Mart for Intensive Care-III (MIMIC-III) datasets, and proprietary synthetic data	Potential reduction in convergence rate due to distributed stochastic gradient descent (SGD)	Applied with federated tensor factorization Minimized communication costs Presented collaborative health data analysis methods Showed that federated generalized tensor factorization can improve communication efficiency problems
39	2021	Fe	Some datasets derived from the MIMIC-III dataset	The model may not be generalizable to an experiment Data correlation is not considered Performance may have degraded due to realistically situational settings	Model performance is proportional to the amount of data. Learning time increases in proportion to the increase in quantity. As the number of nodes increases, the model training speed increases.
40	2021	Fe	Wearable Stress and Affect Detection (WESAD) dataset	Risk of leakage of personal information due to centralized server model Performance degradation of individual models Need additional work to	Empirically demonstrated the compatibility of federated learning models as mobile health data prediction models Demonstrated that using personalization, which

(continued)

Table 2. Continued.

Research paper	Year	Technology	Used/target data	Limitations	Key findings
				provide more powerful privacy	considers individual differences in data, can help improve model accuracy
[41]	2021	Fe	Data of 132 actual patients recruited from seven multinational hospitals	Imbalance between hospitals due to insufficient number The concept shift phenomenon of machine learning from demographics	Used the federated learning approach for personal information Showed that trained convolutional neural networks-based (CNN-based) AI models are effective in detecting CT anomalies in COVID-19 patients
[42]	2021	Fe	Data collected from EHRs of five hospitals in Mount Sinai Health System	Limited data collection Lack of research on various aspects because the study focused on proof of principle Lack of diverse data Lack of algorithmic diversity Lack of architectural optimization	Estimated mortality among COVID-19 patients within 7 days of hospital admission using federated learning Showed that the learning model combining multilayer perceptron and least absolute shrinkage and selection operator outperforms conventional local prediction models
[43]	2020	Fe	Modified National Institute of Standards and Technology (MNIST), MIMIC-III, and electrocardiogram (ECG) datasets	Need to study how to select parameters High communication cost Potential leakage of personal information due to data speculation	Used benchmark datasets to evaluate the reliability and performance of federated learning Improved performance when using unbalanced datasets No difference in performance from existing centralized models
[44]	2020	Fe	Data of 23,203 patients collected from eight healthcare institutions in five countries	Biased data collection Adverse effects of cohort differences on model performance Insufficient value of estimated result time (2 years) Lack of data elements for individual predictions	Provided privacy and proposed quick data analysis algorithms Used algorithms that exchange only regression coefficient and regression coefficients with central servers Suggested ways to overcome the limitations of existing machine learning
[45]	2019	Fe	eICU collaborative research database	No difference from the existing single centralized model Similar to the traditional methods of clustering techniques	Based on data relevance Proposed a federated learning model using a dataset (community) Distributed data based on community clustering It outperforms existing federated learning.

(continued)

Table 2. Continued.

Research paper	Year	Technology	Used/target data	Limitations	Key findings
[46]	2018	Fe	MIMIC-III dataset	Assumed that there is a common functional event in the course of learning and proceeded with the experiment Performance degradation due to data imbalance Time complexity, more memory required	Used hash function and homomorphic encryption Effectively searched for similar patients from different institutions without having to exchange information directly
[47]	2018	Fe	Boston Medical Center's undifferentiated electronic heart recording dataset	Increased communication costs based on the number of participating nodes (hospitals)	Based on the forces described in the EHR Developed a distributed (federated) method for predicting hospital admission Applied distributed binary classification using sparse support vector machine Provided algorithms to solve prediction problems

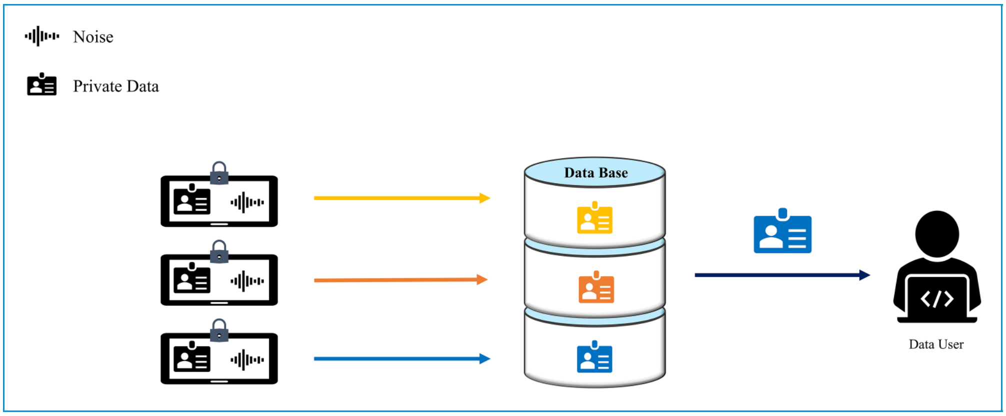

Figure 6. Conceptual diagram of the DP technology.[48]

However, it does not focus on learning anything from the data itself.[55] The primary purpose of DP is to share data. To use the data, the data user first transfers the query to a reliable data provider (data curator or database). Then, a data provider achieves personal information protection by providing noise-added data to the result of the transmitted query.[60] Figure 6 illustrates a simple model depicting the principle of DP.[49]

Research on DP technologies is often aimed at developing privacy capabilities by assisting technologies other than direct and primary use technologies. Major privacy protection technologies include blockchain,[55,61] combined learning,[55–57] deep learning,[58,64] binary classification,[59,62] and data sharing (Table 3).[60,63] DP is applied to these technologies in research because each technology has limitations in protecting personal information. For example, federated learning requires the use of a central server for data learning and data movement, which allows an attacker to launch a single point of failure attack that attacks the central server, and in blockchain technology, it is challenging to overcome attacks by malicious participants.[55] Therefore, two or three technologies are often combined to prevent the leakage of personal information that can occur when only one technology is used.

Table 3. Differential privacy (DP) technology paper list.

Research paper	Year	Technology	Used/target data	Limitations	Key findings
50	2023	DP	Provided real data		Devised low-sensitivity strategies for finding split coordinates Implemented effective privacy budget allocation strategies
51	2023	DP	Synthetic Dataset, Fingerprints Dataset, Molbace Dataset, ECG Dataset, Organ Meshes Dataset	May require additional design that needs to be optimized	strong privacy guarantees and excellent utility Learn similar features in the private and non-private scenarios Can help alleviate social impacts of machine learning
52	2022	DP	dataset of 30,072 WSIs from TCGA		Demonstrated the efficacy of federated learning (using both IID and non-IID data distributions) Private federated learning achieves a comparable result compared to conventional centralized training
53	2022	DP	provided by four clinics, located in three Australian states		Provides a practical solution with high diagnosis accuracy Preserving data privacy for a large-scale deployment.
54	2021	DP	Movielens dataset		Paper propose a privacy-aware real estate recommendation method for elderly care in cloud platforms. The proposal is validated by a real-world dataset.
55	2021	DP	American National Institute of Diabetes and Digestive and Kidney Diseases dataset	Not different from conventional federated learning in prediction accuracy Incomplete security success rate Difficult to balance performance and security	Maintained blockchain so that edge nodes resist a single point of failure Presented a way for medical IoT devices to implement federated learning and utilize distributed clinical data Used adaptive DP and Tilt Verification Infrastructure Agreement Protocol Showed resistance to privacy and external attacks Applied DP to learning slope
56	2021	DP	Various open datasets including MIMIC-III	Presence of distributed characteristics that limit arbitrary data search Reliability reduction due to research validity assessment within existing datasets Limited types of data	Leveraged eight different datasets to form data silos Set up situations similar to those in real multicenter study Applied a federated learning model with integrated DP Applied federated learning model with DP in a distributed environment

(continued)

Table 3. Continued.

Research paper	Year	Technology	Used/target data	Limitations	Key findings
					Not different from conventional centralized models Applied DP to learned parameter information
57	2021	DP	Project NeLL™ Database	Data diversity must be guaranteed Dominant discrimination patterns can appear due to differences in the size of the data Used only patient ICD-9 code	Enabled sequential pattern mining on distributed data sources Utilized the Personal Information Security Framework Suggested a representative pattern discovery method DP mechanisms do not differ from conventional methods in terms of efficiency Applied DP to aggregate pattern information when aggregating extracted patterns
58	2021	DP	Spectralis OCT, Heidelberg Engineering, Germany, Medical Segmentation Decathlon (MSD) Liver segmentation dataset	Performance degradation in classification or subdivision operations Increased training time and memory consumption Neural network architecture incompatible DP-SGD algorithm	Presented deeper for DP deep learning DP-SGD algorithm with DP for image analysis capability is functional reconstruction Demonstrated effective protection of model and personal information against external attacks Applied DP to result gradient vectors
59	2020	DP	SEER dataset	Centralized model design DP mitigation problem	Studied privacy risks when performing survival analysis Proposed a differential personal framework for Kaplan-Meier product-limit estimators DP maintains usefulness in predicting survival curves, avoiding inference of an event-based temporal reasoning for individuals without serious errors Applied DP to published survival analysis results
60	2019	DP	NPS dataset of Health Insurance Review and Assessment	Does not apply to many data types	Utilized attribute-based encryption and local DP Proposed mechanisms for protecting data owners' personal information during data sharing process Showed that more accurate data exchange is possible while increasing reliability among data providers Applied DP to source data

(continued)

Table 3. Continued.

Research paper	Year	Technology	Used/target data	Limitations	Key findings
[61]	2019	DP	Wisconsin breast cancer dataset with a binary class label and MNIST dataset	Need more complementary points in security part	Enabled DP Proposed distributed machine learning model for allowed blockchain Used DP stochastic gradient descent method Showed high elasticity for various attacks on adversarial nodes (model accuracy decline) Applied DP to Gaussian mechanisms
[62]	2019	DP	MIT physionet pcg dataset	No performance advantage due to the versatility of the model used Not compatible with privacy features	Used data-based calculations and DP High-precision identification of non-normal biological signals (80%) Showed that data-based methods with privacy are reliable for both physicians and patients through experiments Applied DP to shared clinical analysis results
[63]	2018	DP	Pamap2 dataset	Performance varies based on data volume	Used local DP Proposed mechanisms for collecting smart health information while protecting personal information Moved only significant data elements identified by local DP instead of moving all data Used the method of reconfiguring data at the data collection destination Demonstrated good data utilization while protecting personal information Applied DP to data from model analysis
[64]	2017	DP	Yesiwell data/MNIST dataset	Risk of re-identification or re-construction of the data under DP	Proposed a new framework for developing DP applicable convolutional deep belief networks Used the Chebyshev equation Indicated that a deep learning model capable of protecting personal information can be created Applied DP to deep learning secret layer stage

The DP technology has been applied to slope-down methods,[55,58,61] original data before sharing,[59–63] reduced data,[56] and aggregated parameters,[57] among others. The reason for this application is that the main technologies of research, such as blockchain and federated learning, were designed to protect personal information. However, there are still limits to the seamless protection of personal information when such technologies work independently. Such vulnerabilities often allow experienced attackers to infer or leak patient information through paradoxical systems.[58]

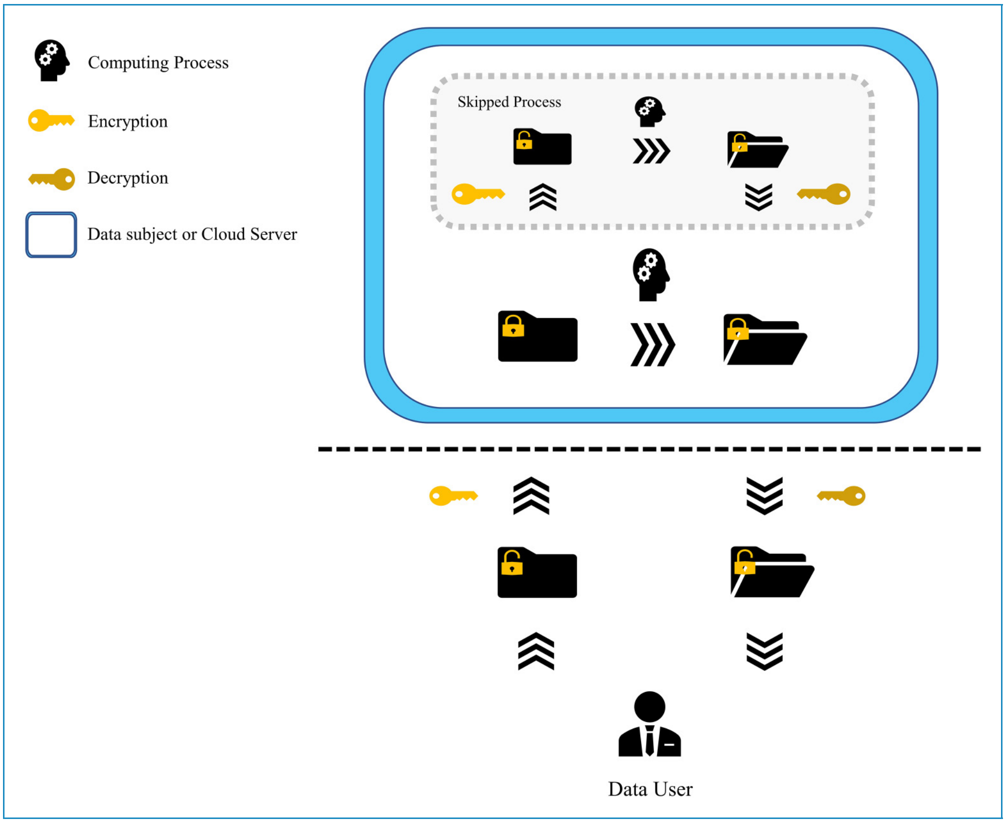

Figure 7. Conceptual diagram of the homomorphic encryption technology.[65]

Therefore, researchers have consistently suggested improving the privacy protection of personal information through the application of DP.

A study on pattern prediction models utilized distributed EHR data to predict patient states.[57] It employed a technique, called sequential pattern mining, which is a method of studying data recorded in chronological order and discovering unexpected patterns. In this paper, they used a method of aggregating pattern data learned by each institution, that is, aggregating the data by placing a server in the center to discover abnormal patterns, which are very sensitive to data leakage of medical evidence and prescriptions received by patients. In contrast, a large amount of EHR data is required to detect patterns, leading to a high risk of data leakage. Therefore, DP was applied to the data to prevent data leakage. According to the relatively recent papers, like the two techniques above, most of the studies are aimed at protecting privacy and utilizing data through the practical application of DP. Studies[50,53] that utilize real-world data for this purpose illustrate this. The recent trend has been to focus on creating technologies that directly utilize DP in real-world applications (or through research that assumes real-world-like situations). However, these practical applications have had the problem of not being able to achieve both privacy and performance, which are difficult to reconcile, for a while. The paper of Mueller et al.[51] actually shows that these two are compatible. In addition to this compatibility, future papers should conduct studies that apply extended functions and performance to real-world data.

Homomorphic encryption. Homomorphic encryption is an encryption method that can perform operations using encrypted data without decrypting the ciphertext. In homomorphic encryption, a stage for encrypting and transmitting data and decrypting and processing the transmitted data is not required. Figure 7 shows a simple schematic representation of homomorphic encryption.[65] Generally, encrypted data are transmitted with an encryption key; the transmitted data are decrypted; and then the data are processed. However, this process is omitted in homomorphic encryption. A data user encrypts the data with homomorphic encryption and transmits the data to a data processor or a cloud server. The transmitted data are processed in an encrypted state, and the data user decrypts the processed encrypted data and records the results.[80]

Homomorphic encryption has also been applied to protect personal information in combination with other technologies, similar to DP (Table 4). Among medical data, studies on the protection of genomic data, such as personal genes, are shown to be studies to which homogeneous encryption has been applied.[70,73,75] In other studies,

Table 4. Homomorphic encryption (He) technology paper list.

Research paper	Year	Technology	Used/target data	Limitations	Key findings
66	2023	He	The data will be provided upon request.	Lack of real-world validation	Offers a safe and open setting for the management and exchange of sensitive patient medical data
67	2022	He	J-HMDB, URFD, Multicam, MPII Human Pose dataset	Possible security issues depending on your environment; optimization issues with CNN models	A 613× speedup over the latency-optimized LoLa (Low-Latency CryptoNets) Achieves an average of 3.1× throughput increase in secure inference"
68	2022	He	The data will be provided upon request.	Security imperfections exist	Improved security and anonymity compared to the benchmark models
69	2022	He	Not available		Approach is resistant to active collusion and replay attacks
70	2021	He	Can see in: http://www.cbioportal.org/study/summary?id = tmb_mskcc_2018		Enabling biomedical insights that are not possible from individual institutions alone.
71	2021	He	NCBI, The Illumino	Lack of real-world application and validation	Methods can practically perform resource-intensive computations for high-throughput genetic data analysis
72	2021	He	Available online: https://github.com/fsumon/BioFusion1	End-to-end security has not been validated.	Offering some significant improvements to the overall security and privacy
73	2021	He	A study of metastatic cancer patients and their tumor mutational burden data and a host genetic study of human immunodeficiency virus type 1 (HIV-1)-infected patients data	Do not combine with anything other than the pricked data set	(Survival analysis + genome-wide association studies [GWASs]) + (federated analysis [FA] + multiparty homomorphic encryption [MHE]) Access method proposed Indicated that MHE can simultaneously protect personal information and share data Demonstrated the ability to achieve analytical workflows in complex biomedical fields
74	2021	He	100 data segments		Used blockchain and homomorphic encryption to address insider threats Proposed PPCCT(privacy

(continued)

Table 4. Continued.

(continued)

Research paper	Year	Technology	Used/target data	Limitations	Key findings
					preserving COVID19 contact tracing) using source encryption method Used the proposed architecture Showed that existing infectious disease contact tracking apps can ease personal information issues and ease disease spread curves
[75]	2021	He	1000 Genomes Project population panel of 2504 individuals	A decrease in relative accuracy compared to non-secure imputation methods for rare variants Existence of inherent limitations of homomorphic encryption Complex algorithm of homomorphic encryption Possibility of excessive optimism	Performed detailed benchmarking of time and memory requirements Used homomorphic encryption Showed that genomic data in genomic sequences (GWASs) can be shared, stored, and analyzed while maintaining large-scale data security Applied homomorphic encryption to shared data
[76]	2021	He	Used open data for a particular paper [On the Feasibility of Low-Cost Wearable Sensors for Multi-Modal Biometric Verification]	End-to-end security has not been validated Possibility of insufficient authentication due to differences in biological signals arising from the authenticator's attitude	Proposed a biological recognition identity management framework based on multi-mode Identity Management System (IDMS). Combined two biological signals and applied homomorphic encryption simultaneously Showed that accurate identification is possible by easing the leakage of personal information
[77]	2020	He	MNIST dataset and X-ray coronary angiography data	Possibility of personal information leakage due to key search attacks	Based on linear transformation Lower security compared to standard schemes but to be used in practical schemes Applied homomorphic encryption to the data
[78]	2019	He	A real heart disease dataset from the University of California Irvine's Machine Learning Repository	Relatively low accuracy for synthetic datasets	Used FHE(Fully Homomorphic Encryption) to protect personal information in source-outsourced biometric data An encrypted analysis indicated that the data owner can retrieve and decrypt it on the security side.

Table 4. Continued.

Research paper	Year	Technology	Used/target data	Limitations	Key findings
79	2019	He	1000-genome dataset	Possibility of personal information leakage Potential information leakage by untrusted service providers	Combining Intel SGX's security hardware features with homomorphic encryption Demonstrated efficient GWAS analysis while protecting privacy on federated genomic datasets
80	2018	He	Aggregate-level data extracted from the i2b2 data model	Potential increase in computational costs to ensure data sharing integrity	Applied homomorphic encryption to apply privacy enhancement technology to tools that enable collaborative medicine in clinical settings Showed efficient use for calculating aggregate queries for ciphertexts

homomorphic encryption has been applied to personal information function in deep learning analysis[74,77] or in an authentication management system that combines biometric information and the personal ID/PW method. Overall, homomorphic encryption is used to add privacy protection capabilities lacking in a single technology.[76] Among them, genetic research is remarkably widely applied in certain fields. Human genetic research, such as Genome Wide Association Studys (GWASs), has become a standard for customized medical treatment.[75] Such extensive studies continue to develop in terms of size and technology. With this development in GWASs, the amount of data required for research has undoubtedly increased. However, the construction of a database through large-scale data exchange implies a risk of personal information leakage and inference and reconstruction of medical and personal data. As a result, this data exchange required for large-scale research is strictly limited by numerous data protection regulations that vary from region to region.[73] According to these regulations, the technologies applied to GWAS research include blockchain and federated learning. However, each technology has its own limitations. One study investigated personal information protection in the exchange of human genetic data by using homomorphic encryption.[73] In particular, the authors simultaneously used combined learning and homomorphic encryption for biomedical research. In addition, unlike a personal information protection technology using existing encryption that may affect the analysis result by encryption, each organization analyzes the data, and then encrypts and shares the analyzed result. Consequently, the survival analysis was capable of protecting personal information, and the study was efficiently executed.[73] More recently, researchers have been working on improving the processing speed of homomorphic encryption, which is the most problematic aspect of homomorphic encryption.[67,71,72] This can be seen as an effort to take the technology described above and apply it to real-world environments to share and analyzing data while protecting privacy. Although the paper in[73] is not very suitable for real data, it shows an example of applying homomorphic encryption to a study on genetic data. This example shows that analysis is possible by applying homomorphic encryption to real data. This can be said to be an example that shows that homomorphic encryption can be utilized for analysis of real data and information protection. If the above technologies can be applied to real-world environments based on the improvements made through these studies, it is expected that a lot of research and improvements will be made.

Discussion

Our study examined the methods, technology, research, and limitations of the most widely used personal information protection technologies for the storage and protection of medical big data in recent years. Many researchers are investigating the protection of personal information in financial and medical fields, where it is critical. However, research papers published up to December 2021 exhibit limitations in security management, such as transferring responsibility for data management to patients, even though the technology has been applied to patient data. Blockchain is a personal information protection technology that has many advantages in ensuring data integrity, reliability, and transparency. Although it has great security advantages, there are ways to destroy the technology,[20] limiting its usability. Therefore, research on actual data exchange applied to multiple institutions to maintain the security and reliability of the privacy protection technology is a pertinent topic for future blockchain research.

When conducting research that highlights the shortcomings of federated learning,[37,38] the research purpose was to complement the shortcomings of federated learning using different methods. In one study, a method, called federated generalized tensor factorization, was used in applications, such as a recommendation system, space–time data analysis, and signal processing, which have the unique capability of expressing high-dimensional data. However, by sharing all related variables, a federated learning method that increases privacy while reducing communication costs by limiting existing tensor usage methods can be achieved.[37] Another study showed how differences in the amount of data, the number of computational nodes, and the distribution of data in federated networks affect the performance of federated learning.[38]

DP and homomorphic encryption are both technologies aimed at encryption; therefore, they have similar limitations. In the case of DP, this study suggests that it is difficult to balance performance and security, as shown in another study.[53] In other words, the more the security increases, the more memory or practice time is needed. Homomorphic encryption can be seen to possess a similar limitation, in that its relative accuracy is reduced,[73] or it is not secured enough.[74]

Despite the fact that technology is applied for security purposes, there are still cases where security weaknesses occur due to technical limitations.[54,59,62,75] Therefore, future technology research on DP and isomorphic encryption must focus on weaknesses or methods that sacrifice accuracy and security. In the case of relatively recent papers, many studies have been conducted to apply the above technologies. Through this, it can be seen that many studies are moving toward the stage where many studies can be conducted while protecting personal information.

Conclusion

Our study investigated the current development of a privacy protection technology that processes and analyzes big data produced in the field of healthcare and healthcare research. Although medical data lead to value creation in research, they contain sensitive personal and medical information and should be used with caution. For this reason, this study was designed to investigate privacy protection technologies and lend valuable insights for the progress of research by using medical data responsibly and safely.

Acknowledgments: We would like to thank all participants for their contribution to this study in all of the ways.

Contributorship: SL contributed to the conception and design of study. KR contributed to the acquisition of data. KR and HS contributed to the analysis and/or interpretation of data and drafting the manuscript. SL and J-YK contributed to revising the manuscript critically for important intellectual content. HS, KR, J-YK, and SL provided approval of the version of the manuscript to be published.

Declaration of conflicting interests: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethics Approval: This paper is a review paper that does not use personal information or personal data and, therefore, does not require consent or permission from the individual. Therefore, we hereby declare that this paper does not require IRB approval or permission.

Funding: This research was supported by the Basic Science Research Program through the National Research Foundation Korea (NRF). This work was supported by the Ministry of Education [NRF-2021R1I1A3044287]. This work was supported by the Gachon University research fund of 2024 [GCU-202404180001].

Guarantor: Suehyun Lee.

ORCID iD: Suehyun Lee https://orcid.org/0000-0003-0651-6481

References

1. Lee J, Kim H and Choi SJ. Do hospital data breaches affect health information technology investment? *Digit Health* 2024; 10: 1–11. DOI: 10.1177/20552076231224164.
2. Personal Data Protection Laws. Korean Law Information Center. https://www.law.go.kr/%EB%B2%95%EB%A0%B9/%EA%B0%9C%EC%9D%B8%EC%A0%95%EB%B3%B4%EB%B3%B4%ED%98%B8%EB%B2%95
3. General Data Protection Regulation. Intersoft consulting. https://gdpr-info.eu (accessed 5 August 2020).
4. Stone MA, Redsell SA, Ling JT, et al. Sharing patient data: competing demands of privacy, trust and research in primary care. *Br J Gen Pract* 2005; 55: 783–789.
5. Hasselgren A, Kralevska K, Gligoroski D, et al. Blockchain in healthcare and health sciences—a scoping review. *Int J Med Inf* 2020; 134: 104040.
6. Agbo C, Mahmoud Q and Eklund J. Blockchain technology in healthcare: a systematic review. *Healthcare* 2019; 7: 56.
7. Xu J, Glicksberg BS, Su C, et al. Federated learning for healthcare informatics. *J Healthc Inform Res* 2021; 5: 1–19.
8. Anwar RW, Abdullah T and Pastore F. Firewall best practices for securing smart healthcare environment: a review. *Appl Sci* 2021; 11: 9183.
9. Kruse CS, Smith B, Vanderlinden H, et al. Security techniques for the electronic health records. *J Med Syst* 2017; 41: 1–9. DOI: 10.1007/s10916-017-0778-4.
10. HIPAA Journal. Largest healthcare data breaches of 2021. https://www.hipaajournal.com/

largest-healthcare-data-breaches-of-2021/ (2014, accessed 30 December 2021)

11. OECD Digital Economy Papers. Emerging privacy enhancing technologies current regulatory and policy approaches, 8 March 2023. https://www.oecd-ilibrary.org/docserver/bf121be4-en.pdf?expires=1710308399&id=id&accname=guest&checksum=5C7BBBF2CB5191D130B816132A31FBF1 (accessed 13 March 2024).
12. Armando A. AVISPA: Automated Validation of Internet Security Protocols and Applications. Ercim News Online. https://www.ercim.eu/publication/Ercim_News/enw64/armando.html, 2005.
13. Nakamoto S. Bitcoin: a peer-to-peer electronic cash system. bitcoin.org (2009, accessed 31 Oct 2008).
14. Youtube. How does a blockchain work, simply explained. https://www.youtube.com/watch?v=SSo_EIwHSd4 (2005, accessed 14 November 2017).
15. Kalra S, Wen J, Cresswell JC, et al. Decentralized federated learning through proxy model sharing. *Nat Commun* 2023; 14: 2899. https://www.nature.com/articles/s41467-023-38569-4#:~:text=Federated%20learning%20(FL)%20is%20a
16. Rehman A, Abbas S, Khan MA, et al. A secure healthcare 5.0 system based on blockchain technology entangled with federated learning technique. *Comput Biol Med* 2022; 150: 106019.
17. Dou Q, So TY, Jiang M, et al. Author correction: federated deep learning for detecting COVID-19 lung abnormalities in CT: a privacy-preserving multinational validation study. *NPJ Digit Med* 2022; 5: 56. https://pubmed.ncbi.nlm.nih.gov/35462562/
18. Wu C, Wu F, Lyu L, et al. Communication-efficient federated learning via knowledge distillation. *Nat Commun* 2022; 13: 1–8.
19. Sadilek A, Liu L, Nguyen D, et al. Privacy-first health research with federated learning. *NPJ Digit Med* 2021; 4: 1–8. DOI: 10.1038/s41746-021-00489-2.
20. Lee D and Song M. MEXchange: a privacy-preserving blockchain-based framework for health information exchange using ring signature and stealth address. *IEEE Access* 2021; 9: 158122–39.
21. Mayer AH, Rodrigues VF, Costa Cd, et al. Fogchain: a fog computing architecture integrating blockchain and internet of things for personal health records. *IEEE Access* 2021; 9: 122723–37.
22. Pawar P, Parolia N, Shinde S, et al. Ehealthchain—a blockchain-based personal health information management system. *Ann Telecommun* 2022; 77: 33–45.
23. Abunadi I and Kumar R. BSF-EHR: blockchain security framework for electronic health records of patients. *Sensors* 2021; 21: 2865.
24. Dubovitskaya A, Baig F, Xu Z, et al. ACTION-EHR: patient-centric blockchain-based electronic health record data management for cancer care. *J Med Internet Res* 2020; 22: e13598.
25. Zhuang Y, Sheets LR, Chen Y-W, et al. A patient-centric health information exchange framework using blockchain technology. *IEEE J Biomed Health Inform* 2020; 24: 2169–2176.
26. Nguyen DC, Pathirana PN, Ding M, et al. Blockchain for secure EHRs sharing of mobile cloud based E-health systems. *IEEE Access* 2019; 7: 66792–66806.
27. Dwivedi A, Srivastava G, Dhar S, et al. A decentralized privacy-preserving healthcare blockchain for IoT. *Sensors* 2019; 19: 26.
28. Dubovitskaya A, Xu Z, Ryu S, et al. Secure and trustable electronic medical records sharing using blockchain. *AMIA Annu Symp Proc* 2018; 2017: 650–659.
29. Guo R, Shi H, Zhao Q, et al. Secure attribute-based signature scheme with multiple authorities for blockchain in electronic health records systems. *IEEE Access* 2018; 6: 11676–11686.
30. Haque EU, Shah A, Iqbal J, et al. A scalable blockchain based framework for efficient IoT data management using lightweight consensus. *Sci Rep* 2024; 14: 7841.
31. Li L, Fan Y, Tse M, et al. A review of applications in federated learning. *Comput Ind Eng* 2020; 149: 106854.
32. Nvidia Developer. Federated learning powered by NVIDIA Clara. https://developer.nvidia.com/blog/federated-learning-clara/ (2022, accessed 1 December 2019).
33. Nanda SK, Panda SK and Dash M. Medical supply chain integrated with blockchain and IoT to track the logistics of medical products. *Multimed Tools Appl* 2023 Mar 4: 1–23. DOI: 10.1007/s11042-023-14846-8.
34. Lakhan A, Mohammed MA, Nedoma J, et al. DRLBTS: deep reinforcement learning-aware blockchain-based healthcare system. *Sci Rep* 2023; 13: 4124. https://www.nature.com/articles/s41598-023-29170-2.
35. Kumar A, Singh AK, Ahmad I, et al. A novel decentralized blockchain architecture for the preservation of privacy and data security against cyberattacks in healthcare. *Sensors* 2022; 22: 5921.
36. Abid A, Cheikhrouhou S, Kallel S, et al. NovidChain: blockchain-based privacy-preserving platform for COVID-19 test/vaccine certificates. *Softw Pract Exper* 2022 Apr; 52: 841–867.
37. Javed IT, Alharbi F, Bellaj B, et al. Health-ID: a blockchain-based decentralized identity management for remote healthcare. *Healthcare* 2021; 9: 12.
38. Ma J, Zhang Q, Lou J, et al. Communication efficient federated generalized tensor factorization for collaborative health data analytics. In: *Proceedings of the web conference*, New York, April 2021, pp.171–182.
39. Budrionis A, Miara M, Miara P, et al. Benchmarking PySyft federated learning framework on MIMIC-III dataset. *IEEE Access* 2021; 9: 116869–78.
40. Liu JC, Goetz J, Sen S, et al. Learning from others without sacrificing privacy: simulation comparing centralized and federated machine learning on mobile health data. *JMIR Mhealth Uhealth* 2021; 9: e23728.
41. Dou Q, So TY, Jiang M, et al. Federated deep learning for detecting COVID-19 lung abnormalities in CT: a privacy-preserving multinational validation study. *NPJ Digit Med* 2021; 4: 1–11. DOI: 10.1038/s41746-021-00431-6.
42. Vaid A, Jaladanki SK, Xu J, et al. Federated learning of electronic health records to improve mortality prediction in hospitalized patients with COVID-19: machine learning approach. *JMIR Med Inform* 2021; 9: e24207.
43. Lee GH and Shin S-Y. Federated learning on clinical benchmark data: performance assessment. *J Med Internet Res* 2020; 22: e20891.
44. Deist TM, Dankers FJWM, Ojha P, et al. Distributed learning on 20,000+ lung cancer patients—the personal health train. *Radiother Oncol* 2020; 144: 189–200.

45. Huang L, Shea AL, Qian H, et al. Patient clustering improves efficiency of federated machine learning to predict mortality and hospital stay time using distributed electronic medical records. *J Biomed Inform* 2019; 99: 103291.
46. Lee J, Sun J, Wang F, et al. Privacy-preserving patient similarity learning in a federated environment: development and analysis. *JMIR Med Inform* 2018; 6: 20.
47. Brisimi TS, Chen R, Mela T, et al. Federated learning of predictive models from federated electronic health records. *Int J Med Inf* 2018; 112: 59–67.
48. Rafi TH, Noor FA, Hussain T, et al. Fairness and privacy preserving in federated learning: a survey. *Inf Fusion* 2024; 105: 102198.
49. Snips.Differential privacy for the rest of us. https://medium.com/snips-ai/differential-privacy-for-the-rest-of-us-665e053cec17 (2012, 30 July 2016).
50. Shaham S, Ghinita G, Ahuja R, et al. HTF: homogeneous tree framework for differentially-private release of large geospatial datasets with self-tuning structure height. *ACM Trans Spat Algorithms Syst* 2023; 9: 1–30.
51. Mueller TT, Paetzold JC, Prabhakar C, et al. Differentially private graph neural networks for whole-graph classification. *IEEE Trans Pattern Anal Mach Intell* 2023; 45: 7308–7318.
52. Adnan M, Kalra S, Cresswell JC, et al. Federated learning and differential privacy for medical image analysis. *Sci Rep* 2022; 12: 1–10. DOI: 10.1038/s41598-022-05539-7.
53. Ngo T, Nguyen DC, Pathirana PN, et al. Federated deep learning for the diagnosis of cerebellar ataxia: privacy preservation and auto-crafted feature extractor. *IEEE Trans Neural Syst Rehabil Eng* 2022; 30: 803–811.
54. Liu Y, Zhang Y, Huang W, et al. Privacy-aware real estate recommendation in cloud for elderly care based on historical consumption behaviors. *IEEE Access* 2021; 9: 41558–41565.
55. Chang Y, Fang C and Sun W. A blockchain-based federated learning method for smart healthcare. *Comput Intell Neurosci* 2021; 2021: 1–12.
56. Sadilek A, Liu L, Nguyen D, et al. Privacy-first health research with federated learning. *NPJ Digit Med* 2021; 4: 1–8. DOI: 10.1038/s41746-021-00489-2.
57. Lee EW, Xiong L, Hertzberg VS, et al. Privacy-preserving sequential pattern mining in distributed EHRs for predicting cardiovascular disease. *AMIA Jt Summits Transl Sci Proc* 2021; 2021: 384–393.
58. Ziller A, Usynin D, Braren R, et al. Medical imaging deep learning with differential privacy. *Sci Rep* 2021; 11: 1–8. DOI: 10.1038/s41598-021-93030-0.
59. Bonomi L, Jiang X and Ohno-Machado L. Protecting patient privacy in survival analyses. *J Am Med Inform Assoc* 2020; 27: 366–375.
60. Kim JW, Edemacu K and Jang B. MPPDS: multilevel privacy-preserving data sharing in a collaborative eHealth system. *IEEE Access* 2019; 7: 109910–23.
61. Kim H, Kim S-H, Hwang JY, et al. Efficient privacy-preserving machine learning for blockchain network. *IEEE Access* 2019; 7: 136481–95.
62. Ukil A, Jara AJ and Marin L. Data-driven automated cardiac health management with robust edge analytics and de-risking. *Sensors* 2019; 19: 2733.
63. Kim JW, Jang B and Yoo H. Privacy-preserving aggregation of personal health data streams. *PLOS One* 2018; 13: e0207639.
64. Phan N, Wu X and Dou D. Preserving differential privacy in convolutional deep belief networks. *Mach Learn* 2017; 106: 1681–1704.
65. PURE AI.. Homomorphic encryption makes slow but steady progress, 16 Jul 2020. https://pureai.com/articles/2020/07/13/homomorphic-encryption.aspx?m=1 (2018, accessed 11 January 2022).
66. Ali A, Al-rimy BAS, Alsubaei FS, et al. Healthlock: blockchain-based privacy preservation using homomorphic encryption in Internet of Things healthcare applications. *Sensors* 2023; 23: 6762. https://www.mdpi.com/1424-8220/23/15/6762/pdf.
67. ProQuest. Link to external site this link will open in a new window, Link to external site this link will open in a new window, Link to external site this link will open in a new window. Secure human action recognition by encrypted neural network inference. https://www.proquest.com/docview/2702359477?parentSessionId=Rzke7oO4LJrwB4XTJQSvep7oY3CCtU8WxXze7V7JyCo%3D&pq-origsite=summon&accountid=14650 (2022, accessed 27 April 2023).
68. Ali A, Almaiah MA, Hajjej F, et al. An industrial IoT-based blockchain-enabled secure searchable encryption approach for healthcare systems using neural network. *Sensors* 2022; 22: 572. https://doaj.org/article/bb5b9762660741f388cd508199895288.
69. Ali A, Pasha MF, Ali J, et al. Deep learning based homomorphic secure search-able encryption for keyword search in blockchain healthcare system: a novel approach to cryptography. *Sensors* 2022; 22: 28.
70. Froelicher D, Troncoso-Pastoriza JR, Raisaro JL, et al. Truly privacy-preserving federated analytics for precision medicine with multiparty homomorphic encryption. *Nat Commun* 2021; 12: 1–10. https://www.nature.com/articles/s41467-021-25972-y.pdf.
71. Kim M, Harmanci AO, Bossuat JP, et al. Ultrafast homomorphic encryption models enable secure outsourcing of genotype imputation. *Cell Syst* 2021; 12(11): 1108–1120.e4. https://www.sciencedirect.com/science/article/pii/S240547122100288X
72. Farid F, Elkhodr M, Sabrina F, et al. A smart biometric identity management framework for personalised IoT and cloud computing-based healthcare services. *Sensors* 2021; 21: 52.
73. Froelicher D, Troncoso-Pastoriza JR, Raisaro JL, et al. Truly privacy-preserving federated analytics for precision medicine with multiparty homomorphic encryption. *Nat Commun* 2021; 12: 1–10. DOI: 10.1038/s41467-021-25972-y.
74. Tahir S, Tahir H, Sajjad A, et al. Privacy-preserving COVID-19 contact tracing using blockchain. *J Commun Netw* 2021; 23: 360–373.
75. Kim M, Harmanci AO, Bossuat J-P, et al. Ultrafast homomorphic encryption models enable secure outsourcing of genotype imputation. *Cell Syst* 2021; 12: 1108–1120.e4.
76. Farid F, Elkhodr M, Sabrina F, et al. A smart biometric identity management framework for personalised IoT and cloud computing-based healthcare services. *Sensors* 2021; 21: 52.
77. Vizitiu A, Niţă CI, Puiu A, et al. Applying deep neural networks over homomorphic encrypted medical data. *Comput Math Methods Med* 2020; 2020: 1–26.

78. Alabdulatif A, Khalil I, Yi X, et al. Secure edge of things for smart healthcare surveillance framework. *IEEE Access* 2019; 7: 31010–31021.
79. Sadat MN, Al Aziz MM, Mohammed N, et al. SAFETY: secure GWAS in federated environment through a hybrid solution. *IEEE/ACM Trans Comput Biol Bioinf* 2019; 16: 93–102.
80. Raisaro JL, Klann JG, Wagholikar KB, et al. Feasibility of homomorphic encryption for sharing I2B2 aggregate-level data in the cloud. *AMIA Jt Summits Transl Sci Proc* 2018; 2017: 176–185.