Privacy-preserving Artificial Intelligence Techniques in Biomedicine

Reihaneh Torkzadehmahani[1], Reza Nasirigerdeh[1], David B. Blumenthal[1], Tim Kacprowski[1], Markus List[1], Julian Matschinske[1], Julian Späth[1], Nina Kerstin Wenke[1], Béla Bihari[2], Tobias Frisch[3], Anne Hartebrodt[3], Anne-Christin Hausschild[4], Dominik Heider[4], Andreas Holzinger[5], Walter Hötzendorfer[6], Markus Kastelitz[6], Rudolf Mayer[7], Cristian Nogales[8], Anastasia Pustozerova[7], Richard Röttger[3], Harald H.H.W. Schmidt[8], Ameli Schwalber[9], Christof Tschohl[6], Andrea Wohner[9], and Jan Baumbach[1]

[1]Chair of Experimental Bioinformatics, Technical University of Munich, Freising, Germany
[2]Gnome Design SRL, Sfântu Gheorghe, Romania
[3]Institute of Mathematics and Computer Science, University of Southern Denmark, Odense, Denmark
[4]Department of Mathematics and Computer Science, Philipps-University of Marburg, Marburg, Germany
[5]Institute for Medical Informatics/Statistics, Medical University Graz, Graz, Austria
[6]Research Institute AG & Co KG, Vienna, Austria
[7]SBA Research Gemeinnützige GmbH, Vienna, Austria
[8]Department of Pharmacology and Personalised Medicine, MeHNS, FHML, Maastricht University, Maastricht, the Netherlands
[9]Concentris Research Management GmbH, Fürstenfeldbruck, Germany

Abstract

Artificial intelligence (AI) has been successfully applied in numerous scientific domains including biomedicine and healthcare. Here, it has led to several breakthroughs ranging from clinical decision support systems, image analysis to whole genome sequencing. However, training an AI model on sensitive data raises also concerns about the privacy of individual participants. Adversary AIs, for example, can abuse even summary statistics of a study to determine the presence or absence of an individual in a given dataset. This has resulted in increasing restrictions to access biomedical data, which in turn is detrimental for collaborative research and impedes scientific progress. Hence there has been an explosive growth in efforts to harness the power of AI for learning from sensitive data while protecting patients' privacy. This paper provides a structured overview of recent advances in privacy-preserving AI techniques in biomedicine. It places the most important state-of-the-art approaches within a unified taxonomy, and discusses their strengths, limitations, and open problems.

Introduction

AI strives to emulate human intelligence and to develop intelligent algorithms that undertake complicated tasks. For many complex tasks, AI already surpasses humans in terms of accuracy, speed and cost. Recently, the rapid adoption of AI and its subfields, specifically machine learning and deep learning, has led to substantial progress in applications such as autonomous driving [1], text translation [2] and voice assistance [3]. At the same time, AI is becoming essential in biomedicine, where it has increasingly captured the attention of researchers. In particular, the rise of big data in healthcare makes it necessary to develop techniques that help scientists to gain understanding from it [4].

Success stories such as acquiring the compressed representation of drug-like molecules [5], modeling the hierarchical structure and function of a cell [6] and translating magnetic resonance images to computed tomography [7] using deep learning models illustrate the remarkable performance of these AI approaches. AI has not only achieved remarkable success in analyzing biomedicine data [8–18], but also has surpassed humans in applications such as sepsis prediction [19], malignancy detection on mammography [20] and mitosis detection in breast cancer [21].

Despite these AI-fueled advancements, important privacy concerns have been raised regarding the individuals who contribute to the training datasets. While taking care of the confidentiality and privacy of sensitive biological data is crucial, several studies showed that AI techniques often do not maintain data privacy [22–24]. In general, attacks known as membership inference can be used to infer an individual's membership by querying over the dataset [25] or the trained model [22], or by having access to certain statistics about the dataset [26–28]. Homer et al. [26] showed that under some assumptions, adversaries can use the genomic statistics published as the results of genome-wide association studies (GWAS) to find out if an individual was a part of the study. Another example of this kind of attack was demonstrated by attacks on Genomics Beacons [25, 29], in which an adversary (an attacker who attempts to invade data privacy) could identify the presence of an individual in the dataset by simply querying the presence of a particular allele. Moreover, the attacker could identify the relatives of those individuals and obtain sensitive disease information [28]. Besides targeting the training dataset, an adversary may attack a fully-trained AI model to extract individual-level membership by training an adversarial inference model that learns the behaviour of the target model [22].

As a result of the aforementioned studies, health research centers such as the National Institutes of Health (NIH) as well as hospitals have restricted access to the pseudonymized data [30–32]. Furthermore, data privacy laws such as those enforced by the Health Insurance Portability and Accountability Act (HIPAA), and the Family Educational Rights and Privacy Act (FERPA) in the US as well as the EU General Data Protection Regulation (GDPR) restrict the use of sensitive data [33, 34]. Consequently, everyone who needs access to these datasets has to go through a difficult approval process, which significantly impedes collaborative research. Therefore, both industry and academia urgently need to apply privacy-preserving techniques to respect individual privacy and comply with these laws.

This paper provides a systematic overview over various recently proposed privacy-preserving AI techniques, which facilitate the collaboration between health research institutes while ensuring data privacy. Several efforts exist to tackle the privacy concerns in the biomedical domain, some of which have been examined in a couple of surveys [35–37]. Aziz *et al.* [35] investigated previous studies which employed differential privacy and cryptographic techniques for human genomic data. Kaissis *et al.*[37] briefly reviewed federated learning, differential privacy and cryptographic techniques applied in medical imaging. Xu *et al.* [36] surveyed the general solutions to challenges in federated learning including communication efficiency, optimization, as well as privacy and discussed possible applications for federated learning including a few examples in healthcare. Our review differs from previous works in several aspects. Compared to [35] and [37], this paper covers a broader set of privacy preserving techniques including federated learning and hybrid approaches but also a wider range of problems such as privacy-preserving medical image segmentation and electronic health record classification. In contrast to [36] that only surveyed federated learning and hybrid approaches, this paper discusses cryptographic techniques and differential privacy approaches and their applications in healthcare too. Moreover, it covers a wider range of studies which employed four different privacy-preserving techniques for healthcare applications and compares the approaches using different criteria such as privacy, accuracy and efficiency.

The presented approaches in this review, are divided into four categories, namely, cryptographic techniques, differential privacy, federated learning, and hybrid approaches. First, we describe how cryptographic techniques — in particular, homomorphic encryption (HE) and secure multiparty computation (SMPC) — ensure secrecy of sensitive data by carrying out computations on encrypted biological data. Next, we illustrate the differential privacy approach and its capability in quantifying individuals' privacy in published summary statistics of, for instance, GWAS data and deep learning models trained on clinical data. Then, we elaborate on federated learning, which allows health institutes to train AIs locally and to share only selected parameters without sensitive data with a coordinator, who aggregates them and builds a global model. Following that, we discuss the hybrid approaches which enhance data privacy by combining multiple privacy-preserving techniques. We elaborate on the strengths and drawbacks of each approach as well as its applications in biomedicine and healthcare. Next, we provide a comparison among the approaches from different perspectives such as computational and communication efficiency, accuracy, and privacy. Afterwards, we discuss the most realistic approaches from practical viewpoint and provide a list of open problems and challenges to the adoption of these techniques in real-world healthcare applications.

Cryptographic Techniques

In the healthcare domain and GWAS in particular, cryptograohic techniques have been used to collaboratively compute result statistics while preserving the data privacy [38–48]. These cryptographic approaches are based on HE [49] or SMPC [50]. HE enables the computation of addition and multiplication over encrypted data.

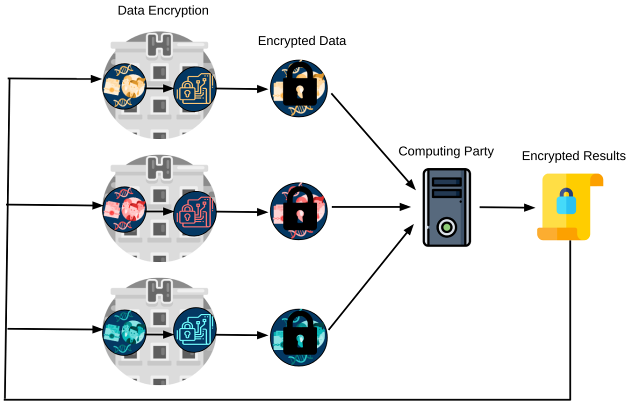


Figure 1: **Homomorphic encryption**: The participants encrypt the private data and share it to a computing party, which computes the aggregated (and encrypted) results over the encrypted data from the participants.

HE-based approaches share three steps (Figure 1):

1. Participants (e.g. hospitals or medical centers) encrypt their private data and send the encrypted data to a computing party.

2. The computing party calculates the statistics over the encrypted data and shares the statistics (which are encrypted) with the participants.

3. The participants access the results by decrypting them.

In SMPC, there are multiple participants as well as a couple of computing parties which perform computations on secret shares from the participants. Given M participants and N computing parties, SMPC-based approaches follow three steps (Figure 2):

1. Each participant sends a separate and different secret to each of the N computing parties.

2. Each computing party computes the intermediate results on the M secret shares from the participants and shares the intermediate results with the other $N-1$ computing parties.

3. Each computing party aggregates the intermediate results from all computing parties including itself to calculate the final (global) results. In the end, the final results computed by all computing parties are the same and can be shared by the participants.

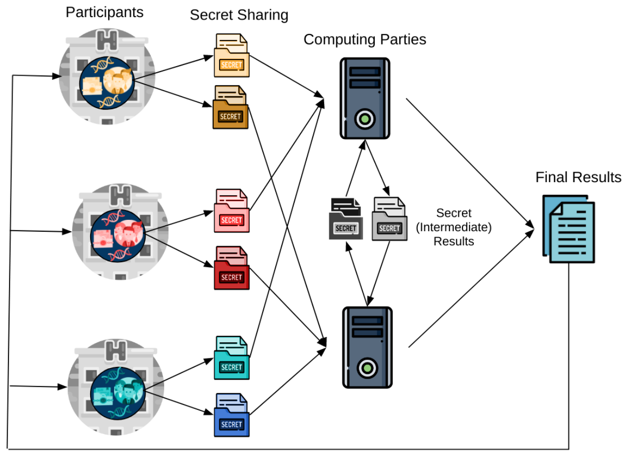


Figure 2: **Secure multi-party computation**: Each participant shares a separate, different secret with each computing party. The computing parties calculate the intermediate results, secretly share them with each other, and aggregate all intermediate results to obtain the final results.

To clarify the concepts of secret sharing [51] and multi-party computation, consider a scenario [52] with two participants P_1 and P_2 and two computing parties C_1 and C_2. P_1 and P_2 possess the private data X and Y, respectively. The aim is to compute $X+Y$, where neither P_1 nor P_2 reveals its data to the computing parties. To this end, P_1 and P_2 generate random numbers R_X and R_Y, respectively; P_1 reveals R_X to C_1 and $(X-R_X)$ to C_2; likewise, $P2$ shares R_Y with C_1 and $(Y-R_Y)$ with C_2; R_X, R_Y, $(X-R_X)$ and $(Y-R_Y)$ are secret shares. C_1 computes (R_X+R_Y) and sends it to C_2 and C_2 calculates $(X-R_X)+(Y-R_Y)$ and reveals it to C_1. Both C_1 and C_2 sum the result they computed and the result each obtained from the other computing party. The sum is in fact $(X+Y)$, which can be shared with P_1 and P_2.

It is worth mentioning that to preserve data privacy, the computing parties C_1 and C_2 must be non-colluding. That is, C_1 must not send R_X and R_Y to C_2 and C_2 must not share $(X-R_X)$ and $(Y-R_Y)$ with C_1. Otherwise, the computing parties can compute X and Y, revealing the participants' data. In general, in a SMPC with N computing parties, data privacy is protected as long as at most $N-1$ computing parties collude with each other. The larger N, the stronger the privacy but the higher the communication overhead and processing time.

Several studies use HE to develop secure, privacy-aware algorithms for healthcare data. Kim *et al.* [41] and Lu *et al.* [43] implemented a secure χ^2 test for GWAS data using HE. Lauter *et al.* [42] developed privacy-preserving versions of common statistical tests in in GWAS, such as Pearson good of fit test, tests for linkage disequilibrium, and the Cochran Armitage trend test. Kim *et al.* [53] and Morshed *et al.* [54] presented a secure logistic for GWAS and linear regression algorithm for healthcare data, based on HE.

Other studies mainly capitalized on SMPC to implement different privacy-preserving algorithms applicable to healthcare data. Zhang *et al.* [47], Constable *et al.* [46], and Kamm *et al.* [45] developed a secure χ^2 test based on SMPC for GWAS data. Shi *et al.* [55] developed a secure logistic regression algorithm using SMPC. Bloom [56] implemented a secure linear regression test based on SMPC for GWAS data. Cho *et al.* [38] introduced a SMPC based framework to facilitate quality control and population stratification correction for large-scale GWAS and showed that their framework is scalable to one million individuals and half million single nucleotide polymorphisms (SNPs).

Despite the promises of privacy-preserving algorithms leveraging cryptographic techniques (Table 1), the road for the wide adoption of these algorithms in the biomedicine and healthcare community is long [57]. The major limitations of HE are few supported operations and computational overhead [58]. HE supports only addition and multiplication operations, and as a result, developing complex AI models with non-linear operations such as deep neural networks (DNNs) using HE is very challenging. Moreover, HE incurs remarkable computational overhead since it performs operations on encrypted data. The main constraints of SMPC are computational overhead and network bottleneck [59]. Similar to HE, SMPC suffers from high overhead which comes from operating on secret shares from a large number of participants or large amount of data. Additionally, SMPC consumes high network bandwidth because participants need to send a large number of secret shares to the computing parties, which in turn, send the intermediate results to the other parties. Unlike HE, SMPC is flexible in terms of operations. On the other hand, HE is more communication-efficient compared to SMPC. Both HE and SMPC based algorithms are not scalable due to their computational overhead, which hinders their adoption for large-scale biomedical and healthcare data [57].

Table 1: Literature for **cryptographic techniques** in biomedicine. HE: homomorphic encryption, SMPC: secure multiparty computation

Authors	**Year**	**Privacy Technique**	**Model**	**Application**
Kim *et al.* [41]	2015	HE	χ^2 statistics minor allele frequency Hamming Distance Edit distance	genetic associations DNA comparison
Lu *et al.* [43]	2015	HE	χ^2 statistics D' measure	genetic associations
Lauter *et al.* [42]	2014	HE	D' and r^2 measures Pearson goodness-of-fit expectation maximization Cochran-Armitage	genetic associations
Kim *et al.* [53]	2018	HE	logistic regression	medical decision making
Morshed *et al.* [54]	2018	HE	linear regression	medical decision making
Kamm *et al.* [45]	2013	SMPC	χ^2 statistics	genetic associations
Constable *et al.* [46] Zhang *et al.* [47]	2015 2015	SMPC	χ^2 statistics minor allele frequency	genetic associations
Shi *et al.* [55]	2016	SMPC	logistic regression	genetic associations
Bloom [56]	2019	SMPC	linear regression	genetic associations
Cho *et al.* [38]	2018	SMPC	quality control population stratification	genetic associations

Differential Privacy

One of the state-of-the-art concepts for eliminating and quantifying the chance of information leakage that has gained considerable attention in recent years is *differential privacy* [60–66]. Differential privacy [67–69] is a mathematical model that encapsulates the idea of injecting enough randomness or noise to sensitive data. So, even a strong adversary with arbitrary auxiliary information about the data will still be uncertain in identifying any of the individuals in the dataset. It's primary goal is to camouflage the contribution of every single individual by inserting uncertainty into the learning process. It has become standard in data protection and has been effectively deployed by Google [70] and Apple [71] as well as agencies such as the United States Census Bureau. Furthermore, it has drawn the attention of researchers in privacy-sensitive fields such as biomedicine and healthcare [66, 72–86].

Differential privacy ensures that the model we train does not overfit the sensitive data of a particular user. In particular, the model trained on a dataset containing information of a specific individual should be statistically indistinguishable from a model trained without the individual (Figure 3). As an example, assume that a patient would like to give consent to his/her doctor to include his/her personal health record in a medical dataset to study the coordination between age and Cardiovascular disease. Differential privacy provides a mathematical guarantee which captures the privacy risk associated with the patient's participation in the study and explains to what extent the analyst or the potential adversary can learn about a particular individual in the dataset.

More formally, a randomized algorithm (an algorithm that has randomness in its logic and its output can vary even on a fixed input) $A : D^n \rightarrow Y$ is (ε, δ)-differentially private if for all subsets $y \subseteq Y$ and for all adjacent datasets $D, D' \in D^n$ that differ in at most one record the following inequality holds:

$$Pr[A(D) \in y] \leq e^{\varepsilon} Pr[A(D') \in y] + \delta$$

Here, ε and δ are privacy loss parameters where lower values imply stronger privacy guarantees. δ is an exceedingly small value (e.g. 10^{-5}) indicating the probability of an uncontrolled breach, where the algorithm produces a specific output only in the presence of a specific individual and not otherwise. ε represents the worst case privacy breach in the absence of any such rare breach. If you assume $\delta = 0$, you will have a pure (ε)-differentially private algorithm, while if you consider $\delta > 0$ to approximate the case in which pure differential privacy is broken, you will have an approximate (ε, δ)-differentially private algorithm.

Two important properties of differential privacy are composability [87] and resilience to post-processing. Composability means that combining multiple differentially private algorithms yields another differentially private algorithm. More precisely, if you combine k (ε, δ)-differentially private algorithms, the composed algorithm is at least ($k\varepsilon$, $k\delta$)-differentially private. Differential privacy also assures resistance to post-processing theorem which states passing the output of an (ε, δ)-differentially private algorithm to any arbitrary randomized algorithm will still uphold the (ε, δ)-differential privacy guarantee.

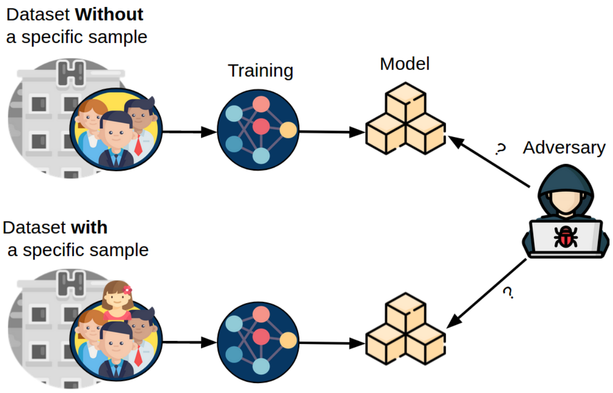


Figure 3: **Differential privacy**; The model trained on a dataset including a specific individual and the one trained on the same dataset excluding that individual, looks statistically indistinguishable to the adversary.

The community efforts to ensure the privacy of sensitive biomedicine data using differential privacy can be grouped into four categories according to the problem they address (Table 2):

1. Approaches to query genomics databases [66, 85, 86].
2. Statistical and AI modeling techniques in biomedicine [78–83].
3. Data release, i.e., releasing summary statistics such as p-values and χ^2 contingency tables [73–75, 84].
4. Training privacy-preserving generative models [63, 88, 89].

Studies in the first category proposed solutions to reduce the privacy risks of genomics databases such as GWAS databases and genomics beacon service [90]. The Beacon Network [29] is an online web service developed by the Global Alliance for Genomics and Health (GA4GH) through which the users can query the data provided by owners or research institutes, ask about the presence of a genetic variant in the database, and get a YES/NO as response. Studies have shown that an attacker can detect membership in the Beacon or GWAS by querying these databases multiple times and asking different questions [25, 91, 92]. In a recent work, Aziz *et al.* [86] proposed two lightweight algorithms to make the Beacon's response inaccurate by controlling a bias variable. These algorithms decide when to answer the query correctly/incorrectly according to specific conditions in the bias variable so that it gets harder for the attacker to succeed. In another work, Johnson *et al.* [66] developed a differentially private query answering framework. With this framework the analysts can explore the GWAS data without any prior knowledge of the number and location of SNPs in the DNA sequence. The analysts can retrieve statistical properties such as the correlation between SNPs and get an almost accurate answer while the GWAS dataset is protected against privacy risks.

Some of the efforts in the second category addressed the privacy concerns in GWAS data analysis by introducing differentially private logistic regression to identify associations between SNPs and diseases [81] or associations among multiple SNPs [79]. Honkela *et al.* [80] improve drug sensitivity prediction by effectively employing differential privacy for Bayesian linear regression. Moreover, Simmons *et al.* [83] presented a differentially private EIGENSTRAT (PrivSTRAT) [93] and linear mixed model (PrivLMM) [94] while correcting for population stratification. In another work, Simmons *et al.* [82] tackled the problem of finding significant SNPs by modeling it as an optimization problem. Solving this problem provides a differentially private estimate of the neighbor distance for all SNPs, such that high scoring SNPs can be found.

The third category focused on releasing summary statistics such as p-values, χ^2 contingency tables, and minor allele frequencies in a differentially private fashion. The common approach in these works is to add Laplacian noise to the true value of the statistics, so that sharing the perturbed statistics preserves privacy of the individuals. They vary in the sensitivity of the algorithms (that is, the maximum change on the output of an algorithm in presence or absence of a specific data point) and hence require different amounts of injected noise [73, 74, 84].

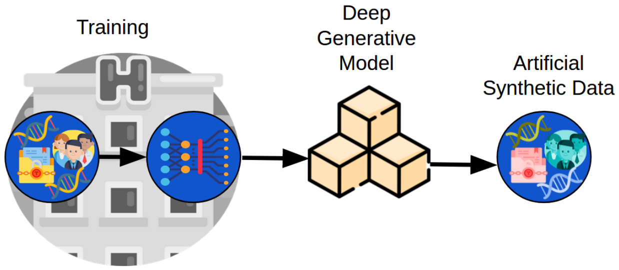


Figure 4: **Differentially private deep generative models:** The sensitive data holder (e.g. health institutes) train a differentially private generative model locally and share just the trained data generator with the outside world (e.g. researchers). The shared data generator can then be used to produce artificial data with the same characteristics as the sensitive data.

The forth category proposed novel privacy-protecting methods to generate synthetic healthcare data leveraging differentially private generative models (Figure 4). Deep generative models, such as generative adversarial networks (GANs), can be trained on sensitive biomedical data to capture its properties and generate artificial data with similar characteristics as the original data.

Abay *et al.* [88] presented a differentially private deep generative model, DP-SYN, a generative autoencoder that splits the input data into multiple partitions, then learns and simulates the representation of each partition while maintain-

Table 2: Literature for **differentially private** (DP) techniques in biomedicine

Authors	Year	Model	Application
Aziz *et al.* [86]	2017	eliminating random positions biased random response	querying genomics database
Johnson *et al.* [66]	2013	distance-score mechanism p-value and χ^2 statistics	querying genomics database
Han *et al.* [81] Yu *et al.* [79]	2019 2014	logistic regression	genetic associations
Honkela *et al.* [80]	2018	bayesian linear regression	drug sensitivity prediction
Simmons *et al.* [83]	2016	EIGENSTRAT linear mixed model	genetic associations
Simmons *et al.* [82]	2016	nearest neighbor optimization	genetic associations
Fienberg *et al.* [73] Uhlerop *et al.* [74] Yu *et al.* [75] Wang *et al.* [84]	2011 2013 2014 2014	statistics such as p-value, χ^2 and contingency table	genetic associations
Abay *et al.* [88]	2018	deep autoencoder	generating artificial medical data
Beaulieu *et al.* [63]	2019	GAN	simulating SPRINT trial
Jordon *et al.* [89]	2018	GAN	generating artificial medical data

ing the privacy of input data. They assessed the performance of DP-SYN on sensitive datasets of breast cancer and diabetes. Beaulieu *et al.* [63] trained an auxiliary classifier GAN (AC-GAN) in a differentially private manner to simulate the participants of the SPRINT trial (Systolic Blood Pressure Trial), so that the clinical data can be shared while respecting participants' privacy. In another approach, Jordon *et al.* [89] introduced a differentially private GAN, PATE-GAN, and evaluated the quality of synthetic data on Meta-Analysis Global Group in Chronic Heart Failure (MAGGIC) and the United Network for Organ Transplantation (UNOS) datasets.

Despite the aforementioned achievements in adopting differential privacy in the field, several challenges remain to be addressed. Although differential privacy involves less network communication, memory usage and time complexity compared to cryptographic techniques, it still struggles with giving highly accurate results within a reasonable privacy budget, namely, intended ε and δ, on large scale datasets such as genomics datasets [35, 95]. In more details, since the genomics datasets are huge, the sensitivity of the applied algorithms on these datasets is large. Hence, the amount of distortion required for anonymization increases significantly, sometimes to the extent that the results will not be meaningful anymore [96]. Therefore, to make differential privacy more practical in the field, balancing a trade off between privacy and utility demands more attention than it has received [76–78, 84].

Federated Learning

Federated learning [97] is a type of distributed learning where multiple clients (e.g. hospitals) collaboratively learn a model under the coordination of a central server while preserving the privacy of their data [98], [99]. Instead of sharing its private data with the server or the other clients, each client extracts knowledge (that is, model parameters) from its data and transfers it to the server for aggregation (Figure 5).

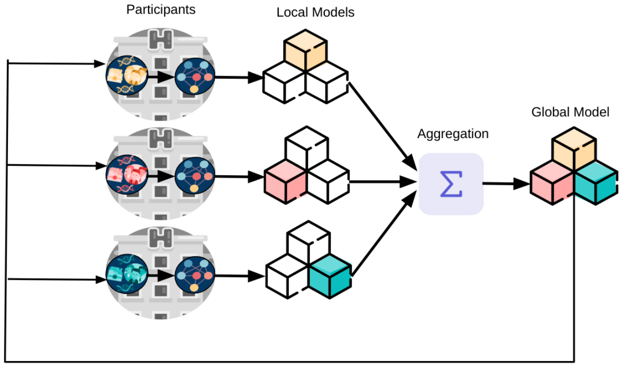


Figure 5: **Federated Learning**: Each participant downloads the global model from the server, computes the local model given its private data and the global model, and finally sends its local model to the server for aggregation and for updating the global model.

Federated learning is an iterative process in which each iteration consists of the following steps [99]:

1. The server chooses a set of clients to participate in the current iteration of the model.

2. The selected clients obtain the current model from the server.

3. Each selected client computes the local parameters using the current model and its private data (e.g., runs

gradient descent algorithm initialized by the current model on its local data to obtain the local gradient updates).

4. The server collects the local parameters from the selected clients and aggregates them to update the current model.

The data of the clients can be considered as a table, where rows represent samples (e.g., individuals) and columns represent features or labels (e.g., age, blood pressure, case vs. control). We refer to the set of samples, features, and labels of the data as *sample space*, *feature space*, and *label space*, respectively. Federated learning can be categorized into three types based on the distribution characteristics of the clients' data:

- **Horizontal (sample-based) federated learning** [100]: Data from different clients shares similar feature space but is very different in sample space. As an example, consider two hospitals in two different cities which collected similar information such as age or sex. In this case, the feature spaces are similar; but because the people who participated in the hospitals' data collections are from different cities, their intersection is most probably very small, and the sample spaces are hence very different.

- **Vertical (feature-based) federated learning** [100] : Clients' data is similar in sample space but very different in feature space. For example, two hospitals with different expertise in the same city might collect different information (different feature space) from almost the same people (similar sample space).

- **Hybrid federated learning**: Both feature space and sample space are different in the data from the clients. For example, consider a medical center with expertise in brain image analysis located in New York and a research center with expertise in protein research based in Berlin. Their data is completely different (image vs. protein data) and disjoint groups of individuals participated in the data collection of each center.

To illustrate the concept of federated learning, consider a scenario with two hospitals A and B. A and B possess lists X and Y, containing the age of their cancer patients, respectively. A simple federated mean algorithm to compute the average age of cancer patients in both hospitals without revealing the real values of X and Y works as follows: For the sake of brevity, we assume that both hospitals are selected in the first step and that the current global model parameters in the second step are zero (see federated learning steps).

- Hospital A computes the average age (M_X) and number of its cancer patients (N_X). Hospital B does the same, resulting in M_Y, N_Y. Here, X and Y are private data while M_X, N_X, M_Y, N_Y are the parameters extracted from the private data.

- The server obtains the values of local model parameters from the hospitals and computes the global mean as follows:

$$M_G = \frac{M_X \times N_X + M_Y \times N_Y}{N_X + N_Y}$$

Two well-known concepts in machine learning are also related to federated learning: transfer learning [101] and multi-task learning [102, 103]. In transfer learning, there are source and destination tasks. The aim is to transfer the knowledge from the source to the destination task. As an example of federated transfer learning, suppose that hospital A has a DNN model trained on its rich dataset of medical images (source task). On the other hand, the hospital B wants to train a DNN model on a dataset containing brain images (a special kind of medical images) of its cancer patients (destination task) but the dataset does not have enough samples. Hospital B can take advantage of hospital A's DNN model by incorporating some parts of the source model into its own DNN model (knowledge transfer) instead of training the model from scratch on its dataset [104].

In multi-task learning, there are multiple tasks and the goal is to exchange the knowledge among the tasks to improve the performance (accuracy) of all tasks. As an example of federated multi-task learning, assume hospitals A and B again, where hospital A has a task of training a DNN model on its cancer image dataset and hospital B's task is to train a logistic regression model on a dataset including the age, sex, and genetic variants of its cancer patients. Here, both DNN and logistic regression models are trained concurrently (and iteratively) and the knowledge (weights) from both models are exchanged in each iteration to improve (tune the weights) both models.

A crucial consideration in both transfer and multi-task learning is task relatedness. Employing unrelated tasks can lead to transferring negative knowledge and deteriorating the performance of the model(s). To learn more about transfer/multi-task learning, interested readers are referred to [101–103, 116]. Moreover, federated transfer/multi-task learning can be a horizontal or hybrid federated learning approach. In the example provided for federated transfer learning, if the shape of the images in the source and destination tasks (feature space) are the same, it is considered as a horizontal approach. Otherwise, it is a hybrid federated learning approach similar to the example given for federated multi-task learning.

The emerging demand for federated learning gave rise to a wealth of both simulation [117, 118] and production-oriented [119, 120] open source frameworks. Additionally, there are AI platforms whose goal is to apply federated learning in real-world healthcare settings [121, 122]. In the following, we survey works on federated AI techniques in biomedicine and healthcare (Table 3). The recent studies in this regard mainly focused on horizontal federated learning and there are a few vertical federated learning and federated transfer/multi-task learning algorithms applicable to healthcare and biomedical data.

Table 3: Summary of **federated learning** (FL) approaches in healthcare and biomedicine

Authors	Year	Model	Application
Sheller *et al.* [105]	2018	DNN	medical image segmentation
Chang *et al.* [106]	2018	single weight transfer	medical image classification
Balachandar *et al.* [107]	2020	cyclical weight transfer	
Nasirigerdeh *et al.* [108]	2020	linear regression, chi-square, logistic regression	GWAS
Wu *et al.* [109]	2012	logistic regression	GWAS
Wang *et al.* [110]	2013		
Li *et al.* [111]	2016		
Brisimi *et al.* [112]	2018	support vector machine	classifying electrical health records
Huang *et al.* [113]	2018	adaptive boosting ensemble	classifying medical data
Liu *et al.* [114]	2018	autonomous deep learning	classifying medical data
Chen *et al.* [115]	2019	transfer learning	training wearable healthcare devices

A number of the studies provided solutions for the lack of sufficient data due to the the privacy challenges in the medical imaging domain [105–107, 123–125]. For instance, Sheller *et al.* developed a supervised DNN in a federated way for semantic segmentation of brain Gliomas from magnetic resonance imaging scans [105]. Chang *et al.* [106] simulated a distributed DNN in which multiple participants collaboratively update model weights using training heuristics such as single weight transfer and cyclical weight transfer (CWT). They evaluated this distributed model using image classification tasks on medical image datasets such as mammography and retinal fundus image collections, which were evenly distributed among the participants. Balachandar *et al.* [107] optimized CWT for cases where the datasets are unevenly distributed across participants. They assessed their optimization methods on simulated diabetic retinopathy detection and chest radiograph classification.

Federated linear/logistic regression or chi-square test have been developed for sensitive biological data that is vertically or horizontally distributed [108–111]. The grid binary logistic regression (GLORE) [109] and the expectation propagation logistic regression (EXPLORER) [110] are horizontal federated learning approaches designed for clinical data. Unlike GLORE, EXPLORER supports asynchronous communication and online learning functionality so that the system can continue collaborating in case a participant is absent or if communication is interrupted. Li *et al.* presented VERTIGO [111], a vertical grid logistic regression algorithm designed for vertically distributed biological datasets such as breast cancer genome and myocardial infarction data. Nasirigerdeh *et al.* [108] developed a horizontally federated tool set for GWAS, called *sPLINK*, which supports chi-square test, linear regression, and logistic regression. Notably, federated results from *sPLINK* on distributed datasets are the same as those from aggregated analysis conducted with *PLINK* [126]. Moreover, they showed that *sPLINK* is robust against heterogeneous (imbalanced) data distributions across clients and does not lose its accuracy in such scenarios.

Moreover, there are studies in the literature that combine federated learning with other traditional AI modeling techniques such as ensemble learning, support vector machines (SVMs) and principle component analysis (PCA) [112–115, 127]. Brisimi *et al.* [112] presented a federated soft-margin support vector machine (sSVM) for distributed electronic health records. Huang *et al.* [113] introduced LoAdaBoost, a federated adaptive boosting method for learning medical data such as intensive care unit data from distinct hospitals [128] while Liu *et al.* [114] trained a federated autonomous deep learner to this end. There have also been a couple of attempts at incorporating federated learning into multi-task learning and transfer learning in general [129–131]. However, to the best of our knowledge, FedHealth [115] is the only federated transfer learning framework specifically designed for healthcare applications. It enables users to train personalized models for their wearable healthcare devices by aggregating the data from different organizations without compromising privacy.

One of the major challenges for adopting federated learning in large scale healthcare applications is the significant network communication overhead, especially for complex AI models such as DNNs that contain millions of model parameters and require thousands of iterations to converge. A rich body of literature exists to tackle this challenge, known as communication-efficient federated learning. These approaches can be categorized into three categories: gradient quantification [132], gradient sparsification [133], and more local updates in clients than global model update [134].

The main idea behind gradient quantification is to use less bytes for each model parameter (gradient), e.g., 2 bytes instead of 8. In gradient sparsification, instead of sending all parameters, a fraction, e.g. 10%, of parameters is exchanged between the server and clients, saving 90% network bandwidth. In the last category of communication-efficient approaches, the clients update their local parameters multiple times before sending them to the server to reduce the number of total iterations, and as a result, decrease the total network bandwidth usage.

There is a trade-off between communication efficiency and model convergence (accuracy). Employing

Table 4: Summary of the **hybrid** privacy-preserving approaches in healthcare and biomedicine

Authors	Year	Privacy Technique	Model	Application
Li *et al.* [135]	2019	FL+DP	DNN	medical image segmentation
Li *et al.* [136]	2020	FL+DP	domain adoption	medical image pattern recognition
Choudhury *et al.* [137]	2019	FL+DP	perceptron neural network support vector machine logistic regression	classifying electronic health records
Constable *et al.* [46]	2015	FL+SMPC	statistical analysis (e.g. χ^2 statistics)	genetic associations
Lee *et al.* [138]	2018	FL+HE	context-specific hashing	learning patient similarity
Kim *et al.* [139]	2019	FL+DP+HE	logistic regression	classifying medical data

communication-efficient approaches reduces the network overhead but might jeopardize the model convergence. Consequently, one should keep in mind that communication-efficient approaches should be leveraged as long as they keep the accuracy of the model acceptable. Interested readers are referred to relevant publications [134, 140] for detailed descriptions.

Another challenge in federated learning is the possible accuracy loss from the aggregation process if the data distribution across the clients is not independent and identically distributed (IID). More specifically, federated learning can deal with non-IID data while preserving the model accuracy if the learning model is simple such as ordinary least squares (OLS) linear regression (*sPLINK* [108]). However, when it comes to learning complex models such as DNNs, the global model might not converge on non-IID data across the clients. *Zhao et al.* [141] showed that simple averaging of the model parameters in the server significantly diminishes the accuracy of a convolutional neural network model in highly skewed non-IID settings. To solve this problem, they train a warm-up model on an IID dataset and share the model as well as a portion of the dataset with all clients. Each client uses its local data and the shared dataset to train the local model and the simple averaging is employed in the server to aggregate the model parameters. Developing the aggregation strategies which are robust against non-IID scenarios is still an open and interesting problem in federated learning.

Finally, federated learning is based on the assumption that the centralized server is honest and not compromised, which is not necessarily the case in real applications. To relax this assumption, differential privacy or cryptographic techniques can be leveraged in federated learning, which is covered in the next section. For further reading on future directions of federated learning in general, we refer the reader to comprehensive surveys [99, 142, 143].

Hybrid Privacy-preserving Techniques

The hybrid techniques combine federated learning with the other paradigms (cryptographic techniques and differential privacy) to enhance privacy or provide privacy guarantees (Table 4). Federated learning preserves privacy to some extent because it does not require the health institutes to share the patients' data with the central server. However, the model parameters that participants share with the server might be abused to reveal the underlying private data if the coordinator is compromised [144]. To handle this issue, the participants can leverage differential privacy and add noise to the model parameters before sending them to the server (FL+DP) [135, 136, 145, 146] or they employ HE (FL+HE) or SMPC (FL+SMPC) to securely share the parameters with the server [46, 138].

In the biomedical field, several hybrid approaches have been presented recently. Li *et al.* [135] presented a federated deep learning framework for magnetic resonance brain image segmentation in which the client side provides differential privacy guarantees on selecting and sharing the local gradient weights with the server for imbalanced data. A recent study [136] extracted neural patterns from brain functional magnetic resonance images by developing a privacy-preserving pipeline that analyzes image data of patients having different psychiatric disorders using federated domain adaption methods. Choudhury *et al.* [137] developed a federated differential privacy mechanism for gradient-based classification on electronic health records. There are also some studies that incorporate federate learning with cryptographic techniques. For instance, Constable *et al.* [46] implemented a privacy-protecting structure for federated statistical analysis such as χ^2 statistics on GWAS while maintaining privacy using SMPC. In a slightly different approach, Lee *et al.* [138] presented a privacy-preserving platform for learning patient similarity in multiple hospitals using a context-specific hashing approach which employs homomorphic encryption to limit the privacy leakage. Moreover, Kim *et al.* [139] presented a privacy-preserving federated logistic regression algorithm for horizontally distributed diabetes and intensive care unit datasets. In this approach, the logistic regression ensures privacy by making the aggregated weights differentially private and encrypting the local weights using homomorphic encryption.

Incorporating HE, SMPC, and differential privacy into federated learning brings about enhanced privacy but it combines the limitations of the approaches, too. FL+HE puts much more computational overhead on the server, since

Table 5: Comparison among the privacy-preserving techniques including homomorphic encryption (HE), secure multiparty computation (SMPC), federated learning (FL), differential privacy (DP) and the hybrid approaches (FL+DP, FL+HE and FL+SMPC); The generic ranking (lowest =1 to highest = 6) is used for comparison purposes such that having a higher score for a criteria, represents performing better on that metric.

	HE	SMPC	DP	FL	FL+DP	FL+HE	FL+SMPC
Accuracy	2	6	1	5	3	4	5
Computational efficiency	1	2	6	6	5	3	4
Network communication efficiency	5	4	6	3	3	2	1
Privacy of exchanged traffic	4	3	NA	1	2	4	3
Exchanging low sensitive traffic	✗	✗	NA	✓	✓	✓	✓
Privacy guarantee	✗	✗	✓	✗	✓	✗	✗

it requires to perform aggregation on the encrypted model parameters from the clients. The network communication overhead is exacerbated in FL+SMPC, because clients need to securely share the model parameters with multiple computing parties instead of one. FL+DP might result in inaccurate models because of adding noise to the model parameters in the clients.

Comparison

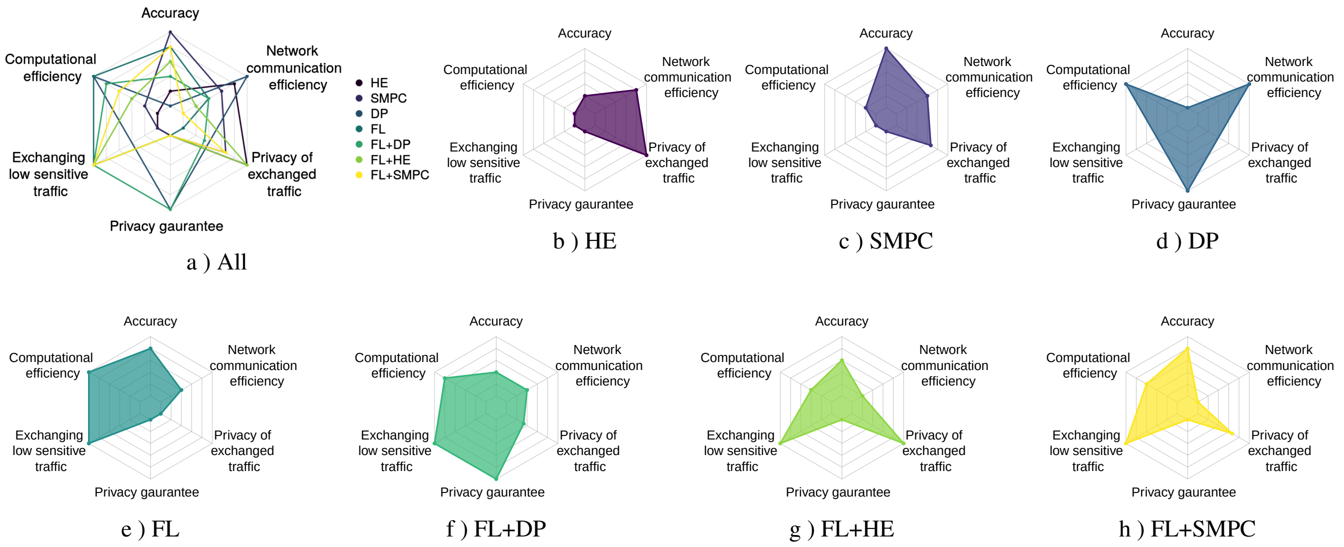
We compare the privacy-preserving techniques (HE, SMPC, differential privacy, federated learning, and the hybrid approaches) using various performance and privacy criteria such as *computational/communication efficiency*, *accuracy*, *privacy guarantee*, *exchanging sensitive traffic through network* and *privacy of exchanged traffic* (Table 5 and Figure 6). We employ a generic ranking (lowest =1 to highest = 6) [35] for all comparison criteria except for *privacy guarantee* and *exchanging sensitive traffic through network*, which are binary criteria. This comparison is made under the assumption of applying a complex model (e.g. DNN with a huge number of model parameters) on the large sensitive biomedical datatsets distributed across dozens of clients in IID configuration. Additionally, there are a few computing parties in SMPC (practical configuration).

Computational efficiency is an indicator of the extra computational overhead an approach incurs to preserve the privacy. According to Table 5 and Figure 6, differential privacy and federated learning are the best from this perspective. This is because the noise injection procedure in differential privacy is not computationally expensive and federated learning follows the paradigm of bringing computation to data, distributing computational overhead among the clients. HE and SMPC are based on the paradigm of moving data to computation. In HE, encryption of the whole private data in the clients and carrying out computation on encrypted data by the computing party cause a huge amount of overhead. In SMPC, a couple of computing parties process the secret shares from dozens of clients, incurring considerable computational overhead. Among the hybrid approaches, FL+DP has the best computational efficiency given the lower overhead of the two approaches whereas FL+HE has the highest overhead because aggregation process on encrypted parameters is computationally expensive.

Network communication efficiency indicates how efficient an approach utilizes the network bandwidth. The less data traffic is exchanged in the network, the more communication-efficient the approach is. Federated learning is the least efficient approach from the communication aspect since exchanging a large number of model parameter values between the clients and the server generates a huge amount of network traffic. Notice that network bandwidth usage of federated learning is independent of the clients' data because federated learning does not move data to computation but depends on the model complexity (i.e. the number of model parameters). The next approach in this regard is SMPC, where not only each participant sends a large traffic (almost as big as its data) to each computing party but also each computing party exchanges intermediate results (which might be large) with the other computing parties through the network. The network overhead of homomorphic encryption comes from sharing the encrypted data of the clients (as big as the data itself) with the computing party, which is small compared to network traffic generated by federated learning and SMPC. The best approach is differential privacy with no network overhead. Accordingly, FL+DP and FL+SMPC are the best and worst among the hybrid approaches from communication efficiency viewpoint, respectively.

Accuracy of the model in a privacy-preserving approach is a crucial factor in whether to adopt the approach. SMPC and federated learning are the most accurate approaches incurring no or a little bit accuracy loss in the final model. The next is homomorphic encryption whose accuracy loss is due to approximating the non-linear operations using addition and multiplication (e.g. least squares approximation [53]). The worst approach is differential privacy where the added noise can considerably affect the model accuracy. In the hybrid approaches, FL+SMPC is the best and FL+DP is the worst considering the accuracy of SMPC and differential privacy approaches.

The rest of the comparison measures are privacy-related. The traffic transferred from the clients (participants) to the server (computing parties) is highly sensitive if it carries



a) All

b) HE

c) SMPC

d) DP

e) FL

f) FL+DP

g) FL+HE

h) FL+SMPC

Figure 6: Comparison radar plots for all (a) and each of (b-h) the privacy preserving approaches including homomorphic encryption (HE), Secure multiparty computation (SMPC), differential privacy (DP), federated learning (FL) and hybrid techniques (FL+DP, FL+HE and FL+SMPC)

the private data of the clients. The less sensitive the exchanged traffic is, the more robust the approach is from the privacy perspective. HE and SMPC send the encrypted and anonymous form of the clients' private data to the server, respectively. Federated learning and hybrid approaches share only the model parameters with the server. In HE, if the server has the key to decrypt the traffic from the clients, the whole private data of the clients will be revealed. The same holds if the computing parties in SMPC collude with each other. This might or might not be the case for the other approaches (e.g. federated learning) depending on the exchanged model parameters and whether they can be abused to infer the underlying private data.

Privacy of the exchanged traffic indicates how much the traffic is kept private from the server. In HE/SMPC, the data is encrypted/anonymized first and then shared with the server, which is reasonable since it is the clients' private data. In federated learning, the traffic (model parameters) is directly shared with the server assuming that it does not reveal any details regarding individual samples in the data. The aim of the hybrid approaches is to hide the real values of the model parameters from the server to minimize the possibility of inference attacks using the model parameters. FL+HE is the best among the hybrid approaches from this viewpoint.

Privacy guarantee is a metric which quantifies the degree to which the privacy of the clients' data can be preserved. Differential privacy and the corresponding hybrid approach (FL+DP) are the only approaches providing a privacy guarantee, whereas all other approaches can only protect the privacy under a set of certain assumptions. HE assumes that the server does not have the decryption key; The underlying assumption in SMPC is that the computing parties do not collude with each other; federated learning supposes that the model parameters do not give any detail about a sample in the clients' data.

Discussion and open problems

From a practical point of view, homomorphic encryption and SMPC that follow the paradigm of ”move data to computation” do not scale as the number of clients or data size in clients become large. This is because they put the computational burden on a single or a few computing parties. Federated learning, on the other hand, distributes the computation across the clients (aggregation on the server is not computationally heavy) but the communication overhead between the server and clients is the major challenge to scalability of federated learning. The hybrid approaches inherit this issue and it is exacerbated in FL+SMPC. Combining homomorphic encryption with federated learning (FL+HE) adds another obstacle (computational overhead) to scalability of federated learning. There is a growing body of literature on communication-efficient approaches to federated learning, which we already discussed. These approaches can dramatically improve the scalability of federated learning and make it suitable for large-scale applications including those in biomedicine.

Given that federated learning is the most realistic approach from a scalability viewpoint, it can be used as a standalone approach as long as inferring the clients' data from the model parameters is practically impossible. Otherwise, it should be combined with differential privacy to avoid possible inference attacks and exposure of clients' private data and to provide privacy guarantee. The accuracy of the model will be satisfactory in federated learning but it might be deteriorated in FL+DP. A realistic trade-off needs to be considered depending on the application of interest.

Moreover, differential privacy can have many practical applications in biomedicince as a standalone approach. It works very well for low-sensitivity queries such as counting queries (e.g number of patients with a specific disease) on biomedical databases and its generalizations (e.g. histograms) since the presence or absence of an individual changes the query's response by at most one. Moreover, it can be employed to release summary statistics such as χ^2 and p-values in a differentially private manner while keeping the accuracy acceptable. A novel promising research direction is to incorporate differential privacy in deep generative models to generate synthetic biomedical data.

Future studies can investigate how to reach a compromise between scalability, privacy, and accuracy in real-world settings. The communication overhead of federated learning is still an open and interesting problem since although state-of-the-art approaches considerably reduce the network overhead, they adversely affect the accuracy of the model. Hence, novel approaches are required to preserve the accuracy, which is of great importance in biomedicine applications, while making federated learning communication-efficient.

Adopting federated learning in non-IID settings, where biomedical datasets across different hospitals/medical centers are heterogeneous, is another important challenge to address. This is because typical aggregation procedures such as simple averaging do not work well for these settings, yielding inaccurate models. Hence, new aggregation procedures are required to tackle non-IID scenarios. Moreover, current communication-efficient approaches which were developed for an IID setting might not be applicable to heterogeneous scenarios. Consequently, new techniques are needed to reduce network overhead in these settings, while keeping the model accuracy satisfactory.

Combining differential privacy with federated learning to enhance privacy and to provide a privacy guarantee is still a challenging issue in the field. It becomes even more challenging for healthcare applications, where accuracy of the model is of crucial importance. Moreover, the concept of privacy guarantee in differential privacy has been defined for local settings. In distributed scenarios, a dataset might be employed multiple times to train different models with various privacy budgets. Therefore, a new formulation of privacy guarantee should be proposed for distributed settings.

Conclusion

The advent of AI in biomedicine has brought about indispensable progress in the field and is expected to result in even more impressive advances in the future [147]. For AI techniques to succeed, big biomedical or healthcare data needs to be available and accessible. However, the more AI models are trained on sensitive biological data, the more pressing privacy concerns become, which, in turn, necessitate strategies for shielding the data [148]. Hence, privacy-enhancing techniques are crucial to allow AI to benefit from the sensitive biological data.

Cryptographic techniques, differential privacy and federated learning can be considered as the prime strategies for protecting personal data privacy. Broadly, these emerging techniques are based on either securing sensitive data, perturbing it or not moving it off site. In particular, cryptographic techniques securely share the data with a single (HE) or multiple computing parties (SMPC), differential privacy adds noise to sensitive data and quantifies privacy loss accordingly, while federated learning enables collaborative learning under orchestration of a centralized server without moving the private data outside local environments.

All of these techniques have their own strengths and limitations. HE and SMPC are more communication efficient compared to federated learning but they are computationally expensive since they move data to computation and put the computational burden on a server or a few computing parties. Federated learning, on the other hand, distributes computation across the clients but suffers from high network communication overhead. Differential privacy is an efficient approach from a computational and a communication perspective but it introduces accuracy loss by adding noise to data or model parameters. Hybrid approaches are studied to combine the advantages or to overcome the disadvantages and limitations of the individual techniques. We argued that federated learning as a standalone approach or in combination with differential privacy is the most realistic approach to be adopted in healthcare applications. We discussed the open problems and challenges in this regard including the balance of communication efficiency and model accuracy in non-IID settings, and need for a new notion of privacy guarantee for distributed biomedical datasets.

Incorporating privacy into the analysis of biomedical and healthcare data is still an open challenge, yet preliminary accomplishments are promising to bring practical privacy even closer to real-world healthcare settings. Future research should investigate how to make a trade-off between scalability, privacy, and accuracy in real healthcare settings.

Acknowledgement

This work has received funding from the European Union's Horizon2020 research and innovation program under grant agreement nr. 826078. The work of JB, HS and TK was also supported by H2020 project REPO-TRIA (nr. 777111). ML, TK, and JB have further been supported by BMBF projects SYS_Care (01ZX1908A) and SYMBOD (01ZX1910D). JB's contribution was also supported by his VILLUM Young Investigator grant (nr. 13154). This paper reflects only the author's view and the Commission is not responsible for any use that may be made of the information it contains.

References

1. Schwarting, W., Alonso-Mora, J. & Rus, D. Planning and Decision-Making for Autonomous Vehicles. *An-*

nual Review of Control, Robotics, and Autonomous Systems **1,** 187–210 (2018).

2. Gehring, J., Auli, M., Grangier, D., Yarats, D. & Dauphin, Y. N. *Convolutional sequence to sequence learning* in *Proceedings of the 34th International Conference on Machine Learning-Volume 70* (2017), 1243–1252.

3. Xiong, W. *et al. The Microsoft 2017 conversational speech recognition system* in *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (2018), 5934–5938.

4. Holzinger, A., Kieseberg, P., Weippl, E. & Tjoa, A. M. in *Springer Lecture Notes in Computer Science LNCS 11015* 1–8 (Springer, Cham, 2018).

5. Gómez-Bombarelli, R. *et al.* Automatic chemical design using a data-driven continuous representation of molecules. *ACS central science* **4,** 268–276 (2018).

6. Ma, J. *et al.* Using deep learning to model the hierarchical structure and function of a cell. *Nature methods* **15,** 290 (2018).

7. Nie, D. *et al.* Medical image synthesis with deep convolutional adversarial networks. *IEEE Transactions on Biomedical Engineering* **65,** 2720–2730 (2018).

8. Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H. & Aerts, H. J. Artificial intelligence in radiology. *Nature Reviews Cancer* **18,** 500–510 (2018).

9. Beam, A. L. & Kohane, I. S. Big data and machine learning in health care. *Jama* **319,** 1317–1318 (2018).

10. Yu, K.-H., Beam, A. L. & Kohane, I. S. Artificial intelligence in healthcare. *Nature biomedical engineering* **2,** 719–731 (2018).

11. Michael, K. Y. *et al.* Visible machine learning for biomedicine. *Cell* **173,** 1562–1565 (2018).

12. Chen, H., Engkvist, O., Wang, Y., Olivecrona, M. & Blaschke, T. The rise of deep learning in drug discovery. *Drug discovery today* **23,** 1241–1250 (2018).

13. Wainberg, M., Merico, D., Delong, A. & Frey, B. J. Deep learning in biomedicine. *Nature biotechnology* **36,** 829–838 (2018).

14. Min, S., Lee, B. & Yoon, S. Deep learning in bioinformatics. *Briefings in bioinformatics* **18,** 851–869 (2017).

15. Litjens, G. *et al.* A survey on deep learning in medical image analysis. *Medical image analysis* **42,** 60–88 (2017).

16. Shen, D., Wu, G. & Suk, H.-I. Deep learning in medical image analysis. *Annual review of biomedical engineering* **19,** 221–248 (2017).

17. Jiang, F. *et al.* Artificial intelligence in healthcare: past, present and future. *Stroke and vascular neurology* **2,** 230–243 (2017).

18. Libbrecht, M. W. & Noble, W. S. Machine learning applications in genetics and genomics. *Nature Reviews Genetics* **16,** 321–332 (2015).

19. Nemati, S. *et al.* An interpretable machine learning model for accurate prediction of sepsis in the ICU. *Critical care medicine* **46,** 547–553 (2018).

20. Teare, P., Fishman, M., Benzaquen, O., Toledano, E. & Elnekave, E. Malignancy detection on mammography using dual deep convolutional neural networks and genetically discovered false color input enhancement. *Journal of digital imaging* **30,** 499–505 (2017).

21. Veta, M. *et al.* Assessment of algorithms for mitosis detection in breast cancer histopathology images. *Medical image analysis* **20,** 237–248 (2015).

22. Shokri, R., Stronati, M., Song, C. & Shmatikov, V. *Membership inference attacks against machine learning models* in *2017 IEEE Symposium on Security and Privacy (SP)* (2017), 3–18.

23. Papernot, N., McDaniel, P., Sinha, A. & Wellman, M. P. *SoK: Security and privacy in machine learning* in *2018 IEEE European Symposium on Security and Privacy (EuroS&P)* (2018), 399–414.

24. Zhang, C., Bengio, S., Hardt, M., Recht, B. & Vinyals, O. *Understanding deep learning requires rethinking generalization* in *Proceedings of the International Conference on Learning Representations (ICLR)* (2017).

25. Shringarpure, S. S. & Bustamante, C. D. Privacy risks from genomic data-sharing beacons. *The American Journal of Human Genetics* **97,** 631–646 (2015).

26. Homer, N. *et al.* Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS genetics* **4** (2008).

27. Harmanci, A. & Gerstein, M. Analysis of sensitive information leakage in functional genomics signal profiles through genomic deletions. *Nature communications* **9,** 1–10 (2018).

28. Wang, R., Li, Y. F., Wang, X., Tang, H. & Zhou, X. *Learning your identity and disease from research papers: information leaks in genome wide association study* in *Proceedings of the 16th ACM conference on Computer and communications security* (2009), 534–544.

29. For Genomics, G. A. & Health*. A federated ecosystem for sharing genomic, clinical data. *Science* **352,** 1278–1280 (2016).

30. Zerhouni, E. A. & Nabel, E. G. Protecting aggregate genomic data. *Science* **322,** 44–44 (2008).

31. Erlich, Y. & Narayanan, A. Routes for breaching and protecting genetic privacy. *Nature Reviews Genetics* **15,** 409–421 (2014).

32. Naveed, M. *et al.* Privacy in the genomic era. *ACM Computing Surveys (CSUR)* **48,** 1–44 (2015).

33. *General Data Protection Regulation(GDPR)* `https://gdpr-info.eu/`. 2020.

34. Cohen, A. & Nissim, K. Towards formalizing the GDPR's notion of singling out. *Proceedings of the National Academy of Sciences* (2020).

35. Aziz, M. M. A. *et al.* Privacy-preserving techniques of genomic data—a survey. *Briefings in bioinformatics* **20,** 887–895 (2019).

36. Xu, J. & Wang, F. Federated Learning for Healthcare Informatics. *arXiv preprint arXiv:1911.06270* (2019).

37. Kaissis, G. A., Makowski, M. R., Rückert, D. & Braren, R. F. Secure, privacy-preserving and federated machine learning in medical imaging. *Nature Machine Intelligence* (June 2020).

38. Cho, H., Wu, D. J. & Berger, B. Secure genome-wide association analysis using multiparty computation. *Nature biotechnology* **36,** 547–551 (2018).

39. Bonte, C. *et al.* Towards practical privacy-preserving genome-wide association study. *BMC bioinformatics* **19,** 1–12 (2018).

40. Jagadeesh, K. A., Wu, D. J., Birgmeier, J. A., Boneh, D. & Bejerano, G. Keeping patient phenotypes and genotypes private while seeking disease diagnoses. *bioRxiv,* 746230 (2019).

41. Kim, M. & Lauter, K. *Private genome analysis through homomorphic encryption* in *BMC medical informatics and decision making* **15** (2015), S3.

42. Lauter, K., López-Alt, A. & Naehrig, M. *Private computation on encrypted genomic data* in *International Conference on Cryptology and Information Security in Latin America* (2014), 3–27.

43. Lu, W.-J., Yamada, Y. & Sakuma, J. *Privacy-preserving genome-wide association studies on cloud environment using fully homomorphic encryption* in *BMC medical informatics and decision making* **15** (2015), S1.

44. Zhang, Y., Dai, W., Jiang, X., Xiong, H. & Wang, S. *Foresee: Fully outsourced secure genome study based on homomorphic encryption* in *BMC medical informatics and decision making* **15** (2015), S5.

45. Kamm, L., Bogdanov, D., Laur, S. & Vilo, J. A new way to protect privacy in large-scale genome-wide association studies. *Bioinformatics* **29,** 886–893 (2013).

46. Constable, S. D., Tang, Y., Wang, S., Jiang, X. & Chapin, S. *Privacy-preserving GWAS analysis on federated genomic datasets* in *BMC medical informatics and decision making* **15** (2015), S2.

47. Zhang, Y., Blanton, M. & Almashaqbeh, G. *Secure distributed genome analysis for GWAS and sequence comparison computation* in *BMC medical informatics and decision making* **15** (2015), S4.

48. Mohassel, P. & Zhang, Y. *Secureml: A system for scalable privacy-preserving machine learning* in *2017 IEEE Symposium on Security and Privacy (SP)* (2017), 19–38.

49. Gentry, C. *Fully homomorphic encryption using ideal lattices* in *Proceedings of the forty-first annual ACM symposium on Theory of computing* (2009), 169–178.

50. Cramer, R., Damgård, I. B. & Nielsen, J. B. *Secure multiparty computation* (Cambridge University Press, 2015).

51. Shamir, A. How to share a secret. *Communications of the ACM* **22,** 612–613 (1979).

52. Jagadeesh, K. A., Wu, D. J., Birgmeier, J. A., Boneh, D. & Bejerano, G. Keeping patient phenotypes and genotypes private while seeking disease diagnoses. *bioRxiv,* 746230 (2019).

53. Kim, M., Song, Y., Wang, S., Xia, Y. & Jiang, X. Secure logistic regression based on homomorphic encryption: Design and evaluation. *JMIR medical informatics* **6,** e19 (2018).

54. Morshed, T., Alhadidi, D. & Mohammed, N. *Parallel linear regression on encrypted data* in *2018 16th Annual Conference on Privacy, Security and Trust (PST)* (2018), 1–5.

55. Shi, H. *et al.* Secure multi-pArty computation grid LOgistic REgression (SMAC-GLORE). *BMC medical informatics and decision making* **16,** 89 (2016).

56. Bloom, J. M. Secure multi-party linear regression at plaintext speed. *arXiv preprint arXiv:1901.09531* (2019).

57. Berger, B. & Cho, H. *Emerging technologies towards enhancing privacy in genomic data sharing* 2019.

58. Chialva, D. & Dooms, A. Conditionals in homomorphic encryption and machine learning applications. *arXiv preprint arXiv:1810.12380* (2018).

59. Alexandru, A. B. & Pappas, G. J. Secure Multi-party Computation for Cloud-Based Control. *Privacy in Dynamical Systems,* 179.

60. Su, D., Cao, J., Li, N., Bertino, E. & Jin, H. *Differentially private k-means clustering* in *Proceedings of the sixth ACM conference on data and application security and privacy* (2016), 26–37.

61. Abadi, M. *et al. Deep learning with differential privacy* in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (2016), 308–318.

62. Phan, N., Wang, Y., Wu, X. & Dou, D. *Differential privacy preservation for deep auto-encoders: an application of human behavior prediction* in *Thirtieth AAAI Conference on Artificial Intelligence* (2016).

63. Beaulieu-Jones, B. K. *et al.* Privacy-preserving generative deep neural networks support clinical data sharing. *Circulation: Cardiovascular Quality and Outcomes* **12,** e005122 (2019).

64. Ren, X. *et al.* LoPub: High-Dimensional Crowdsourced Data Publication With Local Differential Privacy. *IEEE Transactions on Information Forensics and Security* **13,** 2151–2166 (2018).

65. Cormode, G., Kulkarni, T. & Srivastava, D. *Marginal release under local differential privacy* in *Proceedings of the 2018 International Conference on Management of Data* (2018), 131–146.

66. Johnson, A. & Shmatikov, V. *Privacy-preserving data exploration in genome-wide association studies* in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining* (2013), 1079–1087.

67. Dwork, C., McSherry, F., Nissim, K. & Smith, A. Calibrating noise to sensitivity in private data analysis. *Journal of Privacy and Confidentiality* **7,** 17–51 (2016).

68. Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I. & Naor, M. *Our data, ourselves: Privacy via distributed noise generation* in *Annual International Conference on the Theory and Applications of Cryptographic Techniques* (2006), 486–503.

69. Nissim, K. *et al. Differential privacy: A primer for a non-technical audience* in *Privacy Law Scholars Conf* (2017).

70. Erlingsson, Ú., Pihur, V. & Korolova, A. *Rappor: Randomized aggregatable privacy-preserving ordinal response* in *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security* (2014), 1054–1067.

71. Thakurta, A. G. *et al. Learning new words* US Patent 9,594,741. 2017.

72. Beaulieu-Jones, B. K., Yuan, W., Finlayson, S. G. & Wu, Z. S. Privacy-preserving distributed deep learning for clinical data. *arXiv preprint arXiv:1812.01484* (2018).

73. Fienberg, S. E., Slavkovic, A. & Uhler, C. *Privacy preserving GWAS data sharing* in *2011 IEEE 11th International Conference on Data Mining Workshops* (2011), 628–635.

74. Uhlerop, C., Slavković, A. & Fienberg, S. E. Privacy-preserving data sharing for genome-wide association studies. *The Journal of privacy and confidentiality* **5,** 137 (2013).

75. Yu, F. & Ji, Z. Scalable privacy-preserving data sharing methodology for genome-wide association studies: an application to iDASH healthcare privacy protection challenge. *BMC medical informatics and decision making* **14,** S3 (2014).

76. Han, Z., Liu, H. & Wu, Z. *A Differential Privacy Preserving Framework with Nash Equilibrium in Genome-Wide Association studies* in *2018 International Conference on Networking and Network Applications (NaNA)* (2018), 91–96.

77. Tramèr, F., Huang, Z., Hubaux, J.-P. & Ayday, E. *Differential privacy with bounded priors: reconciling utility and privacy in genome-wide association studies* in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security* (2015), 1286–1297.

78. Vu, D. & Slavkovic, A. *Differential privacy for clinical trial data: Preliminary evaluations* in *2009 IEEE International Conference on Data Mining Workshops* (2009), 138–143.

79. Yu, F., Rybar, M., Uhler, C. & Fienberg, S. E. *Differentially-private logistic regression for detecting multiple-SNP association in GWAS databases* in *International Conference on Privacy in Statistical Databases* (2014), 170–184.

80. Honkela, A., Das, M., Nieminen, A., Dikmen, O. & Kaski, S. Efficient differentially private learning improves drug sensitivity prediction. *Biology direct* **13,** 1 (2018).

81. Han, Z., Lu, L. & Liu, H. *A Differential Privacy Preserving Approach for Logistic Regression in Genome-Wide Association Studies* in *2019 International Conference on Networking and Network Applications (NaNA)* (2019), 181–185.

82. Simmons, S. & Berger, B. Realizing privacy preserving genome-wide association studies. *Bioinformatics* **32,** 1293–1300 (2016).

83. Simmons, S., Sahinalp, C. & Berger, B. Enabling privacy-preserving GWASs in heterogeneous human populations. *Cell systems* **3,** 54–61 (2016).

84. Wang, S., Mohammed, N. & Chen, R. Differentially private genome data dissemination through top-down specialization. *BMC medical informatics and decision making* **14,** S2 (2014).

85. Wan, Z., Vorobeychik, Y., Kantarcioglu, M. & Malin, B. Controlling the signal: Practical privacy protection of genomic data sharing through Beacon services. *BMC medical genomics* **10,** 39 (2017).

86. Al Aziz, M. M., Ghasemi, R., Waliullah, M. & Mohammed, N. Aftermath of bustamante attack on genomic beacon service. *BMC medical genomics* **10,** 43 (2017).

87. Kairouz, P., Oh, S. & Viswanath, P. The composition theorem for differential privacy. *IEEE Transactions on Information Theory* **63,** 4037–4049 (2017).

88. Abay, N. C., Zhou, Y., Kantarcioglu, M., Thuraisingham, B. & Sweeney, L. *Privacy preserving synthetic data release using deep learning* in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (2018), 510–526.

89. Jordon, J., Yoon, J. & van der Schaar, M. *PATE-GAN: Generating Synthetic Data with Differential Privacy Guarantees* in (2018).

90. Fiume, M. *et al.* Federated discovery and sharing of genomic data using Beacons. *Nature biotechnology* **37,** 220–224 (2019).

91. Raisaro, J. L. *et al.* Protecting privacy and security of genomic data in I2B2 with homomorphic encryption and differential privacy. *IEEE/ACM transactions on computational biology and bioinformatics* **15,** 1413–1426 (2018).

92. Hardt, M., Ligett, K. & McSherry, F. *A simple and practical algorithm for differentially private data release* in *Advances in Neural Information Processing Systems* (2012), 2339–2347.

93. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics* **38,** 904–909 (2006).

94. Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M. & Price, A. L. Advantages and pitfalls in the application of mixed-model association methods. *Nature genetics* **46,** 100 (2014).

95. Wang, S. *et al.* Genome privacy: challenges, technical approaches to mitigate risk, and ethical considerations in the United States. *Annals of the New York Academy of Sciences* **1387,** 73 (2017).

96. Kieseberg, P., Hobel, H., Schrittwieser, S., Weippl, E. & Holzinger, A. in *Interactive Knowledge Discovery and Data Mining in Biomedical Informatics, Lecture Notes in Computer Science, LNCS 8401* (eds Holzinger, A. & Jurisica, I.) 301–316 (Springer, Berlin Heidelberg, 2014).

97. McMahan, H. B., Moore, E., Ramage, D., Hampson, S., *et al.* Communication-efficient learning of deep networks from decentralized data. *arXiv preprint arXiv:1602.05629* (2016).

98. Malle, B., Giuliani, N., Kieseberg, P. & Holzinger, A. in *Machine Learning and Knowledge Extraction, IFIP CD-MAKE, Lecture Notes in Computer Science LNCS 10410* 367–374 (Springer, Cham, 2017).

99. Kairouz, P. *et al.* Advances and open problems in federated learning. *arXiv preprint arXiv:1912.04977* (2019).

100. Yang, Q., Liu, Y., Chen, T. & Tong, Y. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)* **10,** 1–19 (2019).

101. Pan, S. J. & Yang, Q. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* **22,** 1345–1359 (2009).

102. Caruana, R. Multitask learning. *Machine learning* **28,** 41–75 (1997).

103. Zhang, Y. & Yang, Q. A survey on multi-task learning. *arXiv preprint arXiv:1707.08114* (2017).

104. Holzinger, A., Haibe-Kains, B. & Jurisica, I. Why imaging data alone is not enough: AI-based integration of imaging, omics, and clinical data. *European Journal of Nuclear Medicine and Molecular Imaging* **46,** 2722–2730 (2019).

105. Sheller, M. J., Reina, G. A., Edwards, B., Martin, J. & Bakas, S. *Multi-institutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation* in *International MICCAI Brainlesion Workshop* (2018), 92–104.

106. Chang, K. *et al.* Distributed deep learning networks among institutions for medical imaging. *Journal of the American Medical Informatics Association* **25,** 945–954 (2018).

107. Balachandar, N., Chang, K., Kalpathy-Cramer, J. & Rubin, D. L. Accounting for data variability in multi-institutional distributed deep learning for medical imaging. *Journal of the American Medical Informatics Association* (2020).

108. Nasirigerdeh, R. *et al.* sPLINK: A Federated, Privacy-Preserving Tool as a Robust Alternative to Meta-Analysis in Genome-Wide Association Studies. <`https://www.exbio.wzw.tum.de/splink/`> (2020).

109. Wu, Y., Jiang, X., Kim, J. & Ohno-Machado, L. Grid Binary LOgistic REgression (GLORE): building shared models without sharing data. *Journal of the American Medical Informatics Association* **19,** 758–764 (2012).

110. Wang, S. *et al.* EXpectation Propagation LOgistic REgRession (EXPLORER): distributed privacy-preserving online model learning. *Journal of biomedical informatics* **46,** 480–496 (2013).

111. Li, Y., Jiang, X., Wang, S., Xiong, H. & Ohno-Machado, L. Vertical grid logistic regression (vertigo). *Journal of the American Medical Informatics Association* **23,** 570–579 (2016).

112. Brisimi, T. S. *et al.* Federated learning of predictive models from federated electronic health records. *International journal of medical informatics* **112,** 59–67 (2018).

113. Huang, L. *et al.* Loadaboost: Loss-based adaboost federated machine learning on medical data. *arXiv preprint arXiv:1811.12629* (2018).

114. Liu, D., Miller, T., Sayeed, R. & Mandl, K. D. Fadl: Federated-autonomous deep learning for distributed electronic health record. *arXiv preprint arXiv:1811.11400* (2018).

115. Chen, Y., Wang, J., Yu, C., Gao, W. & Qin, X. FedHealth: A Federated Transfer Learning Framework for Wearable Healthcare. *arXiv preprint arXiv:1907.09173* (2019).

116. Smith, V., Chiang, C.-K., Sanjabi, M. & Talwalkar, A. S. *Federated multi-task learning* in *Advances in Neural Information Processing Systems* (2017), 4424–4434.

117. The TFF Authors. *TensorFlow Federated* 2019. <https://www.tensorflow.org/federated>.

118. Ryffel, T. *et al.* A generic framework for privacy preserving deep learning. *arXiv preprint arXiv:1811.04017* (2018).

119. The FATE Authors. *Federated AI technology enabler* 2019. <https://www.fedai.org/>.

120. The PaddleFL Authors. *PaddleFL* 2019. <https://github.com/PaddlePaddle/PaddleFL>.

121. The FeatureCloud Authors. *FeatureCloud* 2019. <https://featurecloud.eu/>.

122. The Clara Training Framework Authors. *NVIDIA Clara* 2019. <https://developer.nvidia.com/clara>.

123. Vepakomma, P., Gupta, O., Swedish, T. & Raskar, R. Split learning for health: Distributed deep learning without sharing raw patient data. *arXiv preprint arXiv:1812.00564* (2018).

124. Vepakomma, P., Gupta, O., Dubey, A. & Raskar, R. Reducing leakage in distributed deep learning for sensitive health data. *arXiv preprint arXiv:1812.00564* (2019).

125. Poirot, M. G. *et al.* Split Learning for collaborative deep learning in healthcare. *arXiv preprint arXiv:1912.12115* (2019).

126. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics* **81,** 559–575 (2007).

127. Silva, S. *et al. Federated learning in distributed medical databases: Meta-analysis of large-scale subcortical brain data* in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)* (2019), 270–274.

128. Pollard, T. J. *et al.* The eICU Collaborative Research Database, a freely available multi-center database for critical care research. *Scientific data* **5,** 180178 (2018).

129. Smith, V., Chiang, C.-K., Sanjabi, M. & Talwalkar, A. S. *Federated multi-task learning* in *Advances in Neural Information Processing Systems* (2017), 4424–4434.

130. Corinzia, L. & Buhmann, J. M. Variational federated multi-task learning. *arXiv preprint arXiv:1906.06268* (2019).

131. Liu, Y., Chen, T. & Yang, Q. Secure federated transfer learning. *arXiv preprint arXiv:1812.03337* (2018).

132. Gupta, S., Agrawal, A., Gopalakrishnan, K. & Narayanan, P. *Deep learning with limited numerical precision* in *International Conference on Machine Learning* (2015), 1737–1746.

133. Aji, A. F. & Heafield, K. *Sparse Communication for Distributed Gradient Descent* in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (2017), 440–445.

134. McMahan, H. B., Moore, E., Ramage, D., Hampson, S., *et al.* Communication-efficient learning of deep networks from decentralized data. *arXiv preprint arXiv:1602.05629* (2016).

135. Li, W. *et al. Privacy-preserving Federated Brain Tumour Segmentation* in *International Workshop on Machine Learning in Medical Imaging* (2019), 133–141.

136. Li, X. *et al.* Multi-site fMRI Analysis Using Privacy-preserving Federated Learning and Domain Adaptation: ABIDE Results. *arXiv preprint arXiv:2001.05647* (2020).

137. Choudhury, O. *et al.* Differential Privacy-enabled Federated Learning for Sensitive Health Data. *arXiv preprint arXiv:1910.02578* (2019).

138. Lee, J. *et al.* Privacy-preserving patient similarity learning in a federated environment: development and analysis. *JMIR medical informatics* **6,** e20 (2018).

139. Kim, M., Lee, J., Ohno-Machado, L. & Jiang, X. Secure and Differentially Private Logistic Regression for Horizontally Distributed Data. *IEEE Transactions on Information Forensics and Security* **15,** 695–710 (2019).

140. Tang, Z., Shi, S., Chu, X., Wang, W. & Li, B. Communication-Efficient Distributed Deep Learning: A Comprehensive Survey. *arXiv preprint arXiv:2003.06307* (2020).

141. Zhao, Y. *et al.* Federated learning with non-iid data. *arXiv preprint arXiv:1806.00582* (2018).

142. Li, Q., Wen, Z. & He, B. Federated learning systems: Vision, hype and reality for data privacy and protection. *arXiv preprint arXiv:1907.09693* (2019).

143. Rieke, N. *et al.* The Future of Digital Health with Federated Learning. *arXiv preprint arXiv:2003.08119* (2020).

144. Melis, L., Song, C., De Cristofaro, E. & Shmatikov, V. *Exploiting unintended feature leakage in collaborative learning* in *2019 IEEE Symposium on Security and Privacy (SP)* (2019), 691–706.

145. Geyer, R. C., Klein, T. & Nabi, M. Differentially private federated learning: A client level perspective. *arXiv preprint arXiv:1712.07557* (2017).

146. Truex, S. *et al. A hybrid approach to privacy-preserving federated learning* in *Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security* (2019), 1–11.

147. Sanders, S. F., Terwiesch, M., Gordon, W. J. & Stern, A. D. How Artificial Intelligence Is Changing Health Care Delivery. *NEJM Catalyst* **5** (2019).

148. Berger, B. & Cho, H. *Emerging technologies towards enhancing privacy in genomic data sharing* 2019.