Can Our Health Data Stay Private? A Review and Future Directions for IS Research on Privacy-Preserving AI in Healthcare

Aycan Aslan
University of Goettingen, aycan.aslan@uni-goettingen.de

Maike Greve
University of Goettingen, maike.greve@uni-goettingen.de

Till Ole Diesterhöft
University of Goettingen, tillole.diesterhoeft@uni-goettingen.de

Lutz M. Kolbe
University of Goettingen, lkolbe@uni-goettingen.de

Can Our Health Data Stay Private? A Review and Future Directions for IS Research on Privacy-Preserving AI in Healthcare

Aycan Aslan[1], Maike Greve[1], Till Ole Diesterhöft[1], and Lutz M. Kolbe[1]

[1] University of Goettingen, Chair of Information Management, Goettingen, Germany
{aycan.aslan,maike.greve,tillole.diesterhoeft,lkolbe}@uni-goettingen.de

Abstract. The generation of data has become one of the main drivers of modern healthcare. Like other industries, we see that the total amount of healthcare data is growing and in diversity. Thus, Artificial Intelligence (AI) is being used increasingly as a tool to turn this body of healthcare data into real value. But with AI and big data comes big risk, especially in terms of data privacy. Privacy-preserving AI techniques are gaining in popularity to prevent patient privacy compromises while utilizing the potentials offered by AI. However, there is no clear understanding of the current research space of applying such privacy-preserving techniques in healthcare. This paper aims to provide an understanding of these techniques and investigates the emerging research field of privacy-preserving AI and its use in healthcare by reviewing the current multidisciplinary research to synthesize knowledge and derive future research directions in this regard.

Keywords: Data Privacy, Privacy-Preserving AI, Private AI, Healthcare

1 Introduction

The 21st century is the century of big data affecting all aspects of human life, including healthcare. Historically, the healthcare industry always generated a large amount of data due to complex regulatory requirements such as continuous record keeping. In the last couple of years, we experienced the digitization of this traditionally generated data, for example, in the forms of Electronic Health Records (EHR) [1]. Additional to this development of digitizing already available data, various new data sources emerged; for example, the improvement of medical technology created various new data sources, such as genomics data or biometric sensor readings [2]. Further, with new emerging data sources, the body of healthcare data is getting significantly more heterogeneous. Nowadays, healthcare data types are ranging from unstructured data (e.g., clinical notes and pathology reports) to structured and semi-structured data (e.g., genomic data or medical images) [3, 4].

Due to this increase in volume and variety of healthcare data, traditional data management tools are inadequate, while Artificial Intelligence (AI) is gaining popularity. For example, in order to realize 'precision medicine', the tailored treatment of each patient, a body of data sources ranging from Genomics, Biomarkers, EHRs, and wearables have to be analyzed [5–7]. Traditional tools, such as Online Analytical

17th International Conference on Wirtschaftsinformatik,
February 2022, Nürnberg, Germany

Processing (OLAP), are not designed for such tasks, especially when predictive analysis is needed [8, 9]. Here, AI models seem to be a great tool to take advantage of the ever-growing, diverse set of healthcare data [10, 11]. Especially, more sophisticated techniques, such as Deep Learning models, can leverage large healthcare datasets to deliver accurate and reliable results [12]. Recently published research studies point out the potential of AI models in healthcare, for example, in the form of medical image analysis [13] or the use of Natural Language Processing to extract information from clinical notes [14].

While the opportunities for the use of AI in healthcare are promising, one cannot ignore arising dangers for implementing discussed techniques. No matter how useful for healthcare, AI can only be used if data privacy issues are addressed and resolved due to the high need for data [15, 16]. This applies particularly to sensitive individual health data since it represents personally identifiable information [14].

To ensure privacy for sensitive health data while capitalizing on the potentials offered by AI, the field of privacy-preserving AI (PP-AI) is gaining in popularity. Hereby, the term PP-AI represents an umbrella term for different techniques [17]. The overall goal being, to ensure data privacy while still exploiting the potentials offered by AI. In comparison to traditional systems, PP-AI also addresses unique privacy needs of AI-based systems, such as training data privacy, model weight privacy or the possible memorization of data points in the training process [17, 18]. For instance, prior research has shown that so called model inversion attacks can be performed to reconstruct training data only from model parameters [18]. Despite these unique challenges, there is limited understanding of the current research space of PP-AI, especially in terms of using these techniques in healthcare since the field is still emerging and of multidisciplinary nature. Various disciplines contribute to the field, such as Computer Science, Mathematics and Encryption, Healthcare, and Information Systems (IS). Here, all fields have different approaches and therefore make different contributions. Nevertheless, moving from a multidisciplinary character to truly interdisciplinary work requires input from all involvement fields. This also applies to the IS community, which is still beginning to understand and discuss PP-AI in the IS context.

Therefore, the objectives of this work are to introduce the field of PP-AI and structure and assess the current status quo of multidisciplinary research on PP-AI in healthcare to synthesize knowledge and derive future research directions in this regard. This paper aims to answer the following three research questions (RQ) to address these objectives:

***RQ1**: Which PP-AI techniques exist, and how do they differ in suitability for the healthcare context?*

***RQ2**: What is the current research space, regarding the theories, methods, and domains, of PP-AI literature in healthcare?*

***RQ3:** What are future directions for IS scholars to advance PP-AI in healthcare?*

Our work follows a threefold procedural approach to answer these questions: First, we *introduce* the most common PP-AI techniques, provide a high-level explanation, and

discuss their suitability in the healthcare context (RQ1). Second, we *review* and present a comprehensive overview of how PP-AI techniques are discussed in healthcare. Based on this review, we categorize the use of PP-AI techniques in healthcare for healthcare domains and data types & sources (RQ2). Finally, we discuss how IS scholars can use this understanding and categorization of PP-AI techniques in healthcare to *enrich further* the PP-AI literature from the IS perspective (RQ3).

2 State of the Art: What is Privacy-Preserving AI?

To answer RQ1, in this section we will explain and analyze the field of PP-AI. The term PP-AI is an umbrella term and is used for many different techniques and methods [17]. The overall goal is to ensure data privacy while still exploiting the potential offered by AI methods [17]. Here, it is important to note that PP-AI is not a clearly defined field yet. Some researchers incorporate and discuss a broad range of encryption schemes (e.g., Garbled Circuits) [19, 20], while others include secure hardware implementations [17]. Nevertheless, the techniques discussed and used in most studies and therefore represent the contemporary consensus are: 1) Federated Learning, 2) Differential Privacy, 3) Homomorphic encryption, and 4) Secure multi-party computation [10, 17, 19–21]. While Federated Learning can be seen as an AI technique itself, Differential Privacy, Homomorphic encryption, and Secure multi-party computation are privacy and encryption schemes. However, in the context of PP-AI, the privacy enhancing characteristics of these techniques can be embedded or combined with AI. Therefore, they are included in this classification as PP-AI and when discussed in the context of this work, we refer to these privacy and encryption schemes always as part or embedded in AI-systems. Hereafter, these most common PP-AI techniques will be explained in general and in the context of healthcare.

Federated Learning. Federated Learning (FL) is a paradigm belonging to the class of distributed systems. Instead of transferring sensitive data from sites to sites or servers, FL consists of participating parties that train a model collaboratively without exchanging the underlying dataset [22]. In practice, copies of a Machine Learning algorithm are distributed to the sites and devices where the data is kept [17]. These remote devices are also called nodes, which perform the training iterations locally and only return the results of the computation (e.g., updated neural network weights) to the central repository. Therefore, the model benefits from the knowledge accumulated across all participating institutions without moving the data.

In the context of healthcare, the distributed nature of FL opens the possibility for a wide range of applications [12, 21]. For example, FL can make it possible that several healthcare institutions (e.g., local hospitals) can collaboratively work on training an algorithm without sharing their patient data [22]. Hence, the data is not moved beyond the firewalls of the respective healthcare institution.

Differential Privacy. Differential Privacy (DP) is the systematic randomized modification of a dataset or algorithm to reduce information about single individuals

while retaining the capability of statistical reasoning about the dataset [25]. Thus, outside observers cannot infer whether a specific individual was used for the result of an analysis or not [17]. The robustness of DP lies in its rigorous mathematical proof. Therefore, it can resist various forms of adversary attacks with the maximum background knowledge of the attacker. The conducted modification makes it hard for an adversary to tell which behavioral aspects of the given model come from randomness and which from the actual training data. The modification can be simple random shuffling or more sophisticated forms, such as noise adding (e.g., Laplacian, or Gaussian noise). There are different forms of implementing or combining such modifications with AI. For example, differentially private stochastic gradient descent (DPSGD), which is used for the differentially private learning of Deep Learning and applies noise to the gradients of the model [26]

DP is gaining in popularity for healthcare due to its mathematical rigor and possibility of being easily combined with other privacy methods [25]. With this, DP is explored since it can ensure privacy at the source of the data, which puts the data owner in control. This might be useful for mHealth settings, where noise can be added to the data of wearable healthcare devices before being sent to a central server.

Homomorphic encryption. Homomorphic encryption (HE) is an encryption scheme that provides rigorous guarantees while enabling operations over encrypted data. Simply put, HE allows computation on encrypted data as if it was unencrypted. Hence, computations on the encrypted data would yield a result, and once decrypted, match the result of the non-encrypted computation [17]. When applied to Machine Learning classification tasks, it is possible to realize secure classification over encrypted data. HE offers the possibility to entrust a third party (e.g., an off-site cloud computing server) with the encrypted version of the dataset since the data owner has the mathematical certainty that the third party cannot decrypt the dataset itself nor the result of a given analysis. In the context of AI-systems, the encryption characteristics of HE can be used to, for example encrypt the gradients of a Deep Learning model, hence increasing the privacy [27].

These characteristics are also advantageous in healthcare. In comparison to DP, for example, there is no trade-off between privacy and utility. This is beneficial in healthcare, where sacrificing even some model performance can lead to major negative effects in terms of human lives [28]. Further, the openness for distributed systems offers advantages for healthcare institutions that want to collaborate with each other. In such cases, HE can enable the secure aggregation of encrypted algorithm updates between the institutions [29].

Secure multi-party computation. Secure multi-party computation (SMPC) is an encryption scheme that includes multiple parties, which form a governance model built on secret-sharing. SMPC ensures that a function can only be computed when all parties jointly provide their needed inputs while the content of the inputs stays private [17]. In practice, the data is being divided into data shares among the participating parties. Processing is then done based on the encrypted data shares. Hence, no single party can retrieve the total dataset on their own. Therefore, the computation results can be

announced without any single party having seen the undivided data. Recovering the original data is only possible by the consensus of all participating parties. So, without the permission of all parties, the data shares stay encrypted and therefore unusable for adversaries, yielding a shared governance model. In combination with AI-systems, these characteristics can be used to facilitate training of AI models on data sets owned by different parties [30].

These secret-sharing characteristics are also utilized in healthcare. SMPC is gaining popularity in healthcare, intending to enhance private collaboration among healthcare institutions [17]. For example, when multiple hospitals want to execute a collaborative analysis, where the goal is to join respective datasets and gain knowledge from more extensive and diverse data. Here, SMPC reduces the risk of inadvertent or malicious leaks and facilitates collaboration between hospitals. This solution is especially advantageous for the collaboration between healthcare institutions with no prior relationship or a low-trust environment [17].

3 Methodology

To understand the current state of PP-AI use in healthcare and answer the research questions, we conducted a systematic literature review suggested by [31]. Hence, the procedure follows three phases: literature search, literature evaluation and selection, and literature analysis and synthesis.

3.1 Literature Search

Due to the stated multidisciplinary nature of the topic, databases that cover multiple academic fields are included. Hence, we searched the following databases: ProQuest/EBSCO Host/Science Direct (for Basket of Eight), AIS eLibrary, ACM Digital Library, IEEE Explore, and PubMed. The language of the articles is limited to English, and only peer-reviewed articles are included to validate the quality of research. The final search string was applied on 'title' and 'abstract'. Besides journal articles, we include conference papers in recognition of the novelty of the topic.

To reflect the multidisciplinary character of the topic, we split our keyword search into three parts. The first two represent the 'privacy' and 'AI' nature, while the last part represents the 'healthcare' domain. This resulted in the following search query: *(privacy-preserving OR privacy OR private OR privacy-protection OR privacy-aware) AND (AI OR 'artificial intelligence' OR 'machine learning' OR 'deep learning') AND (healthcare OR health OR e-health OR medical OR medicine OR hospital).*

3.2 Literature Evaluation and Selection

By following the steps described, our search resulted in an initial set of 1097 contributions. To identify relevant articles, we followed the evaluation process depicted in Figure 1. First, the initial set was reduced by applying the search string only to the field's 'title' and 'abstract'. This reduced the set to 432 studies. Second, we scanned

the titles and abstracts of the articles for a content fit. This phase reduced the set to 194 articles. Third, we conducted a full-text analysis, where we checked the remaining articles based on the inclusion criteria defined.

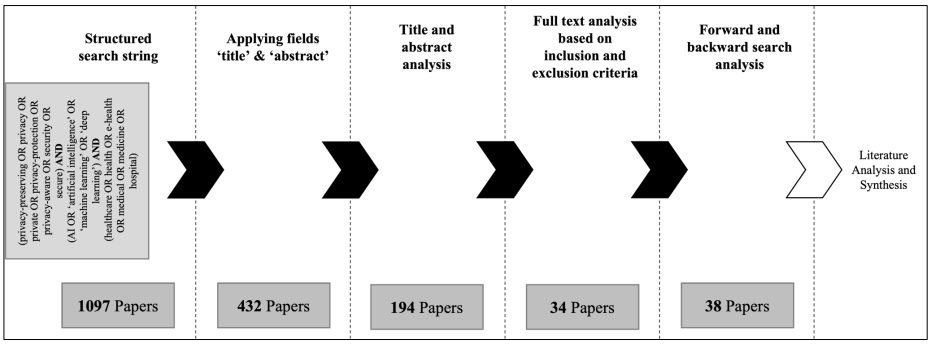

Figure 1. The conducted systematic literature procedure

For the selection of the encountered articles, we established the following inclusion criteria, whereby all three must be fulfilled:

1. *Articles must discuss or apply a privacy-preserving technique.*
2. *Articles must discuss or apply the technique specifically in healthcare.*
3. *Articles must discuss or apply AI-based techniques.*

Those criteria reduced the number of articles to 34. Finally, a forward and backward search lead to four additional articles, which results in 38 relevant contributions in total.

3.3 Literature Analysis and Synthesis

To examine the current research space of PP-AI in healthcare, the 38 studies were classified with the help of the framework proposed by [32]. Applying the conceptualization, we distinguish the domains 'Theory' type, 'Methods' used and the 'Context' of the analyzed studies.

In order to analyze the theory, we adapt the 'Taxonomy of Theory Types in Information Systems Research' by [33]. The taxonomy classifies five theory types: I) Analysis, II) Explanation, III) Prediction, IV) Explanation and prediction (EP), and V) Design and action. Here, Analysis describes *what* is, Explanation *what* is, *how*, *why, when,* and *where*, Prediction *what is* and *what will* be, EP *what is*, *how*, *why*, *when*, *where*, and *what will* be, and Design and action *how* to do something.

In terms of the Methods used, we analyze two levels. The first level is 'Data generation,' which describes the methods of data production and generation [32]. Here, we will analyze the healthcare data types & sources which are used by the 38 studies. For this goal, we utilized the classification by [34]. They analyzed data types & sources in healthcare and identified four groups for data types & sources: 1) Clinical data, 2) Patient behavior and sentiment data, 3) Administrative data and cost data, and 4) R&D data. The first group of 'Clinical data' includes data that derives from patients in clinics, such as EHRs or medical images. Next, the group 'Patient behavior and sentiment data'

includes all data that is collected in a distributed manner, such as wearables and social sites. The data type group 'Administrative data and cost data' is data that describes costs, bills, reimbursement categories, and other patient characteristics. Last, the group 'R&D data' includes data that derived from common R&D activities such as genomic data. It is important to note, that based on the framework of [32], in terms of 'Data generation' we only focus on the source of data. The second level is 'Data analysis' and describes the type of data analysis conducted. Here, our analysis will check which of the PP-AI techniques introduced in Chapter 2 are used by the respective study.

For context, we decide to focus on the investigate context, meaning to understand where and from whom data is collected. In our analysis, we will look at the specific healthcare domain in which the respective PP-AI technique was used. For this, we adopted the classification proposed by [19]. They analyzed the general use of AI in healthcare and created three major domain groups. These are 1) Diagnosis & Prognosis, 2) Treatment, and 3) Clinical workflow and management. Here, 'Diagnosis & Prognosis' describes the extraction of clinical features to diagnose diseases and abnormalities. Further, this domain group also describes the process of the expected development of a disease. Next, the domain 'Treatment' characterizes activities for the treatment of patients after the diagnosis of diseases. Hence, the generation of documents that are used in the treatment are included as well. Lastly, the domain 'Clinical workflow and management' concerns the clinic's administrative processes and operational tasks.

4 Analysis of the Research Space

To answer RQ2, this section discusses the results of the conducted literature analysis and synthesis, as depicted in Table 1. It gives an overview of the 38 analyzed studies in terms of theory, method, and context.

Table 1. The synthesis of our analysis

	Theory					**Method: Data generation**				**Method: Data analysis**				**Investigative context: Healthcare domain**		
Paper	I. Analysis	II. Explanation	III. Prediction	IV. Explanation and prediction	V. Design and action	Clinical data	Patient behav. & sentiment data	Administrative and cost data	R&D data	Federated Learning	Differential Privacy	Homomorphic Encryption	Secure Multi-Party Computation	Diagnosis & Prognosis	Treatment	Clinical workflow & mgmt.
Zhang et al. 2021 [35]					X	X					X			X		
Jiang et al. 2019 [29]					X	X						X		X		
Chen et al. 2019 [36]					X	X					X			X		
Suriyakumar et al. 2021 [37]					X			X			X			X		X
Can and Ersoy 2021 [38]					X		X			X				X		
Cheng et al. 2021 [39]					X	X					X			X		

Chen et al. 2020 [40]					X				X				X		X	
Yuan et al. 2019 [41]					X	X					X			X		
Yasumura et al. 2019 [42]					X	X						X		X		
Vignesh et al. 2019 [43]					X	X							X	X		
Hakak et al. 2020 [44]					X		X			X				X		
Sun et al. 2019 [45]					X	X					X			X		
Müftüoglu et al. 2020 [46]					X	X					X			X		
Kim et al. 2019 [47]					X	X					X			X		
Guo et al. 2020 [48]				X		X				X	X			X		
Leboe-McGowan et al. 2020 [49]					X				X				X	X		
Ying et al. 2020 [50]					X	X							X		X	
Jarin and Eshete et al. 2021 [51]					X			X			X		X			X
Kumar et al. 2021 [52]					X	X				X				X		
Liu et al. 2021 [24]					X	X				X					X	
Ying et al. 2021 [53]					X	X							X		X	
Zhang et al. 2020 [54]					X	X						X		X		
Imtiaz et al. 2020 [23]					X		X			X	X				X	
Ma et al. 2020 [55]					X	X						X		X		
Rahman et al. 2020 [56]					X		X			X		X		X		
Thwal et al. 2021 [57]					X	X				X				X		
Kwabena et al. 2019 [58]					X	X						X		X		
Wu et al. 2020 [59]					X		X			X					X	
Zhang et al. 2021 [60]					X	X				X				X		
Zhang et al. 2020 [61]					X	X				X				X		
Zhao et al. 2020 [62]					X		X			X					X	
Li and Qin 2018 [4]					X	X					X			X		
Pfitzner et al. 2021 [63]		X					X			X					X	
Brisimi et al. 2018 [64]				X		X				X				X		
Passerat-Palmbach et al. 2020 [22]				X				X		X						X
Rieke al. 2020 [65]				X				X		X					X	X
Qayyum et al. 2021 [19]				X		X				X	X	X	X	X		X
Hu et al. 2019 [25]				X							X			X		
Sum	0	1	0	6	31	25	7	4	2	17	14	7	7	28	9	5

4.1 Theory

As noted, to classify the theory type of the analyzed studies, we adopted the taxonomy by [33]. The results can be seen in Figure 2a.

We note a clear trend for applying *design and action* theory since 31 out of the analyzed 38 studies can be classified as such. Hence, these 31 studies describe PP-AI techniques in healthcare and propose explicit frameworks on how to build an artifact. The second largest group can be classified as *explanation and prediction*. These six studies explain the application of PP-AI techniques in healthcare and give predictions based on their explanation and analysis [19, 22, 25, 48, 64, 65]. Only one study can be classified as the theory type *explanation*. This study explains applying FL in healthcare but does not aim to predict future developments [63]. None of the 38 reviewed studies can be categorized as *analysis* or *prediction*.

4.2 Method

To analyze the methods deployed, we investigated data generation and data analysis. In terms of data generation, we categorized the reviewed studies into four groups: Clinical data, Patient behavior and sentiment data, administrative data and cost data, and R&D data. To analyze the deployed data analysis technique, we classified the reviewed studies based on the PP-AI techniques presented in Chapter 2.

As depicted in Figure 2b, most of the reviewed studies used *clinical data* in their analysis. Among these 25 studies, six studies used disease attribute records for breast cancer [29, 36, 42, 47, 48, 51], four studies used diabetes patient records [45, 50, 53, 58], three studies used Electronic Health Records [4, 36, 51], and one study utilized an ECG dataset [61]. In terms of medical images, studies used X-Rays [35, 39, 41, 43, 46, 60], CT scans [52, 60], SPECT scans of the heart [29], and dermoscopic images [39, 54]. The second most popular data type group is *patient behavior and sentiment data*. Here, three studies used smartphones & smartphone apps [23, 56, 59], two studies utilized smartwatches [38, 62], and one studies smart bands [38]. In our analysis, only four studies worked with *administrative and cost data*. Exemplary, these utilized ICU data [37], or critical care databases [51]. Lastly, only two studies worked with *R&D data*. Of those, one study used HIV sequence data [40] and the other used gene expression data [49].

As shown in Figure 2c, the most popular PP-AI technique is FL, with 17 reviewed studies applying it. For the studies, which stated the specific details regarding the FL setup, the most popular algorithm was the Federated Averaging algorithm, used by [23, 24, 38, 44, 56, 59]. The second most used PP-AI technique is DP. DP was used by 14 of the reviewed studies. Here, the most popular mechanism for adding noise is the Gaussian mechanism, implemented by [23, 35, 39, 41, 45, 47, 48]. Compared to FL and DP, we find that HE and SMPC are much less used in the reviewed studies. In our analysis, HE was only applied by seven studies [19, 29, 42, 54–56, 58], while SMPC was applied by seven studies as well [19, 40, 43, 49–51, 53].

4.3 Context

Lastly, we classified the investigative context of the reviewed studies by classifying them based on the healthcare domain they discussed in their study.

Looking at the distribution over the three defined healthcare domains in Figure 2d, it becomes evident that most studies focus on *Diagnosis & Prognosis*, followed by *Treatment* and *Clinical Workflow & Management*. In terms of *Diagnosis & Prognosis*, many studies discuss the detection and classification of diseases, e.g., breast cancer [29, 36, 42, 47–49, 51]. Other studies discuss the diagnosis of respiratory diseases such as pneumonia [41, 46, 52, 57, 60], tuberculosis [57], bronchial asthma [57], and pleural effusions [39]. Besides the detection, studies also discuss the progression of diseases, e.g., the progression of diabetes [58]. Additional to detecting and predicting the progression of diseases, studies also look at detecting abnormalities. These include abnormalities such as cardiac arrhythmia [36, 55, 61] and skin tone [39]. In terms of the treatment of patients, nine relevant studies were detected. These discuss, for example the efficacy of antiviral drugs [40], depression treatment [24], optimal insulin dosage [53], or the forecast of dietary habits and health monitoring [23, 59, 62]. As noted, the least number of studies focus on applying PP-AI techniques in *Clinical Workflow and Management*. Among the five studies discussing the application for *Clinical Workflow and Management*, studies for example look at ICU mortality prediction [37], LOS prediction in hospitals [37, 51], or predicting the intervention onset for vasopressor administration [37].

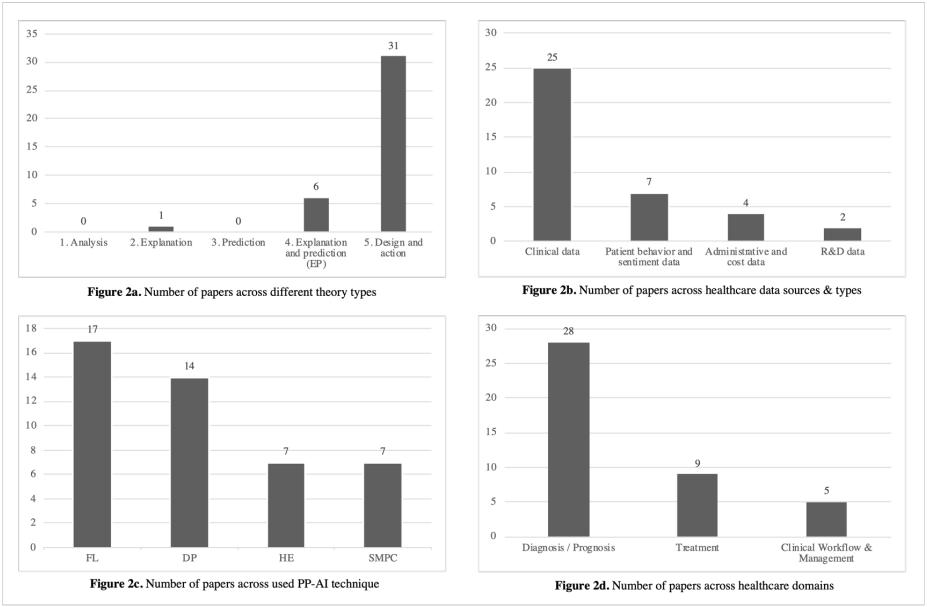

Figure 2a. Number of papers across different theory types

Figure 2b. Number of papers across healthcare data sources & types

Figure 2c. Number of papers across used PP-AI technique

Figure 2d. Number of papers across healthcare domains

Figure 2. Quantitative analysis of the reviewed literature

5 Future Directions for IS Research

Based on our understanding of the PP-AI techniques and the conducted research of current use in healthcare, to answer RQ3 we now identify three key findings which imply research gaps and future directions for IS research on PP-AI use in healthcare, depicted in Table 2.

First, our findings reveal that most of the studies reviewed are published in Computer Science outlets and focus on creating artifacts; hence there is an uneven distribution of outlets and scholars working on PP-AI in healthcare and consequently the research methods applied. Here we note that there is particularly little research done by the IS community since only one of 38 studies was published in an IS affiliated outlet [4]. Hence, IS research should participate more actively in researching PP-AI techniques in general and their use in healthcare. On the one hand, IS researchers are challenged to conduct more research in this field, since the IS perspective can be crucial in further developing and establishing such techniques in healthcare. Primarily by analyzing the field more holistically and the interdependence between PP-AI techniques, healthcare stakeholders, and healthcare organizations. To achieve this goal, familiar research methods to IS researchers (e.g., Survey research) can be deployed to address questions such as the understanding and acceptance of patients of new PP-AI techniques or the implications for the increase in collaboration between healthcare organizations. On the other hand, IS outlets should be more receptive and encourage this research, for example by offering special issues which focus on privacy concerns of AI methods.

This could sign the topics relevance to the IS community and reinforce IS researchers to engage with the topic.

Second, we found that the overwhelming majority of studies discuss the use of PP-AI techniques in healthcare in isolation. This is surprising when considering that prior PP-AI research already established that considered alone, all PP-AI techniques have major limitations, both technical and regarding applicability to real-life scenarios [17, 19, 65]. Therefore, a combination of techniques is recommended to utilize the orthogonal privacy protections offered by the different techniques. For example, FL mainly resolves data governance and ownership [63], while HE conducts encryption of the data itself [29]. Here, future IS research could help to advance PP-AI techniques in terms of practical applications by, for example, conducting case study research to test PP-AI techniques in a more natural setting. Such studies would generate in-depth insights about the benefits and limitations of combining specific PP-AI techniques for more realistic healthcare applications. This understanding could bridge the gap between the academic discussion to actual application of PP-AI techniques in healthcare.

Lastly, our findings show that most studies use clinical data, rely on FL or DP, and apply their artifact in Diagnosis & Prognosis, yielding a disproportionate distribution in terms of user data types & sources, applied PP-AI techniques, and healthcare domain. Future research should examine the potentials and limitations of applying PP-AI techniques in less studied healthcare fields. For this purpose, the IS community is especially well-suited due to its vast history in analyzing and discussing the use of Information Technology in various healthcare fields [66–70]. Through the healthcare industry's progress and ongoing digital transformation, we can expect more healthcare data to be generated and more healthcare processes getting digitized. Here, the IS community can leverage prior research in terms of digital transformation to analyze and understand privacy problems in newly digitized healthcare fields and how PP-AI techniques might pose solutions to those problems [1, 71, 72].

Table 2. Findings, research gaps, and future directions for IS research

Reference in our analysis	**Findings**	**Research gaps**	**Future directions for IS research**
Theory	Most of the studies reviewed are from the Computer Science community and focus on creating artifacts.	There is an uneven distribution of outlets and scholars, and consequently, research methods applied.	IS research should participate more actively in researching PP-AI in general and its use in healthcare.
Method (Data analysis)	The overwhelming majority of studies discuss the respective PP-AI techniques in isolation.	Considered alone, the discussed PP-AI techniques still have some major limitations.	Future IS studies should address the need to understand the effects of combining different PP-AI techniques in more realistic settings.

Method (Data generation), Method (Data analysis), Investigative context	Most studies use clinical data, rely on Federated Learning, or Differential Privacy, and apply their artifact in the context of Diagnosis & Prognosis.	We note a disproportionate distribution in terms of used data types & sources, applied PP-AI techniques, and healthcare domains.	Future IS research should examine PP-AI techniques in less-studied healthcare fields by leveraging prior knowledge in digital healthcare.

6 Concluding Remarks

This paper aimed to provide a comprehensive explanation of the most important PP-AI techniques, their suitability in healthcare, and analyze the current research space of studies on PP-AI use in healthcare. We showed that PP-AI techniques have enormous potential for application in healthcare but vary in suitability for different healthcare fields. Further, the current research space has been analyzed by looking into the theory types applied, data generation methods, data analysis techniques, and investigative context. Based on these findings, we have established that the IS community is just at the beginning in terms of PP-AI research. Hence, we derived future directions for IS scholars and showed which aspect they can add to the growing multidisciplinary research on PP-AI. Nevertheless, we must note two important limitations to our work. First, we have strongly focused on the potentials of PP-AI techniques, but not the technical limitations that exist. A complete picture of the applicability of PP-AI techniques would require a more in-depth analysis of the limitations. A more thorough understanding of the technical limitations would also pave the way for implementing PP-AI techniques in actual healthcare practice. Second, we looked at the future directions which can be derived for the IS community, while ignoring the learnings of this work for other domains. As stated, PP-AI is a multidisciplinary field, so the findings generated in this work opens a range of research avenues for multiple fields. To establish real advancement in the field of PP-AI, additional domains besides the IS community, should build on the findings generated in this work. These limitations should be considered and extended in future (multidisciplinary) work.

In conclusion, we hope that this paper provides a comprehensive understanding of the current state of PP-AI research in healthcare which stimulates future research and motivates scholars to engage and collaborate in this emerging field to enable the private use of healthcare data for AI methods.

References

1. Kohli, R., Tan, S.S.L.: Electronic health records: How can is researchers contribute to transforming healthcare? MIS Q. Manag. Inf. Syst. 40, 553–573 (2016).
2. Raghupathi, W., Raghupathi, V.: Big data analytics in healthcare: promise and potential. Heal. Inf. Sci. Syst. 2, 1–10 (2014).
3. Andreu-Perez, J., Poon, C.C.Y., Merrifield, R.D., Wong, S.T.C., Yang, G.Z.: Big Data for Health. IEEE J. Biomed. Heal. Informatics. 19, 1193–1208 (2015).
4. Li, X., Qin, J.: Protecting Privacy When Releasing Search Results from Medical Document Data. Proc. 51st Hawaii Int. Conf. Syst. Sci. 3770–3778 (2018).
5. Ginsburg, G.S., Phillips, K.A.: Precision medicine: From science to value. Health Aff. 37, 694–701 (2018).
6. Rumsfeld, J.S., Joynt, K.E., Maddox, T.M.: Big data analytics to improve cardiovascular care: Promise and challenges. Nat. Rev. Cardiol. 13, 350–359 (2016).
7. Filipp, F. V.: Opportunities for Artificial Intelligence in Advancing Precision Medicine. Curr. Genet. Med. Rep. 7, 208–213 (2019).
8. Gordon, B.D., Asplin, B.R.: Using online analytical processing to manage emergency department operations. Acad. Emerg. Med. 11, 1206–1212 (2004).
9. Dash, S., Shakyawar, S.K., Sharma, M., Kaushik, S.: Big data in healthcare: management, analysis and future prospects. J. Big Data. 6, (2019).
10. Torkzadehmahani, R., Nasirigerdeh, R., Blumenthal, D.B., Kacprowski, T., List, M., Matschinske, J., Späth, J., Wenke, N.K., Bihari, B., Frisch, T., Hartebrodt, A., Hausschild, A.-C., Heider, D., Holzinger, A., Hötzendorfer, W., Kastelitz, M., Mayer, R., Nogales, C., Pustozerova, A., Röttger, R., Schmidt, H.H.H.W., Schwalber, A., Tschohl, C., Wohner, A., Baumbach, J.: Privacy-preserving Artificial Intelligence Techniques in Biomedicine. (2020).
11. Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., Wang, Y.: Artificial intelligence in healthcare: Past, present and future. Stroke Vasc. Neurol. 2, 230–243 (2017).
12. Raja, R., Mukherjee, I., Sarkar, B.K., Ali, S.: A Systematic Review of Healthcare Big Data. Sci. Program. 2020, (2020).
13. Mehta, N., Pandit, A.: Concurrence of big data analytics and healthcare: A systematic review. Int. J. Med. Inform. 114, 57–65 (2018).
14. Roski, J., Bo-Linn, G.W., Andrews, T.A.: Creating value in health care through big data: Opportunities and policy implications. Health Aff. 33, 1115–1122 (2014).
15. Abouelmehdi, K., Beni-Hessane, A., Khaloufi, H.: Big healthcare data: preserving security and privacy. J. Big Data. 5, 1–18 (2018).
16. Li, X., Zhang, T.: An exploration on artificial intelligence application: From security, privacy and ethic perspective. 2017 2nd IEEE Int. Conf. Cloud Comput. Big Data Anal. ICCCBDA 2017. 416–420 (2017).
17. Kaissis, G.A., Makowski, M.R., Rückert, D., Braren, R.F.: Secure, privacy-preserving and federated machine learning in medical imaging. Nat. Mach. Intell. 2, 305–311 (2020).
18. Fredrikson, M., Jha, S., Ristenpart, T.: Model inversion attacks that exploit confidence information and basic countermeasures. Proc. ACM Conf. Comput. Commun. Secur.

2015-Octob, 1322–1333 (2015).
19. Qayyum, A., Qadir, J., Bilal, M., Al-Fuqaha, A.: Secure and Robust Machine Learning for Healthcare: A Survey. IEEE Rev. Biomed. Eng. 14, 156–180 (2021).
20. Al-Rubaie, M., Chang, J.M.: Privacy-Preserving Machine Learning: Threats and Solutions. IEEE Secur. Priv. 17, 49–58 (2019).
21. Azencott, C.A.: Machine learning and genomics: Precision medicine versus patient privacy. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. 376, (2018).
22. Passerat-Palmbach, J., Farnan, T., McCoy, M., Harris, J.D., Manion, S.T., Flannery, H.L., Gleim, B.: Blockchain-orchestrated machine learning for privacy preserving federated learning in electronic health data. Proc. - 2020 IEEE Int. Conf. Blockchain, Blockchain 2020. 550–555 (2020).
23. Imtiaz, S., Horchidan, S.F., Abbas, Z., Arsalan, M., Chaudhry, H.N., Vlassov, V.: Privacy Preserving Time-Series Forecasting of User Health Data Streams. Proc. - 2020 IEEE Int. Conf. Big Data, Big Data 2020. 3428–3437 (2020).
24. Liu, Y., Yang, R.: Federated Learning Application on Depression Treatment Robots (DTbot). 2021 IEEE 13th Int. Conf. Comput. Res. Dev. ICCRD 2021. 121–124 (2021).
25. Hu, Y., Ge, L., Zhang, G., Qin, D.: Research on differential privacy for medical health big data processing. Proc. - 2019 20th Int. Conf. Parallel Distrib. Comput. Appl. Technol. PDCAT 2019. 140–145 (2019).
26. Song, S., Chaudhuri, K., Sarwate, A.D.: Stochastic gradient descent with differentially private updates. 2013 IEEE Glob. Conf. Signal Inf. Process. Glob. 2013 - Proc. 245–248 (2013).
27. Fang, H., Qian, Q.: Privacy preserving machine learning with homomorphic encryption and federated learning. Futur. Internet. 13, 1–20 (2021).
28. Kelly, C.J., Karthikesalingam, A., Suleyman, M., Corrado, G., King, D.: Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 17, 1–9 (2019).
29. Jiang, Y., Hamer, J., Wang, C., Jiang, X., Kim, M., Song, Y., Xia, Y., Mohammed, N., Sadat, M.N., Wang, S.: SecureLR: Secure Logistic Regression Model via a Hybrid Cryptographic Protocol. IEEE/ACM Trans. Comput. Biol. Bioinforma. 16, 113–123 (2019).
30. Knott, B., Venkataraman, S., Hannun, A., Sengupta, S., Ibrahim, M., van der Maaten, L.: CrypTen: Secure Multi-Party Computation Meets Machine Learning. (2021).
31. Brocke, J. vom, Simons, A., Niehaves, B., Niehaves, B., Reimer, K., Plattfaut, R., Cleven, A.: Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. CIS 2009 Proc. 372, (2009).
32. Berthon, P., Pitt, L., Ewing, M., Carr, C.L.: Potential research space in MIS: A framework for envisioning and evaluating research replication, extension, and generation. Inf. Syst. Res. 13, 416–427 (2002).
33. Gregor, S.: The Nature of Theory in Information Systems. MIS Q. 30, 611–642 (2006).
34. Galetsi, P., Katsaliaki, K., Kumar, S.: Big data analytics in health sector: Theoretical framework, techniques and prospects. Int. J. Inf. Manage. 50, 206–216 (2020).
35. Zhang, X., Ding, J., Wu, M., Wong, S.T.C., Van Nguyen, H., Pan, M.: Adaptive Privacy Preserving Deep Learning Algorithms for Medical Data. 1168–1177 (2021).
36. Chen, X., Wang, X., Yang, K.: Asynchronous Blockchain-based Privacy-preserving Training Framework for Disease Diagnosis. Proc. - 2019 IEEE Int. Conf. Big Data, Big

Data 2019. 5469–5473 (2019).
37. Suriyakumar, V.M., Papernot, N., Goldenberg, A., Ghassemi, M.: Chasing your long tails: Differentially private prediction in health care settings. FAccT 2021 - Proc. 2021 ACM Conf. Fairness, Accountability, Transpar. 723–734 (2021).
38. Can, Y.S., Ersoy, C.: Privacy-preserving Federated Deep Learning for Wearable IoT-based Biomedical Monitoring. ACM Trans. Internet Technol. 21, (2021).
39. Cheng, V., Suriyakumar, V.M., Dullerud, N., Joshi, S., Ghassemi, M.: Can you fake it until you make it?: Impacts of differentially private synthetic data on downstream classification fairness. FAccT 2021 - Proc. 2021 ACM Conf. Fairness, Accountability, Transpar. 149–160 (2021).
40. Chen, H., Ünal, A.B., Akgün, M., Pfeifer, N.: Privacy-preserving SVM on outsourced genomic data via secure multi-party computation. IWSPA 2020 - Proc. 6th Int. Work. Secur. Priv. Anal. 61–69 (2020).
41. Yuan, D., Zhu, X., Wei, M., Ma, J.: Collaborative deep learning for medical image analysis with differential privacy. 2019 IEEE Glob. Commun. Conf. GLOBECOM 2019 - Proc. (2019).
42. Yasumura, Y., Ishimaki, Y., Yamana, H.: Secure naïve bayes classification protocol over encrypted data using fully homomorphic encryption. ACM Int. Conf. Proceeding Ser. (2019).
43. Vignesh, R., Vishnu, R., Raj, S.M., Akshay, M.B., Nair, D.G., Nair, J.R.: An improved method for sharing medical images for privacy preserving machine learning using multiparty computation and steganography. Proc. 2019 9th Int. Conf. Adv. Comput. Commun. ICACC 2019. 42–45 (2019).
44. Hakak, S., Ray, S., Khan, W.Z., Scheme, E.: A Framework for Edge-Assisted Healthcare Data Analytics using Federated Learning. Proc. - 2020 IEEE Int. Conf. Big Data, Big Data 2020. 3423–3427 (2020).
45. Sun, Z., Wang, Y., Shu, M., Liu, R., Zhao, H.: Differential Privacy for Data and Model Publishing of Medical Data. IEEE Access. 7, 152103–152114 (2019).
46. Muftuoglu, Z., Kizrak, M.A., Yildlnm, T.: Differential Privacy Practice on Diagnosis of COVID-19 Radiology Imaging Using EfficientNet. INISTA 2020 - 2020 Int. Conf. Innov. Intell. Syst. Appl. Proc. (2020).
47. Kim, H., Kim, S.H., Hwang, J.Y., Seo, C.: Efficient privacy-preserving machine learning for blockchain network. IEEE Access. 7, 136481–136495 (2019).
48. Guo, Y., Liu, F., Cai, Z., Chen, L., Xiao, N.: FEEL: A Federated Edge Learning System for Efficient and Privacy-Preserving Mobile Healthcare. ACM Int. Conf. Proceeding Ser. 19, (2020).
49. Leboe-Mcgowan, D., Al Aziz, M.M., Mohammed, N.: Simple Approximations for Fast and Secure Deep Learning on Genomic Data. 11th Annu. IEEE Inf. Technol. Electron. Mob. Commun. Conf. IEMCON 2020. 860–866 (2020).
50. Ying, Z., Zhang, Y., Cao, S., Xu, S., Liu, X.: OIDPR: Optimized Insulin Dosage based on Privacy-Preserving Reinforcement Learning. IFIP Netw. 2020 Conf. Work. Netw. 2020. 655–657 (2020).
51. Jarin, I., Eshete, B.: Pricure: Privacy-preserving collaborative inference in a multi-party setting. IWSPA 2021 - Proc. 2021 ACM Work. Secur. Priv. Anal. 25–35 (2021).
52. Kumar, R., Khan, A.A., Kumar, J., Zakria, A., Golilarz, N.A., Zhang, S., Ting, Y.,

Zheng, C., Wang, W.: Blockchain-Federated-Learning and Deep Learning Models for COVID-19 detection using CT Imaging. IEEE Sens. J. 21, 16301–16314 (2021).
53. Ying, Z., Cao, S., Xu, S., Liu, X., Lyu, L., Chen, C., Wang, L.: Privacy-Preserving Optimal Insuling Dosing Decision. IEEE Int. Conf. Acoust. Speech Signal Process. 2640–2644 (2021).
54. Zhang, M., Song, W., Zhang, J.: A Secure Clinical Diagnosis With Privacy-Preserving Multiclass Support Vector Machine in Clouds. IEEE Syst. J. 1–12 (2020).
55. Ma, Z., Ma, J., Miao, Y., Liu, X., Choo, K.-K.R., Yang, R., Wang, X.: Lightweight Privacy-preserving Medical Diagnosis in Edge Computing. IEEE Trans. Serv. Comput. 1374, 1–1 (2020).
56. Rahman, M.A., Shamim Hossain, M., Islam, M.S., Alrajeh, N.A., Muhammad, G.: Secure and provenance enhanced internet of health things framework: A blockchain managed federated learning approach. IEEE Access. 8, 205071–205087 (2020).
57. Thwal, C.M., Thar, K., Tun, Y.L., Hong, C.S.: Attention on personalized clinical decision support system: Federated learning approach. Proc. - 2021 IEEE Int. Conf. Big Data Smart Comput. BigComp 2021. 141–147 (2021).
58. Kwabena, O.A., Qin, Z., Qin, Z., Zhuang, T.: MSCryptoNet: Multi-Scheme Privacy-Preserving Deep Learning in Cloud Computing. IEEE Access. 7, 29344–29354 (2019).
59. Wu, Q., Chen, X., Zhou, Z., Zhang, J.: FedHome: Cloud-Edge based Personalized Federated Learning for In-Home Health Monitoring. IEEE Trans. Mob. Comput. 1233, 1–14 (2020).
60. Zhang, W., Zhou, T., Lu, Q., Wang, X., Zhu, C., Sun, H., Wang, Z., Lo, S.K., Wang, F.Y.: Dynamic Fusion-based Federated Learning for COVID-19 Detection. IEEE Internet Things J. 14, (2021).
61. Zhang, M., Wang, Y., Luo, T.: Federated Learning for Arrhythmia Detection of Non-IID ECG. 1176–1180 (2020).
62. Zhao, Y., Haddadi, H., Skillman, S., Enshaeifar, S., Barnaghi, P.: Privacy-preserving activity and health monitoring on databox. EdgeSys 2020 - Proc. 3rd ACM Int. Work. Edge Syst. Anal. Networking, Part EuroSys 2020. 49–54 (2020).
63. Pfitzner, B., Steckhan, N., Arnrich, B.: Federated Learning in a Medical Context: A Systematic Literature Review. ACM Trans. Internet Technol. 21, 1–31 (2021).
64. Brisimi, T.S., Chen, R., Mela, T., Olshevsky, A., Paschalidis, I.C., Shi, W.: Federated learning of predictive models from federated Electronic Health Records. Int. J. Med. Inform. 112, 59–67 (2018).
65. Rieke, N., Hancox, J., Li, W., Milletarì, F., Roth, H.R., Albarqouni, S., Bakas, S., Galtier, M.N., Landman, B.A., Maier-Hein, K., Ourselin, S., Sheller, M., Summers, R.M., Trask, A., Xu, D., Baust, M., Cardoso, M.J.: The future of digital health with federated learning. npj Digit. Med. 3, 1–7 (2020).
66. Peng, G., Dey, D., Lahiri, A.: Healthcare IT Adoption: An Analysis of Knowledge Transfer in Socioeconomic Networks. J. Manag. Inf. Syst. 31, 7–34 (2014).
67. Yeow, Adrian; Huat Goh, K.: Work Harder or Work Smarter? Information Technology and Resource Allocation in Healthcare Processes. IEEE Wirel. Commun. (2015).
68. Bhattacherjee, A., Hikmet, N., Menachemi, N., Kayhan, V.O., Brooks, R.G.: The differential performance effects of healthcare information technology adoption. Inf. Syst. Manag. 24, 5–14 (2007).

69. Goh, J.M., Gao, G., Agarwal, R.: Evolving work routines: Adaptive routinization of information technology in healthcare. Inf. Syst. Res. 22, 565–585 (2011).
70. Dünnebeil, S., Sunyaev, A., Blohm, I., Leimeister, J.M., Krcmar, H.: Determinants of physicians' technology acceptance for e-health in ambulatory care. Int. J. Med. Inform. 81, 746–760 (2012).
71. Mihailescu, M., Mihailescu, D.: Understanding Healthcare Digitalization: A Critical Realist Approach. ICIS 2017 Transform. Soc. with Digit. Innov. 0–12 (2018).
72. Venkatesh, V., Zhang, X., Sykes, T.A.: "Doctors do too little technology": A longitudinal field study of an electronic healthcare system implementation. Inf. Syst. Res. 22, 523–546 (2011).