World Journal of Advanced Engineering Technology and Sciences
eISSN: 2582-8266
Cross Ref DOI: 10.30574/wjaets
Journal homepage: https://wjaets.com/

(REVIEW ARTICLE)

Check for updates

Federated learning and differential privacy in clinical health: Extensive survey

David Odera *

Tom Mboya University, Computer Science & Information Technology, P. O. Box 199-40300, Homa-Bay, Kenya.

World Journal of Advanced Engineering Technology and Sciences, 2023, 08(02), 305–329

Publication history: Received on 01 March 2023; revised on 08 April 2023; accepted on 11 April 2023

Article DOI: https://doi.org/10.30574/wjaets.2023.8.2.0113

Abstract

Federated Learning (FL) is concept that has been adopted in medical field to analyze data in individual devices through aggregation of machine learning model in global server. It also provides data privacy being that the sampled devices are not allowed to share data among themselves. Therefore, it minimizes computation costs and privacy risks to some extent compared to conventional methods of machine learning. However, federation learning provides a different use case in health as compared to other sectors. Preservation of patients' sensitive information such as electronic health record (EHR) when sharing data among different medical practitioners is of greatest concern. So the question is, how should FL techniques be structured in the current clinical environment where heterogeneity is the order of the day? The EU's General Data Protection Regulation (GDPR) and Health Insurance Portability and Accountability Act of 1996 (HIPPA) regulations recommends health providers to gain authorizations from patients before sharing their private data for medical analytical progression. This leads to some bottlenecks in clinical analysis. Although attempts have been made to address some of the challenges, privacy, performance, implementation, computation and adversaries still pose some threats. This paper provides a comprehensive review that covers literature, mathematical notations, architecture, process flow, challenges and frameworks used to implement FL with respect to healthcare. Possible solutions on how to address privacy challenges in accordance with HIPPA act and GDPR is discussed. Finally, the study gives future direction of FL in clinical health and a list of practical tools to conduct analysis on patients' data.

Keywords: FL; DPSGD; FedAvg; FedProx; RPC; CL-DP; MNIST

1. Introduction

In medical, federated learning (FL) use-case is inherently different from other domains [1]. The models are bigger and with less participants in terms of number of clients. The original domain of Federated Learning (FL) is smart phones [2] [3], which accommodates millions of participants in form of mobile phones unlike in healthcare where you would expect to see a number of hospitals willing to participate [1]. Artificial intelligence (AI) has really transformed the medical field [4],[5]enormously in the last couple of years. Precisely, the advances has been on the areas of machine learning, specifically deep learning which has led to disruptive innovation not only in radiology (CT, MRI, ultrasound or X-ray) but also pathology genomics, dermatology or a completely different data types such as electronic health records (EHR) [1]. Based on this analysis, it appears that across the fields in health domain data crunching is gaining insights with aid of AI innovated medical field. Many applications have been developed which have improved performances and accuracy as well [5], [6], [7].

According to the authors in [8], [9], [10], [11] there are thousands of research papers that leverage machine learning in different fields in medicine. AI has achieved human-level performance on large data and clinical settings [1]. For instance, dermatologist level classification of deep neural networks used by [12] to detect skin cancer among 2032 diseases using 129450 clinical images. A real-time artificial intelligence gastrointestinal cancer detection introduced in [13] reported similar performance to that of endoscopist in dataset consisting of6 hospitals, 84424 individuals and

*Corresponding author: David Odera

1036496 endoscopy images. The authors in [14] and [15] have developed a "Similar image search for Histopathology", while the authors in [16] also introduced a clinically applicable deep learning for diagnosis and referral in retinal disease. This is an indication that AI has contributed towards the development of healthcare innovations especially with respect to data mining.

All these research advances and techniques demands some level of datasets for adequate training [17]. Therefore, the era of data driven medicine that can train robust and accurate deep learning models is here with us. These kind of data is diverse and very large but it has to represent the problem that a particular health domain has to solve. As been alluded by [1], the current research in machine learning is driven by data lakes that centrally places required data for experimentation in a GPU infrastructure [18]. Perhaps due to the complexity of creating a centralized database where data can be publicly accessed for learning especially within medical domain. However, a lot of research is not evident in scenarios where data cannot be accessed through data lakes (think about big data, IoT, for example). Researchers in [19] and [20] have discussed on the emergence of data from pervasive computing and machine learning in small. In clinical health, gathering data in centrally is a highly sensitive issue [1], and it's also subject to regulations such as Health Insurance Portability and Accountability Act (HIPPA) [21][22] and General Data Protection Regulation ('GDPR')[23][24]. These regulations make it almost impossible to share patients' data in a centralized data lake. These regulations are placed for good reasons [1], which relates to ethics, data privacy, data protection and technical ones as well [21], [23], [24].Therefore, the big question is how to enable a data driven analysis that respects the highly sensitive nature [25] of health data and develops machine-learning models that avoids demographic biasness (i.e healthcare in remote areas).

2. Nature of Dataset in Healthcare

Apart from publicly available MIMIC-III dataset [26] developed in Massachusetts Institute of Technology (MIT)'s Laboratory for clinical data of patients admitted at the Beth Israel Deaconess Medical Center (BIDMC) in Boston, Massachusetts during 2001 to 2012 [27].This article also states the following examples of large datasets that can be used for robust models in healthcare as shown in Table 1.

Table 1 Healthcare datasets

Dataset	Description	Category
NHS Scotland's National Safe Have	EHR and Health Data Research UK that covers a number of regional hubs in Scotland [28] & [29]	Federated
French health data hub	French National Health data by [30] , [31]	Centralized
UK Biobank	Presents imaging, genetics, EHR, biomarkers, activity monitoring datasets by [31]	Centralized
Cancer Imaging Archive (TCIA)	Contains data related to images for research by [32]	Centralized
TCIA COVID-19 Datasets	COVID19 related complications images [33] & [34]	Centralized
Cancer Genome Atlas (TCGA)	Contains large cohorts of about 30 human tumors through genome [35]	Centralized
Medical Segmentation Decathlon	biomedical image [36]	Centralized

All these data efforts are great and they foster research and insights in health related issues [37]. That notwithstanding, it becomes very complex to model a clinical system that can integrate databases that are highly fragmented as it is now. Federated learning can be used to circumvent the problem by bringing in the algorithm to patient data (in this case hospital) and only aggregating intermediate model trainings for update of the global server [37], [38]. This will drastically reduce the risk of privacy violations through data communication, centralized and cloud storage [40].

Interestingly, it's still possible for offenders to gain access to private data within individual devices and institutional databases through reconstruction attacks or inference attack [41]. Federated averaging (*FedAvg*) and differential privacy (DP) is the most popular and widely acceptable way of resolving these challenges described above [40].

This paper presents a survey of federated averaging and differential privacy with a focus on resolving privacy and communication challenges in clinical health. The specific contributions of this article are as follows:

- Discussion on various research work on federated learning including improvements on FL
- Assessment of Differential Privacy (DP) mechanisms on reduction of privacy loss
- Analyze some of the bottlenecks that federated learning faces with regards to healthcare
- Assess existing works in federated and differential privacy as well as identifying contributions, methods and gaps that exits

The rest of this paper is organized as follows: Part 2 presents the nature of dataset in healthcare, while Part 3 describes federated averaging. On the other hand, differential privacy is discussed in Part 4, while the current frameworks and techniques are discussed in Part 5. In addition, Part 6 presents some of the frameworks for federated learning, while the technologies for big data security are discussed in Part 7. Similarly, research gaps are explained in Part 8 while the conclusions are described in Part 9.

3. Federated Averaging

According to [2], Federated Averaging (*FedAvg*) is most widely accepted and recognized algorithm for conducting federated learning. The author in [2] states that, *FedAvg* is good in practice but not perfect due to a number of simple assumptions. For instance, it seems like all sample devices will complete in ∈Epochs of local stochastic gradient descent but some devices take longer than others do. Slower devices are known as stragglers which tend to affect speed of convergence. Therefore, *FedAvg* may just opt to drop these stragglers. The question that [2] is posing is what then happens when 90% of devices are stragglers. *FedAvg* weighs devices by proportion of data they hold so it might favor certain devices performances in expense of others.

According to [42] the learning is conducted at the machine so that individual databases do not share data among themselves. The model given by M_{fed} executes in every machine then collaboratively combines them in a trusted server machine. In every machine, there exists data D_i that is never leaks to other data owners F_i [40]. The federated learning systems accuracy given by V_{Fed} should not be far from the performance of aggregated model M_{SUM} and sum of accuracy V_{SUM} [42]. As shown in eq. (1), the loss calculated by finding the difference between sum of accuracies and federated accuracy, if that difference is less than a non-negative number δ [1], [2], [40], [42].

$$| V_{\text{FED}} - V_{\text{SUM}} | < \delta \quad \text{.....eq. (1)}$$

The authors in [1] defined a general federated learning by denoting a global loss function as L which is obtained by combination of weighted K local losses $\{L_k\}$, calculated at individual machine X_K as shown in eq. (2)

$$\min_{\phi} L(x; \phi) \text{ with } L(x; \phi) = \sum_{k=1}^{K} w_k l_k(X_K; \phi) \quad \text{.....eq. (2)}$$

The diagram above illustrates three parties (medical institutions) with private database and federated server that collaboratively trains without centralizing datasets. It addresses privacy and data governance challenges. In position paper written by [1], this FL would create more opportunities by enabling precision medicine at large scale, novel research (e.g. in rare diseases) and medical data will not be duplicated. The concept above models communication using FL for health institutions, which consist of few clients, no control of private institutions datasets and those datasets, are large in terms of volume [1], [44], [44]. As affirmed to the writers [45], [46], [47] and [48] there are many flavors of Federated learning which include the following

- Centralized; there is no collaboration and aggregation of updates from individual private datasets. There exist a central data lakes which trains a pool of data from individual datasets. The models and individual machine communicate through API when requesting for resources available [48], [49].
- On-device inference: Initial communication between devices and cloud is used to send model to each device. Every device will then build its own learning model thereafter, no further communication with cloud is necessary [48]. Its characteristics include many devices, no control on how data is split and devices contains less data volumes [45].

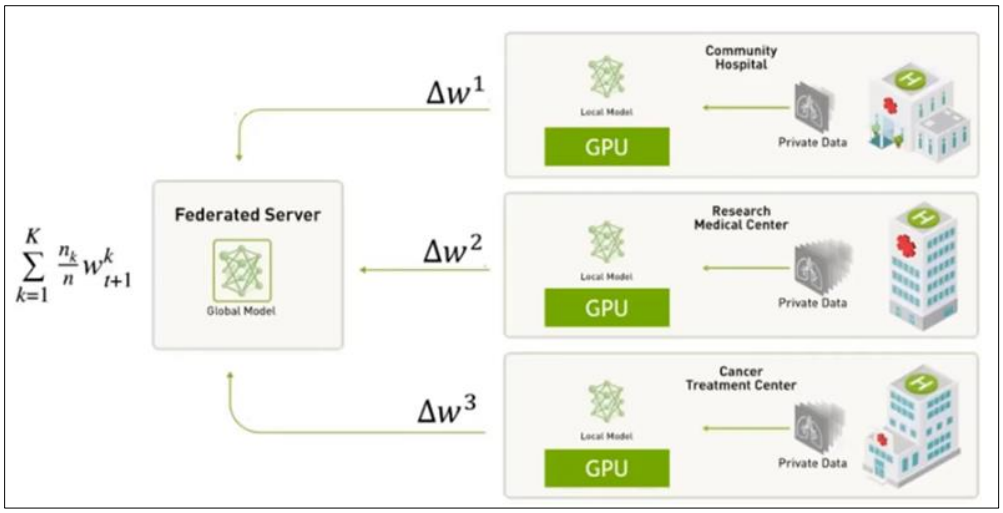

Figure 1 A simple federated learning design [1]

- Aggregation Server: In this case, every individual machine will train a model from the global server, then sends its updates to server for aggregation with other updates. The data is retained in specific machines and knowledge is only shared through aggregation with global server [43], [44], [48].
- Peer-to-peer: uses blockchain technology to enable nodes update models among themselves without sharing data in a distributed environment [49].

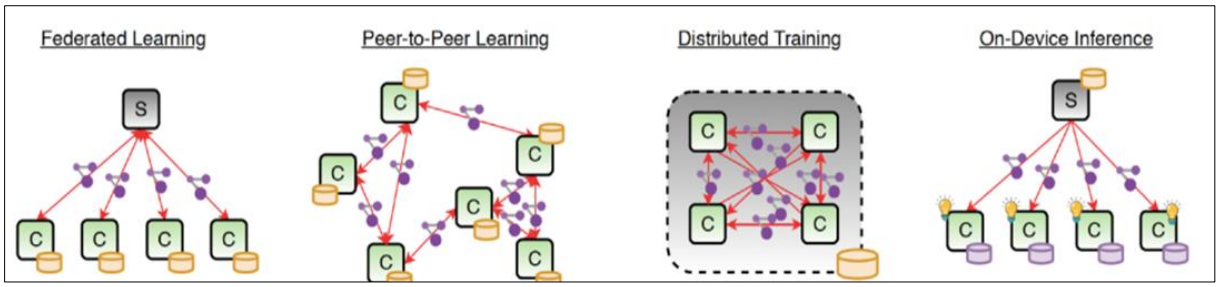

Figure 2 Illustration of various model learning architectures [1], [46], [47], [48], [49], [50], [52], [53]

So, there are many different combinations of FL but the bottom line characteristics are, that local datasets are distributed and there exist collaborative model.

3.1. Federation optimization in Heterogeneous Networks

This approach helps devices to do variable amount of work [29], [39], [50], [51], [52], [53]. Naively you might think this saves devices that can run more set of gradient descent at the same favorable amount of time and therefore change the weight model much more. So *FedProx* algorithm [50] [54] introduces a regularization time

$$\min_{\omega} h_k(\omega; \omega^t) = F_K(\omega) + \frac{\mu}{2} \| \omega - \omega^t \|^2 \ldots\ldots\ldots\ldots\ldots\ldots \text{eq. (3)}$$

where ω is local updated weight for client k, ω^t is shared model's weight at current round t, Proximal term that penalizes large changes in weights this also help convergence in, proximal term penalized model from changing too much on one single device , $\frac{\mu}{2}$ is hyper parameter for tuning [55],[56].

The challenge of this federated optimization is communication and heterogeneity [50] where data in multiple devices vary a lot as different distribution requires different features, in order to fix this *q-FedAvg* [57] is used.

3.2. Fair Resource Allocation in Federated Learning

Here, shared model learn a lot more fair by performing similarly on all devices [57], [58].

$$\min f(w)=\sum_{i=1}^{m} p_k F_k (w) \text{ ..... eq. (4)}$$

where P_k is proportion of overalldata belonging to client k. So rather than weighting devices by proportion of data they have, it penalized worst performing device more. It is the sensitive model to improve performance on these devices, *q-FedAvg* [57].

$$\min_{\omega} f_g (w) = \sum_{i=1}^{m} \frac{P_k}{q+1} F_k^{q+1} (w), \text{..... eq. (5)}$$

Here the loss (F_k^{q+1}) is raised to power *q+1* .This is tuned so that the larger the q, the more this worst performing client dominate average loss, the more fairer it becomes.

3.3. Personalized Federated Learning

This seek to train model that can be personalized to each training device after running few sets of local stochastic gradient descent [59], [60], [61],[62], [63],[64] [65]. So, the loss function change this current weight as shown.

$$\min_{\omega} f(w)=\sum_{k=1}^{m} p_k F_k (w) \text{..... eq. (6)}$$

$$\min_{\omega} f(w)=\sum_{k=1}^{m} p_k F_k (w-\alpha\nabla F_k(w)) \text{ .....eq. (7)}$$

Where F_k (w-α∇Fk(w)) is the weight after one step of gradient descent. This uses Model-Agnostic Meta-Learning MAML approach to formulate federated learning as a multitask problem where each client device distribution is separate task. This concept is known as federated multi-task learning by [66].

3.4. Federated multi-task learning

It is limited to linear models or lose ability to model complex relationships [66], [67], [68]. This concept teaches machine learning how to do multiple thing at the sometime [66]. Many things can be done using neural networks and machine learning models such as image classification, object detection, super-resolution, and text generation and so on. Typically, a model trained to do a single task, which is convenient, but you may want to use a single model to solve multiple problems for various reasons, such as efficiency, better generalization [66]. To improve efficiency, you can share some of the layers between different but related tasks [69]. For example to classify a scene, you can detect objects in it and output a segmentation mask [70]. So the question is, do you need to train these three tasks separately? An understanding needs to be reached so that resultant model can potentially save memory, computation time [71] and energy by developing a unified multitasking model.

Multitask learning can be defined as interpretation of each data set and the task of fitting parameters of each local model to local data set as learning task [72]. So sourcing all these tasks at once and putting them in a single model is multi-tasking.

4. Differential Privacy

Many of the machine learning models require access to private data [63], [73], [74], [75], [76]. The issue is the models have tendency to memorize these private data even if they are not over-fitting [77], [78]. A technique that is used to preserve privacy in federated learning is known as differential privacy (DP) [74], [75], [76]. Privacy is complex and as Netflix found when announcing a recommender challenge [82], simply anonymizing (process of replacing all private IDs with random identification) [77], [78] your dataset is not enough. Researchers found out that you can take these dataset and link across entries in order to de-anonymize individuals [79], [80].Therefore, anonymization still is not enough in preserving privacy [81].

According to authors in [82], every time a query is sent to database for some sort of statistical analysis, we are leaking some set of information about dataset. When they [82] released an anonymized aggregated heat map of location from across the world, it found that they actually released sensitive data about military. So how can model learn from general trends of dataset without revealing individual's private information? This is the idea behind differential privacy [82]. It is possible to build a more intuitive notion of privacy loss and define a mathematical definition for privacy in a dataset called differential privacy. DP analyses any process that checks data and produces output such as database query, training model etc[63].

4.1. Plausible deniability

The problem of plausible deniability [83] is attributed to learning models, which are exposed to more targeted and personal data that has higher chances of prediction outcome [52], [84], [85], [86], [87], [88]. Every time a machine-learning model learns a particular dataset by adding or removing a personal record, the outcome becomes certainly likely however, leakage of information about a particular data record is alive and well [86]. Deep learning can be used to limit privacy loss [73], [89], [90]. The model that is being used should therefore make a similar prediction whether or not there is addition or removal of data.

The key idea about DP is to limit privacy loss, which bounds to privacy budget (epsilon∈). The smaller the budget the more the privacy and the larger the budget the more it learns. Privacy loss mechanism is given by [91], [92]:

$$\text{Log}\frac{P\ (M(D) \in S)}{P(M(D') \in S)} \text{.....eq. (8)}$$

Where *M(D)*, is outcome of learned dataset that differ in one record and *s* is probability for set of outcomes.

$$\text{Log}\frac{P\ (M(D) \in S)}{P(M(D') \in S)} \leq \in \text{, ...eq. (9)}$$

where ∈, is considered privacy budget, bound to episolon. A smaller budget is more private while a larger budget need more learning. Therefore there should be some trade-offs between privacy and learning of data, therefore this inequality can be rearranged using some traditional mission of differential privacy.

$$P\ (M(D)\ \in\ S) \leq e\ P^{\in}\ (M(D')\ \in\ S) \text{ ..... eq. (10)}$$

In practice, epsilon delta ($\in -\delta$) is defined differentially private where delta δ is a failure probability and so long as this probability is much less than any particular individual's probability it won't affect privacy but allows proof of balance much more easily.

$P\ (M(D)\ \in\ S) \leq e\ P^{\in}(M(D')\ \in\ S) + \delta$. This concept prevents leakage of information in practice [92].

4.2. Privacy by addition of noise

This concept of privacy can be achieved by adding noise to the outputs [73]. So think of an image, the more blurry the image the less you can tell about it and the more it requires addition of noise.

4.3. Privacy Amplification Theorem

Another technique to reduce loss is sampling .Rather than looking at the dataset you may sample a fraction q then an ($\in -\delta$)-differentially private mechanism M becomes ($q\ \in -q\delta$)-differentially private [93].

4.4. Fundamental Law of Information Recovery

It states that privacy can actually be eroded by asking enough overly accurate questions [89]. Therefore the more queries you may have the larger the privacy loss [94]. This is represented in "Composition in Differential Privacy" [93]. When ($\in_1 - \delta_1$)-differentially private mechanism is run followed by an ($\in_2 - \delta_2$)-differentially private mechanism, the whole thing turns out to be ($\in_1 + \in_2$)-($\delta_2 + \delta_2$)-differentially private.

Deep learning with differential privacy modifies Stochastic Gradient Descent to become differentially private and the algorithm is known as "Differentially-Private Stochastic Gradient Descent (DP-SGD)" [95]. The aim of the algorithm is to limit privacy loss per gradient update. So rather than updating with raw gradient, clip gradient (first gradient) as a maximum gradient of C (add noise is proportional to clipping) [95]. This intuitively limits the amount of information that model is learning from any given example. Then noise is added, so the sample is from Gaussian distribution [96] and standard deviation of C sigma. As shown in the algorithm, high parameters C and sigma can be tuned to give epsilon delta guarantees for each step of gradient descent. This is individual type of gradient descent technique.

Algorithm : Differential Privacy with SGD

Input: Examples {x_1.....,x_n}, loss function $L(\theta) = \frac{1}{N}\sum_i L(\theta, x_i)$.

Learning rate μ_t , noise scale σ , group size L, gradient norm bound C.

Randomly Initializeθ_0
For t ∈ $[T]$**do**
Get random sample L_t with probability sampling L/N
Calculate gradient
For each i∈ L_t compute $g_t(x_i) \leftarrow \nabla_{\theta t} L(\theta_t x_i)$
Clip gradient
$\bar{g}_t(x_i) \leftarrow \boldsymbol{g_t(x_i)/max}(\mathbf{1}, \frac{||g_t(x_i)||_2}{C})$
Add noise
$\tilde{g}_t \leftarrow \frac{1}{N}(\sum_i \tilde{g}_t(x_i) + N(0, \sigma^2 C^2 \boldsymbol{I}))$
Descent
$\theta_{t+1} \leftarrow \theta_t - \mu_t \tilde{g}_t$
Output
$\boldsymbol{\theta_T} then find the overall privacy cost (\varepsilon, \delta) by use of accounting method$

DP SGD algorithm by [95]

As proposed and evaluated in [63], asymptotic convergence rate is provided as follows O(1/ √ nτT) + O(τσ2/T), where number of devices is given by n, number of communication cycles is T, local iteration is $\tau$and σ^2 is the variance of Gaussian noise that is added for gradients at every local iteration.

In order to track the overall gradient descent, this article studied the papers [96], [97] that describes moments accountant [95] for tracking privacy budget. First they looked at naïve analysis of adding privacy budgets [95]. Since each step is epsilon delta $(\in -\delta)$, but with sampling of mini batch with some probability q, so by the privacy amplification theorem each step is actually $(q \in -q\delta)$-differentially private. If this is repeated in t epochs then all these can be added together so the overall algorithm will be (T ∈ −Tδ)-differentially private. This is quiet loosely bound and there is no strong composition theorem that states you can get a tighter bound of (O(q∈√ (Tlog(1/δ))) − Tqδ) − differentialy private [95]. It actually improve the budget rather than it being proportional to T,its proportional to root T (√(Tlog(1/δ) therefore it can run longer number of epochs. The accounting technique makes it better by bounding it down to O (q∈√T) so this save a factor of log over delta (log(1/δ) and failure probability then becomes δ instead of Tqδ. The key insight in moments technique is that the privacy loss being calculated is a random variable on its own and can be plot in a distribution which will have a long tail. At any point the privacy loss can be clipped to decide privacy budget and a failure probability delta is a bound on the probability of that long tail for values greater than epsilon. So the moments accounts each track of all these bounds the trainings associated with each of the moments and picks the tightest bound [95] The privacy will be preserved as long as the offender gets the sum of machine models are differentially private [98], [99]. As discussed in [98] the approach allows individual machines to upload their encrypted messages to server as the global server decrypts the aggregated messages through secure aggregation protocol.

5. Current frameworks and techniques

Table 2 below contains some of the research works that largely contributes to the literature domain in relation to FL and DP in healthcare.

6. Frameworks of Federated Learning

To implement FL, the following are some of the frameworks you may use [1]:

- Tensorflow Federated (TFF): Machine Learning on Decentralized Data [141]
- PySyft from the open community Open-minded Developed by Open mind community [142]
- Flower from University of Cambridge [143]
- Federated AI Technology Enabler (FATE) from Webank's AI department [144];
- Paddle Federated Learning (PFL) from Baidu [145];
- Federated Learning and Differential Privacy (FL&DP) framework from Sherpa.AI [146].

Table 3 Presents some of the cureent models developed over the recent past

Table 2 Current frameworks and techniques

Framework/technique	Contribution	Challenge/Gap	Method	Category
Privacy preservation in federated learning using gradient perturbation, secure aggregation, and zero-concentrated differential privacy (zCDP)[63]	Propose marginal degradation and perturbation for model utility and loss reduction through Gaussian noise in order to build a novel federated learning model without fully trusted server.	Data pollution attack by local device may supply wrong dataset it cannot prevent attackers who infer private messages from untrusted servers (eavesdropping) [100][101] Analyze the design of secure aggregation rather than optimizing the design They consider their future work to be on performance of learning in other areas where multi-task learning and privacy considerations are paramount	Gaussian noise [95] Encryption and decryption for aggregated protocol [98] pseudorandom function [102] (PRF) for reduction of overhead of protocol during rounds	Framework
Clipping for Federated Learning: Convergence and Client-Level Differential Privacy[46]	Analyzed use of clipping operation in client model Empirical study to show performance of clipping-enabled *FedAvg* Provide relation between client's learning update and clipping bias	Effect of clipping operation on performance of FL is not certain How to create a balance when adding noise to FL algorithm and CL-DP [93]	Sample-level differential privacy (SL-DP) [103], fits in cross-silo Federated Learning which consist of a smaller number of clients, where every client has a large dataset Client-level differential privacy (CL-DP) [103], suitable for cross-device like Google keyboard where bigger distribution of client is necessary	Framework
Privacy-Preserving Federated Brain Tumour Segmentation [104]	Implement privacy-preserving federated learning system for image analysis Compare Federated algorithm in handling momentum optimization and imbalanced trainings	Assume the dataset at local clients are fixed (result in over-fitting) Indicate exploration of differentially private SGD in image analysis as future work [95]	Deep Neural Networks (DNN) for global server node SGD at Local node for privacy preservation Selective parameter sharing, prevents over-fitting by	Technique

	Study sparse vector technique for DP		limiting data that a client can share through clipping threshold. sparse vector technique (SVT) to protect indirect data leakage [104], [105] BraTS 2018 dataset [106]	
RR-LADP: A Privacy-Enhanced Federated Learning Scheme for Internet of Everything[107]	Used randomized response (RR) mechanism to select clients whose data should be learned by model for server Local adaptive differential privacy (LADP) mechanism to add Gaussian noise in every clients update before sending to server	It has considered only accuracy as performance metric Apply the framework in wearable devices and Internet of Vehicles in future	Randomized response mechanism Federated averaging Local adaptive differential privacy (LADP) batch gradient descent (BGD) to optimize loss function MNIST Dataset [108]	Framework
Concentrated Differentially Private Federated Learning With Performance Analysis by [63]	Proposed a periodic averaging with a device sampling without a fully trusted server for private differential FL Use zero-concentrated differential privacy (zCDP) to for end-to-end privacy loss	Trade-offs between privacy and utility considers only accuracy by setting number of iteration and communication rounds in logistic regression and neural networks, other metrics no considered	Uses *FedAvg* and zCDP Logistic regression and Neural networks	Framework
A Comprehensive Survey on Federated Learning Techniques for Healthcare Informatics [109]	Explained various frameworks techniques and challenges in health informatics	How to strike a balance between privacy and performance Identified possibility of data leakage in FL as described by [63], [40]	None	Survey
Federated learning for healthcare domain-Pipeline, applications and challenges[101]	Identified architectural components of FL Discuss privacy and communication challenges [110] in healthcare	Federated evaluation (FedEva) [111] identifies “accuracy, communication, time consumption, privacy, data distribution, available resources and robustness” as evaluation targets in FL systems.	Adversarial attacks [112] Inference poisoning attacks [72] Evasion [113], privacy in Genomic data [114] noise [115]	Survey

			data biases [116] false positives and false negatives (Pollard et. al, 2019) patent variability, failure and dropouts [75] Computational overheads [116], [117] Inadequate model training	
Secure and Efficient Smart Healthcare System Based on Federated Learning[118]	Proposed dynamic secrete sharing mechanism for privacy protection Reduce time overhead through elliptic curve cryptosystem	Only considers time overhead for efficiency while ignoring other metrics such as data distribution, false positives and negatives, recall etc	Secrete sharing, two-mask protocol homogeneous linear recursive equation homomorphic hash function [98] elliptic curve crypto-system [119] full dynamic secret sharing [120]	Framework
Shuffled Check-in: Privacy Amplification towards Practical Distributed Learning[121]	proposed a protocol for privacy using numerical evaluation of Gaussian mechanism	Did not cover fairness with regards to biased data	SGD using Gaussian mechanism Rényi differential privacy (RDP)	Framework
On Privacy and Personalization in Cross-Silo Federated Learning [122]	Provided an empirical and theoretical study on mean-regularized multi-task learning (MR-MTL) as effective model personalization	Understands how privacy in device-silo differs from that in cross-silo FL	MR-MTL DP-SGD for noise reduction	Framework
HealthCare EHR: A Blockchain-Based Decentralized Application[123]	Use Ethereum blockchain to build a peer to peer network platform for distributed database of health entities This was to resolve a challenge of data lake that centrally puts the	Payment transaction through banks have not been institutionalized	Blockchain (Ethereum)	technique

	data in a central point masking heterogeneity			
Federated Learning and Differential Privacy: Software tools analysis, the Sherpa.ai FL framework and methodological guidelines for preserving data privacy [124] in [115]	Analyze software tools for FL and DP implementation Presents Sherpa.ai FL for developing AI using methodological guidelines	Proposed to add RAPPOR [125] as new DP mechanism together with concentrated DP [126] or Rényi DP [127]	FedAvg DP using SGD Sherpa.ai FL	technique
Privacy-Enhanced Federated Learning: A Restrictively Self-Sampled and Data-Perturbed Local Differential Privacy Method [128]	Proposed efficient data perturbation to improve communication Proposed restrictive client self-sampling technology	The enhanced Local Differential Privacy (Optimal LDP-FL) is theoretically analyzed on relationship between client-sampling probability and model [129] accuracy	LDP-FL	Framework
Benchmarking Differential Privacy and Federated Learning for BERT Models [130]	Used web twitter(tweets to detect depression tendency) dataset and sexual harassment cleansed data, trained on four NLP models(*BERT*, RoBERTa, DistillBERT and ALBERT) to implement DP and FL	Only considered accuracy metric Smaller dataset was used, health data is enormous	DP, FL and DP-FL on BERT, RoBERTa, DistillBERT and ALBERT	Framework
Federated Machine Learning: Concept and Applications [42]	Described categories, architectures, techniques used for privacy in FL Identified evaluation steps for vertical FL	The article describes the application areas without demonstrating any specific implementation or even evaluation against any data distribution	Secure Multiparty Computation (SMC) [91][131] Differential Privacy [2] Homomorphic Encryption[132] SGD [2]	Framework
A Hybrid Approach to Privacy-Preserving Federated Learning[91]	The approach uses secure multiparty computation and differential privacy to reduce increase of noise without affecting privacy (limiting extraction attacks and collision)	The choice of three ML (Decision Tree, CNN and Support Vector) algorithms against remaining algorithms such as regression, ANN, RNN, LSTM and others may not conclusively give significant accuracy loss	Decision tree (TD), CNN and Support Vector Machine (SVM) SMC for noise reduction [131] DP using Gaussian Distribution	Technique

`Blockchain Meets COVID-19: A Framework for Contact Information Sharing and Risk Notification System[133]	Used Bluetooth technology to solve Contact tracing of potential COVID19 victims	may affect performance [134] need to optimize, validate against other metrics apart from time	Blockchain and Bluetooth	Framework
Big healthcare data: preserving security and privacy[135]	The approach highlighted how privacy can be preserved in big data healthcare by analyzing unstructured patient data using natural-language	Reconstruction of data when performing privacy among patients data	Developed an invent monitoring system model using Spark	Techniques
Heterogeneous data and big data analytics[136]	Discussed privacy problems [137], scalability [138], lack of structure, storage bottlenecks [139], [140] spurious correlations, incidental endogeneity, noise accumulation, experimental variations, statistical bias	Requires noise and error reduction mechanisms in order to improve performance	*FedAvg* [1]	Framework

Table 3 Current models

Framework	**Strategy**	**Model development**	**Run-Time**	**Security**	**Implementation Status**	**Scheme**
TFF distribution under Apache 2.0 license [141]	Support *FedAvg*, FedSGD Functions that are used include: Sum Mean, *DPQueries*	Neural Networks (NN), Recurrent NN, Convolution NN. Keras and Tensor flow libraries Runs on Google Colaboratory	Native back-end is managed through Executor and Client interactions using *gRPC* technology and proto-object	DP	Lacks Federated mode, data splitting	Centralized
FATE developed by Webank's AI Department [144]	Heterogeneous:Secure Segregation (*SecAgg*), gradient-boosting decision tree (GBDT)	Dense layers of NN, Linear, logistic and Poisson, Decision Tree, K-Means for	Server (model aggregator) Scheduler and executer (implements FL algorithm)	HE, RSA, SPDZ	Both simulation mode and federated mode but lacks core API	Centralized

	Homogeneous: *FedAvg*, Secure [147] Aggregation	vertical data partitioning Runs on PyTorch	Also uses gRPC technology			
PFL with Apache 2.0 license [129]	Horizontal partition: *FedAvg*, *SecAgg* DPSGD *Vertical Partition*: MPC and PSI	RNN, CNN, NN, LR	Server, Worker (trainer) and Scheduler Interaction is via ZeroMQ	DP for horizontal partition and Secrete-Sharing for Vertical partition	Both simulation mode and federated mode. Docker containers	Decentralized (Peer-to-peer)
PySyft with MIT license [142]	Horizontal partition: *FedAvg*, *SecAgg* DPSGD Vertical Partition: MPC and PSI Its Open Minded	NN, RNN, CNN, LR and other deep learning algorithms	Virtual Worker, FL Client, SwiftSyft, kotlinSyftand syft.js (Anaconda Manager)	*PyVertical*, PSI, MPC, HE, DP	Implements both PyTorch and TF libraries	Decentralized
FL&DP developed by Sherpa from University of Granada [146]	*FedAvg*, Weighted *FedAvg*, IOWA, Cluster *FedAvg*	Tensorflow (NN, RNN, CNN) Scikit-learn (LC, LR and K-Means)	Aggregator, Database	AdaptiveDP, Randomized Response Coins	Simulation mode only	Centralized
Flower [143]	*FedAvg*, *FedProx*	NN, CNN, RNN	FL server and PRC server communicates [147],[148], [149-[153]with RPC client through gRPC technology [149]-[150]	DP	TFF, PyTorch, Keras [151]- [160] Client SDK (Java, Python, C++)	Decentralized

7. Technologies for big data security

In terms of security and privacy perspective [161], the authors in [162] argue that security in big data refers to three matters: data security, access control, and information security. In this regards, healthcare organizations must implement security measures and approaches to protect their big data, associated hardware and software, and both clinical and administrative information from internal and external risks. At a project's inception, the data lifecycle [163] must be established to ensure that appropriate decisions are made about retention, cost effectiveness, reuse and auditing of historical or new data. The authors in [164] have proposed privacy preserving data mining techniques in Hadoop. On the other hand, the authors in [165] suggested a big data security lifecycle model extended from the model in [166]. This model is designed to address the phases of the big data lifecycle and correlate threats and attacks [167], [168] that face big data environment within these phases, while [166] address big data lifecycle from user role perspective: data provider, data collector, data miner, and decision maker. The model proposed in [165] comprised of four interconnecting phases: data collection phase, data storage phase, data processing and analysis, and knowledge creation. The authors in [169] introduced also an efficient and privacy-preserving cosine similarity computing protocol. Furthermore, Chronic Conditions Data Warehouse (CCW) follows a formal information security lifecycle model, which consists of four core phases that serve to identify, assess, protect and monitor against patient data security threats. This lifecycle model is continually being improved with emphasis on constant attention and continual monitoring. Moreover, the authors in [170] have suggested a scalable approach to anonymize large-scale data sets.

Authentication is the act of establishing or confirming claims made by or about the subject are true and valid [171], [172], [173]. It serves vital functions within any organization: securing access to corporate networks, protecting the identities of users, and ensuring that the user is really who he is pretending to be. Researchers in [174] proposes a novel and simple authentication model using one time pad algorithm. It provides removing the communication of passwords between the servers. Security monitoring is gathering and investigating network events to catch the intrusions [175], [176]. Audit means recording user activities of the healthcare system in chronological order, such as maintaining a log of every access to and modification of data. These are two optional security metrics to measure and ensure the safety of a healthcare system [177], [178], [179]. The authors in [180] proposed various privacy issues dealing with big data applications, while researchers in [181] proposed an anonymization algorithm to speed up anonymization of big data streams. In addition, authors in [182] suggested a novel framework to achieve privacy-preserving machine learning and paper [183] proposed methodology provides data confidentiality and secure data sharing [184]. All these techniques and approaches have shown some limitations.

Big data network security systems should be find abnormalities quickly and identify correct alerts from heterogeneous data. Therefore, a big data security event monitoring system model has been proposed which consists of four modules: data collection, integration, analysis, and interpretation [185]. Although the current techniques offer patient's privacy, their demerits led to the advent of newer methods.

De-identification is a traditional method to prohibit the disclosure of confidential information by rejecting any information that can identify the patient, either by the first method that requires the removal of specific identifiers of the patient or by the second statistical method where the patient verifies himself that enough identifiers are deleted. Nonetheless, an attacker can possibly get more external information assistance for de-identification in big data. As a result, de-identification is not sufficient for protecting big data privacy. It could be more feasible through developing efficient privacy-preserving algorithms to help mitigate the risk of re-identification. The concepts of k-anonymity [186], l-diversity and t-closeness have been introduced to enhance this traditional technique.

8. Research gaps

Big data has fundamentally changed the way organizations manage, analyze and leverage data in any industry. One of the most promising fields where big data can be applied to make a change is healthcare. Big healthcare data has considerable potential to improve patient outcomes, predict outbreaks of epidemics, gain valuable insights, avoid preventable diseases, reduce the cost of healthcare delivery and improve the quality of life in general. However, deciding on the allowable uses of data while preserving security and patient's right to privacy is a difficult task [135]. Big data, no matter how useful for the advancement of medical science and vital to the success of all healthcare organizations, can only be used if security and privacy issues are addressed. To ensure a secure and trustworthy big data environment, it is essential to identify the limitations of existing solutions and envision directions for future research [187], [188].

Security and privacy [190] in big data are important issues. Privacy is often defined as having the ability to protect sensitive information about personally identifiable health care information. It focuses on the use and governance of individual's personal data like making policies and establishing authorization requirements to ensure that patients' personal information is being collected, shared and utilized in right ways. While security is typically defined as the protection against unauthorized access, with some including explicit mention of integrity and availability. It focuses on protecting data from pernicious attacks and stealing data for profit. Although security is vital for protecting data but it's insufficient for addressing privacy.

The heightened focus on care quality and value; and evidence-based medicine [191] as opposed to subjective clinical decisions—all of which are leading to offer significant opportunities for supporting clinical decision, improving healthcare delivery, management and policy making, surveilling disease, monitoring adverse events, and optimizing treatment for diseases affecting multiple organ systems. As noted above, big data analytics in healthcare carries many benefits, promises and presents great potential for transforming healthcare, yet it raises manifold barriers and challenges. Indeed, the concerns over the big healthcare data security and privacy are increased year-by-year. Additionally, healthcare organizations found that a reactive, bottom-up, technology-centric approach [192] to determining security and privacy requirements is not adequate to protect the organization and its patients.

The research in [193] protects against identity disclosure but failed to protect against attribute disclosure. The authors in [194] have presented p-sensitive anonymity that protects against both identity and attribute disclosure. Other anonymization methods fall into the classes of adding noise to the data, swapping cells within columns and replacing groups of k records with k copies of a single representative. These methods have a common problem of difficulty in anonymizing high dimensional data sets.

The researchers in [195] propose also a cloud-oriented storage efficient dynamic access control scheme cipher-text based on the CP-ABE and a symmetric encryption algorithm (such as AES). To satisfy requirements of fine-grained access control yet security and privacy preserving, suggestions have been given to adopt technologies in conjunction with other security techniques, such as encryption, and access control methods along communication channels [196]. On its part, k-anonymous data can still be helpless against attacks like unsorted matching attack, temporal attack, and complementary release attack. On the bright side, the complexity of rendering relations of private records k-anonymous, while minimizing the amount of information that is not released and simultaneously ensure the anonymity of individuals up to a group of size k, and withhold a minimum amount of information to achieve this privacy level and this optimization problem is NP-hard [197]. Various measures have been proposed to quantify information loss caused by anonymization, but they do not reflect the actual usefulness of data.

L-diversityis a form of group based anonymization that is utilized to safeguard privacy in data sets by diminishing the granularity of data representation [198]. This model (Distinct, Entropy, Recursive) is an extension of the k-anonymity which utilizes methods including generalization and suppression to reduce the granularity of data representation in a way that any given record maps onto at least k different records in the data. However, L-diversity method is also a subject to skewness and similarity attack and thus can't prevent attribute disclosure [199]. On the other hand, identity based anonymization is a type of information sanitization whose intent is privacy protection [200]. It is the process of either encrypting or removing personally identifiable information from data sets, so that the people whom the data describe remain anonymous. The main difficulty with this technique involves combining anonymization, privacy protection, and big data techniques [201] to analyze usage data while protecting the identities.

9. Conclusion

This paper discusses FL as a technique that can support big data analysis in healthcare in compliance to HIPPA act and other regulations. Federated averaging is a widely acceptable strategy used for training of models in multiple machines on private datasets without leakage of data among the individual machines. As shown in the illustration above, its use case in health is different concerning data implementation. In health environment, few clients (example health clinic), full autonomy of private health institutions' datasets and voluminous amount of data are some of unique characteristics. There are various advancements on FL that the paper has highlighted which aims to optimize its performances. For example *FedProx*, *q-FedAvg*, *per-FedAvg* among others. Due to persistent threat against private data such as leakage, de-anonymization etc, Differential Privacy (DP) approach has been used to secure and preserve private dataset residing in the client machine. This article reviews using mathematical notations, concept that is used to prevent leakage of information in practice. Further, the privacy of data is improved by addition of noise, by amplification and through composition, which applies principle of "Fundamental Law of Information Recovery". DP algorithm known as Stochastic Gradient Descent that is used to minimize privacy loss is also explained in this article. According to the literature domain

summarized in the table above, most challenges are on privacy and performance of the model. However, a trade-off between privacy and performance is needed in order to provide an trustable efficient systems. Therefore need to consider more performance metrics when evaluating a model in FL to ensure adequate verification and validation of the model. The study recommends development of Federated Learning techniques that can be applied to conduct analysis in clinical health while preserving the privacy of patients' in a collaborative environment.

Compliance with ethical standards

Acknowledgments

Special appreciation goes to my colleagues who assisted me in one way or another during the drafting of this paper.

References

[1] Rieke N, Hancox J, Li W, Milletari F, Roth HR, Albarqouni S, Bakas S, Galtier MN, Landman BA, Maier-Hein K, Ourselin S. The future of digital health with federated learning. NPJ digital medicine. 2020 Sep 14, 3(1):119.

[2] McMahan B, Moore E, Ramage D, Hampson S, y Arcas BA. Communication-efficient learning of deep networks from decentralized data. InArtificial intelligence and statistics 2017 Apr 10 (pp. 1273-1282). PMLR.

[3] Konečný J, McMahan HB, Ramage D, Richtárik P. Federated optimization: Distributed machine learning for on-device intelligence. arXiv preprint arXiv:1610.02527. 2016 Oct 8.

[4] Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. Future healthcare journal. 2021 Jul, 8(2):e188.

[5] Dwivedi YK, Hughes L, Ismagilova E, Aarts G, Coombs C, Crick T, Duan Y, Dwivedi R, Edwards J, Eirug A, Galanos V. Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management. 2021 Apr 1, 57:101994.

[6] Agarwal P, Swami S, Malhotra SK. Artificial intelligence adoption in the post COVID-19 new-normal and role of smart technologies in transforming business: a review. Journal of Science and Technology Policy Management. 2022 Feb 16.

[7] Nyangaresi VO, El-Omari NK, Nyakina JN. Efficient Feature Selection and ML Algorithm for Accurate Diagnostics. Journal of Computer Science Research. 2022 Jan 25, 4(1):10-9.

[8] Garg A, Mago V. Role of machine learning in medical research: A survey. Computer science review. 2021 May 1, 40:100370.

[9] MacEachern SJ, Forkert ND. Machine learning for precision medicine. Genome. 2021, 64(4):416-25.

[10] Chan HP, Samala RK, Hadjiiski LM, Zhou C. Deep learning in medical image analysis. Deep Learning in Medical Image Analysis: Challenges and Applications. 2020:3-21.

[11] Bhatt C, Kumar I, Vijayakumar V, Singh KU, Kumar A. The state of the art of deep learning models in medical science and their challenges. Multimedia Systems. 2021 Aug, 27(4):599-613.

[12] Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S. Dermatologist-level classification of skin cancer with deep neural networks. nature. 2017 Feb, 542(7639):115-8.

[13] Luo L, Xiong Y, Liu Y, Sun X. Adaptive gradient methods with dynamic bound of learning rate. arXiv preprint arXiv:1902.09843. 2019 Feb 26.

[14] Hegde N, Hipp JD, Liu Y, Emmert-Buck M, Reif E, Smilkov D, Terry M, Cai CJ, Amin MB, Mermel CH, Nelson PQ. Similar image search for histopathology: SMILY. NPJ digital medicine. 2019 Jun 21, 2(1):56.

[15] Komura D, Ishikawa S. Machine learning methods for histopathological image analysis. Computational and structural biotechnology journal. 2018 Jan 1, 16:34-42.

[16] De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, Askham H, Glorot X, O'Donoghue B, Visentin D, Van Den Driessche G. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nature medicine. 2018 Sep, 24(9):1342-50.

[17] Al Sibahee MA, Ma J, Nyangaresi VO, Abduljabbar ZA. Efficient Extreme Gradient Boosting Based Algorithm for QoS Optimization in Inter-Radio Access Technology Handoffs. In2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA) 2022 Jun 9 (pp. 1-6). IEEE.

[18] Cecilia JM, Cano JC, Morales-García J, Llanes A, Imbernón B. Evaluation of clustering algorithms on GPU-based edge computing platforms. Sensors. 2020 Nov 6, 20(21):6335.

[19] Sannara EK, Portet F, Lalanda P, German VE. A federated learning aggregation algorithm for pervasive computing: Evaluation and comparison. In2021 IEEE International Conference on Pervasive Computing and Communications (PerCom) 2021 Mar 22 (pp. 1-10). IEEE.

[20] Becker C, Julien C, Lalanda P, Zambonelli F. Pervasive computing middleware: current trends and emerging challenges. CCF Transactions on Pervasive Computing and Interaction. 2019 May 1, 1:10-23.

[21] Mnjama J, Foster G, Irwin B. A privacy and security threat assessment framework for consumer health wearables. In2017 Information Security for South Africa (ISSA) 2017 Aug 16 (pp. 66-73). IEEE.

[22] Farhadi M, Haddad H, Shahriar H. Static analysis of hippa security requirements in electronic health record applications. In2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC) 2018 Jul 23 (Vol. 2, pp. 474-479). IEEE.

[23] Voigt P, Von demBussche A. The eu general data protection regulation (gdpr). A Practical Guide, 1st Ed., Cham: Springer International Publishing. 2017 Aug, 10(3152676):10-5555.

[24] Tikkinen-Piri C, Rohunen A, Markkula J. EU General Data Protection Regulation: Changes and implications for personal data collecting companies. Computer Law & Security Review. 2018 Feb 1, 34(1):134-53.

[25] Nyangaresi VO. Privacy Preserving Three-factor Authentication Protocol for Secure Message Forwarding in Wireless Body Area Networks. Ad Hoc Networks. 2023 Apr 1, 142:103117.

[26] Purushotham S, Meng C, Che Z, Liu Y. Benchmarking deep learning models on large healthcare datasets. Journal of biomedical informatics. 2018 Jul 1, 83:112-34.

[27] Johnson AE, Pollard TJ, Shen L, Lehman LW, Feng M, Ghassemi M, Moody B, Szolovits P, Anthony Celi L, Mark RG. MIMIC-III, a freely accessible critical care database. Scientific data. 2016 May 24, 3(1):1-9.

[28] Denaxas S, Gonzalez-Izquierdo A, Direk K, Fitzpatrick NK, Fatemifar G, Banerjee A, Dobson RJ, Howe LJ, Kuan V, Lumbers RT, Pasea L. UK phenomics platform for developing and validating electronic health record phenotypes: CALIBER. Journal of the American Medical Informatics Association. 2019 Dec, 26(12):1545-59.

[29] Gao Y, Cai C, Grifoni A, Müller TR, Niessl J, Olofsson A, Humbert M, Hansson L, Österborg A, Bergman P, Chen P. Ancestral SARS-CoV-2-specific T cells cross-recognize the Omicron variant. Nature medicine. 2022 Mar, 28(3):472-6.

[30] Hendolin M. Towards the European health data space: from diversity to a common framework. Eurohealth. 2022, 27(2):15-7.

[31] Horgan D, Hajduch M, Vrana M, Soderberg J, Hughes N, Omar MI, Lal JA, Kozaric M, Cascini F, Thaler V, Solà-Morales O. European Health Data Space—An Opportunity Now to Grasp the Future of Data-Driven Healthcare. InHealthcare 2022 Aug 26 (Vol. 10, No. 9, p. 1629). MDPI.

[32] Kirby J, Prior F, Petrick N, Hadjiski L, Farahani K, Drukker K, Kalpathy-Cramer J, Glide-Hurst C, El Naqa I. Introduction to special issue on datasets hosted in The Cancer Imaging Archive (TCIA).

[33] Tsai EB, Simpson S, Lungren MP, Hershman M, Roshkovan L, Colak E, Erickson BJ, Shih G, Stein A, Kalpathy-Cramer J, Shen J. The RSNA international COVID-19 open radiology database (RICORD). Radiology. 2021 Apr, 299(1):E204-13.

[34] Shiri I, Arabi H, Salimi Y, Sanaat A, Akhavanallaf A, Hajianfar G, Askari D, Moradi S, Mansouri Z, Pakbin M, Sandoughdaran S. COLI-Net: Deep learning-assisted fully automated COVID-19 lung and infection pneumonia lesion detection and segmentation from chest computed tomography images. International journal of imaging systems and technology. 2022 Jan, 32(1):12-25.

[35] Hutter C, Zenklusen JC. The cancer genome atlas: creating lasting value beyond its data. Cell. 2018 Apr 5, 173(2):283-5.

[36] Yang J, Shi R, Ni B. Medmnist classification decathlon: A lightweight automl benchmark for medical image analysis. In2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI) 2021 Apr 13 (pp. 191-195). IEEE.

[37] Nyangaresi VO. Provably Secure Pseudonyms based Authentication Protocol for Wearable Ubiquitous Computing Environment. In2022 International Conference on Inventive Computation Technologies (ICICT) 2022 Jul 20 (pp. 1-6). IEEE.

[38] Hao M, Li H, Luo X, Xu G, Yang H, Liu S. Efficient and privacy-enhanced federated learning for industrial artificial intelligence. IEEE Transactions on Industrial Informatics. 2019 Oct 4, 16(10):6532-42.

[39] Li L, Fan Y, Tse M, Lin KY. A review of applications in federated learning. Computers & Industrial Engineering. 2020 Nov 1, 149:106854.

[40] Li J, Meng Y, Ma L, Du S, Zhu H, Pei Q, Shen X. A federated learning based privacy-preserving smart healthcare system. IEEE Transactions on Industrial Informatics. 2021 Jul 20, 18(3).

[41] Shokri R, Stronati M, Song C, Shmatikov V. Membership inference attacks against machine learning models. In2017 IEEE symposium on security and privacy (SP) 2017 May 22 (pp. 3-18). IEEE.

[42] Yang Q, Liu Y, Chen T, Tong Y. Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology (TIST). 2019 Jan 28, 10(2):1-9.

[43] Kumar Y, Singla R. Federated learning systems for healthcare: perspective and recent progress. Federated Learning Systems: Towards Next-Generation AI. 2021:141-56.

[44] Sattler F, Wiedemann S, Müller KR, Samek W. Robust and communication-efficient federated learning from non-iid data. IEEE transactions on neural networks and learning systems. 2019 Nov 1, 31(9):3400-13.

[45] Shi Y, Shen L, Wei K, Sun Y, Yuan B, Wang X, Tao D. Improving the model consistency of decentralized federated learning. arXiv preprint arXiv:2302.04083. 2023 Feb 8.

[46] Zhang C, Xie Y, Bai H, Yu B, Li W, Gao Y. A survey on federated learning. Knowledge-Based Systems. 2021 Mar 15, 216:106775.

[47] Antunes RS, André da Costa C, Küderle A, Yari IA, Eskofier B. Federated learning for healthcare: Systematic review and architecture proposal. ACM Transactions on Intelligent Systems and Technology (TIST). 2022 May 4, 13(4):1-23.

[48] AbdulRahman S, Tout H, Ould-Slimane H, Mourad A, Talhi C, Guizani M. A survey on federated learning: The journey from centralized to distributed on-site learning and beyond. IEEE Internet of Things Journal. 2020 Oct 12, 8(7):5476-97

[49] Nyangaresi VO, Abd-Elnaby M, Eid MM, Nabih Zaki Rashed A. Trusted authority based session key agreement and authentication algorithm for smart grid networks. Transactions on Emerging Telecommunications Technologies. 2022 Sep, 33(9):e4528.

[50] Li T, Sahu AK, Zaheer M, Sanjabi M, Talwalkar A, Smith V. Federated optimization in heterogeneous networks. Proceedings of Machine learning and systems. 2020 Mar 15, 2:429-50.

[51] Reddi S, Charles Z, Zaheer M, Garrett Z, Rush K, Konečný J, Kumar S, McMahan HB. Adaptive federated optimization. arXiv preprint arXiv:2003.00295. 2020 Feb 29.

[52] Niu X, Wei E. FedHybrid: A hybrid federated optimization method for heterogeneous clients. IEEE Transactions on Signal Processing. 2023 Jan 26, 71:150-63.

[53] Nyangaresi VO. Target Tracking Area Selection and Handover Security in Cellular Networks: A Machine Learning Approach. InProceedings of Third International Conference on Sustainable Expert Systems: ICSES 2022 2023 Feb 23 (pp. 797-816). Singapore: Springer Nature Singapore.

[54] Su L, Xu J, Yang P. A Non-parametric View of FedAvg and FedProx: Beyond Stationary Points. arXiv preprint arXiv:2106.15216. 2021 Jun 29.

[55] Khodak M, Tu R, Li T, Li L, Balcan MF, Smith V, Talwalkar A. Federated hyperparameter tuning: Challenges, baselines, and connections to weight-sharing. Advances in Neural Information Processing Systems. 2021 Dec 6, 34:19184-97.

[56] Shen C, Wang P, Roth HR, Yang D, Xu D, Oda M, Wang W, Fuh CS, Chen PT, Liu KL, Liao WC. Multi-task federated learning for heterogeneous pancreas segmentation. InClinical Image-Based Procedures, Distributed and Collaborative Learning, Artificial Intelligence for Combating COVID-19 and Secure and Privacy-Preserving Machine Learning: 10th Workshop, CLIP 2021, Second Workshop, DCL 2021, First Workshop, LL-COVID19 2021, and First Workshop and Tutorial, PPML 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27 and October 1, 2021, Proceedings 2 2021 (pp. 101-110). Springer International Publishing.

[57] Li T, Sanjabi M, Beirami A, Smith V. Fair resource allocation in federated learning. arXiv preprint arXiv:1905.10497. 2019 May 25.

[58] Mozaffari H, Houmansadr A. E2FL: Equal and equitable federated learning. arXiv preprint arXiv:2205.10454. 2022 May 20.

[59] Tan AZ, Yu H, Cui L, Yang Q. Towards personalized federated learning. IEEE Transactions on Neural Networks and Learning Systems. 2022 Mar 28.

[60] Fallah A, Mokhtari A, Ozdaglar A. Personalized federated learning: A meta-learning approach. arXiv preprint arXiv:2002.07948. 2020 Feb 19.

[61] Kulkarni V, Kulkarni M, Pant A. Survey of personalization techniques for federated learning. In2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4) 2020 Jul 27 (pp. 794-797). IEEE.

[62] Hussain MA, Hussien ZA, Abduljabbar ZA, Ma J, Al Sibahee MA, Hussain SA, Nyangaresi VO, Jiao X. Provably throttling SQLI using an enciphering query and secure matching. Egyptian Informatics Journal. 2022 Dec 1, 23(4):145-62.

[63] Hu R, Guo Y, Li H, Pei Q, Gong Y. Personalized federated learning with differential privacy. IEEE Internet of Things Journal. 2020 Apr 30, 7(10):9530-9.

[64] Tang X, Guo S, Guo J. Personalized federated learning with clustered generalization.

[65] Matsuda K, Sasaki Y, Xiao C, Onizuka M. An Empirical Study of Personalized Federated Learning. arXiv preprint arXiv:2206.13190. 2022 Jun 27.

[66] Smith V, Chiang CK, Sanjabi M, Talwalkar AS. Federated multi-task learning. Advances in neural information processing systems. 2017, 30.

[67] Marfoq O, Neglia G, Bellet A, Kameni L, Vidal R. Federated multi-task learning under a mixture of distributions. Advances in Neural Information Processing Systems. 2021 Dec 6, 34:15434-47.

[68] Dinh CT, Vu TT, Tran NH, Dao MN, Zhang H. Fedu: A unified framework for federated multi-task learning with laplacian regularization. arXiv preprint arXiv:2102.07148. 2021, 400.

[69] Ruder S. An overview of multi-task learning in deep neural networks. arXiv preprint arXiv:1706.05098. 2017 Jun 15.

[70] Amyar A, Modzelewski R, Li H, Ruan S. Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: Classification and segmentation. Computers in biology and medicine. 2020 Nov 1, 126:104037.

[71] Nyangaresi VO, Ogundoyin SO. Certificate based authentication scheme for smart homes. In2021 3rd Global Power, Energy and Communication Conference (GPECOM) 2021 Oct 5 (pp. 202-207). IEEE.

[72] Tian Y, Wan Y, Lyu L, Yao D, Jin H, Sun L. FedBERT: when federated learning meets pre-training. ACM Transactions on Intelligent Systems and Technology (TIST). 2022 Aug 24, 13(4):1-26.

[73] Wei K, Li J, Ding M, Ma C, Yang HH, Farokhi F, Jin S, Quek TQ, Poor HV. Federated learning with differential privacy: Algorithms and performance analysis. IEEE Transactions on Information Forensics and Security. 2020 Apr 17, 15:3454-69.

[74] El Ouadrhiri A, Abdelhadi A. Differential privacy for deep and federated learning: A survey. IEEE Access. 2022 Feb 15, 10:22359-80.

[75] Zhao Y, Zhao J, Yang M, Wang T, Wang N, Lyu L, Niyato D, Lam KY. Local differential privacy-based federated learning for internet of things. IEEE Internet of Things Journal. 2020 Nov 10, 8(11):8836-53.

[76] Kim JW, Edemacu K, Kim JS, Chung YD, Jang B. A survey of differential privacy-based techniques and their applicability to location-based services. Computers & Security. 2021 Dec 1, 111:102464.

[77] Yeom S, Giacomelli I, Fredrikson M, Jha S. Privacy risk in machine learning: Analyzing the connection to overfitting. In2018 IEEE 31st computer security foundations symposium (CSF) 2018 Jul 9 (pp. 268-282). IEEE.

[78] Carlini N, Tramer F, Wallace E, Jagielski M, Herbert-Voss A, Lee K, Roberts A, Brown TB, Song D, Erlingsson U, Oprea A. Extracting Training Data from Large Language Models. InUSENIX Security Symposium 2021 Aug 11 (Vol. 6).

[79] Ravindra V, Grama A. De-anonymization attacks on neuroimaging datasets. InProceedings of the 2021 International Conference on Management of Data 2021 Jun 9 (pp. 2394-2398).

[80] Hirschprung RS, Leshman O. Privacy disclosure by de-anonymization using music preferences and selections. Telematics and Informatics. 2021 Jun 1, 59:101564.

[81] Nyangaresi VO. Lightweight anonymous authentication protocol for resource-constrained smart home devices based on elliptic curve cryptography. Journal of Systems Architecture. 2022 Dec 1, 133:102763.

[82] Narayanan A, Shmatikov V. Robust de-anonymization of large sparse datasets: a decade later. May. 2019 May 21, 21:2019.

[83] Nagar A. Privacy-preserving blockchain based federated learning with differential data sharing. arXiv preprint arXiv:1912.04859. 2019 Dec 10.

[84] Xu J, Glicksberg BS, Su C, Walker P, Bian J, Wang F. Federated learning for healthcare informatics. Journal of Healthcare Informatics Research. 2021 Mar, 5:1-9.

[85] Rückel T, Sedlmeir J, Hofmann P. Fairness, integrity, and privacy in a scalable blockchain-based federated learning system. Computer Networks. 2022 Jan 15, 202:108621.

[86] Lian Z, Yang Q, Zeng Q, Su C. Webfed: Cross-platform federated learning framework based on web browser with local differential privacy. InICC 2022-IEEE International Conference on Communications 2022 May 16 (pp. 2071-2076). IEEE.

[87] Ziller A, Trask A, Lopardo A, Szymkow B, Wagner B, Bluemke E, Nounahon JM, Passerat-Palmbach J, Prakash K, Rose N, Ryffel T. Pysyft: A library for easy federated learning. Federated Learning Systems: Towards Next-Generation AI. 2021:111-39.

[88] Nyangaresi VO, Ahmad M, Alkhayyat A, Feng W. Artificial neural network and symmetric key cryptography based verification protocol for 5G enabled Internet of Things. Expert Systems. 2022 Dec, 39(10):e13126.

[89] Darzidehkalani E, Ghasemi-Rad M, van Ooijen PM. Federated learning in medical imaging: Part II: methods, challenges, and considerations. Journal of the American College of Radiology. 2022 Aug 1, 19(8):975-82.

[90] Lu W, Wang J, Chen Y, Qin X, Xu R, Dimitriadis D, Qin T. Personalized federated learning with adaptive batchnorm for healthcare. IEEE Transactions on Big Data. 2022 May 23.

[91] Truex S, Baracaldo N, Anwar A, Steinke T, Ludwig H, Zhang R, Zhou Y. A hybrid approach to privacy-preserving federated learning. InProceedings of the 12th ACM workshop on artificial intelligence and security 2019 Nov 11 (pp. 1-11).

[92] Balle B, Wang YX. Improving the gaussian mechanism for differential privacy: Analytical calibration and optimal denoising. InInternational Conference on Machine Learning 2018 Jul 3 (pp. 394-403). PMLR.

[93] Sengupta P, Paul S, Mishra S. Learning with differential privacy. InHandbook of Research on Cyber Crime and Information Privacy 2021 (pp. 372-395). IGI Global.

[94] Nyangaresi VO. Masked Symmetric Key Encrypted Verification Codes for Secure Authentication in Smart Grid Networks. In 2022 4th Global Power, Energy and Communication Conference (GPECOM) 2022 Jun 14 (pp. 427-432).

[95] Abadi M, Chu A, Goodfellow I, McMahan HB, Mironov I, Talwar K, Zhang L. Deep learning with differential privacy. InProceedings of the 2016 ACM SIGSAC conference on computer and communications security 2016 Oct 24 (pp. 308-318).

[96] Riaz S, Ali S, Wang G, Anees A. Differentially private block coordinate descent. Journal of King Saud University-Computer and Information Sciences. 2023 Jan 1, 35(1):283-95.

[97] Ain QT, Ali M, Riaz A, Noureen A, Kamran M, Hayat B, Rehman A. Sentiment analysis using deep learning techniques: a review. International Journal of Advanced Computer Science and Applications. 2017, 8(6).

[98] Bonawitz K, Ivanov V, Kreuter B, Marcedone A, McMahan HB, Patel S, Ramage D, Segal A, Seth K. Practical secure aggregation for privacy-preserving machine learning. Inproceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security 2017 Oct 30 (pp. 1175-1191).

[99] Alsamhi SH, Shvetsov AV, Kumar S, Shvetsova SV, Alhartomi MA, Hawbani A, Rajput NS, Srivastava S, Saif A, Nyangaresi VO. UAV computing-assisted search and rescue mission framework for disaster and harsh environment mitigation. Drones. 2022 Jun 22, 6(7):154.

[100] Al-Rubaie M, Chang JM. Reconstruction attacks against mobile-based continuous authentication systems in the cloud. IEEE Transactions on Information Forensics and Security. 2016 Jul 27, 11(12):2648-63.

[101] Shokri R, Stronati M, Song C, Shmatikov V. Membership inference attacks against machine learning models. In2017 IEEE symposium on security and privacy (SP) 2017 May 22 (pp. 3-18). IEEE.

[102] Bogdanov A, Rosen A. Pseudorandom functions: Three decades later. InTutorials on the Foundations of Cryptography: Dedicated to OdedGoldreich 2017 Apr 6 (pp. 79-158). Cham: Springer International Publishing.

[103] Zhang X, Chen X, Hong M, Wu ZS, Yi J. Understanding clipping for federated learning: Convergence and client-level differential privacy. InInternational Conference on Machine Learning, ICML 2022 2022 Jan.

[104] Li W, Milletarì F, Xu D, Rieke N, Hancox J, Zhu W, Baust M, Cheng Y, Ourselin S, Cardoso MJ, Feng A. Privacy-preserving federated brain tumour segmentation. InMachine Learning in Medical Imaging: 10th International Workshop, MLMI 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Proceedings 10 2019 (pp. 133-141). Springer International Publishing.

[105] Nyangaresi VO, Ma J. A Formally Verified Message Validation Protocol for Intelligent IoT E-Health Systems. In2022 IEEE World Conference on Applied Intelligence and Computing (AIC) 2022 Jun 17 (pp. 416-422). IEEE.

[106] Ghaffari M, Sowmya A, Oliver R. Automated brain tumor segmentation using multimodal brain scans: a survey based on models submitted to the BraTS 2012–2018 challenges. IEEE reviews in biomedical engineering. 2019 Oct 11, 13:156-68.

[107] Li Z, Tian Y, Zhang W, Liao Q, Liu Y, Du X, Guizani M. RR-LADP: A privacy-enhanced federated learning scheme for internet of everything. IEEE Consumer Electronics Magazine. 2021 Feb 17, 10(5):93-101.

[108] Cohen G, Afshar S, Tapson J, Van Schaik A. EMNIST: Extending MNIST to handwritten letters. In2017 international joint conference on neural networks (IJCNN) 2017 May 14 (pp. 2921-2926). IEEE.

[109] Dasaradharami Reddy K, Gadekallu TR. A Comprehensive Survey on Federated Learning Techniques for Healthcare Informatics. Computational Intelligence and Neuroscience. 2023 Mar 1, 2023.

[110] Mohammad Z, Nyangaresi V, Abusukhon A. On the Security of the Standardized MQV Protocol and Its Based Evolution Protocols. In2021 International Conference on Information Technology (ICIT) 2021 Jul 14 (pp. 320-325). IEEE.

[111] Chai D, Wang L, Chen K, Yang Q. Fedeval: A benchmark system with a comprehensive evaluation model for federated learning. arXiv preprint arXiv:2011.09655. 2020 Nov 19.

[112] Papernot N, McDaniel P, Jha S, Fredrikson M, Celik ZB, Swami A. The limitations of deep learning in adversarial settings. In2016 IEEE European symposium on security and privacy (EuroS&P) 2016 Mar 21 (pp. 372-387). IEEE.

[113] Geiping J, Bauermeister H, Dröge H, Moeller M. Inverting gradients-how easy is it to break privacy in federated learning?. Advances in Neural Information Processing Systems. 2020, 33:16937-47.

[114] Akgün M, Bayrak AO, Ozer B, Sağıroğlu MŞ. Privacy preserving processing of genomic data: A survey. Journal of biomedical informatics. 2015 Aug 1, 56:103-11.

[115] Rodríguez-Barroso N, Stipcich G, Jiménez-López D, Ruiz-Millán JA, Martínez-Cámara E, González-Seco G, Luzón MV, Veganzones MA, Herrera F. Federated Learning and Differential Privacy: Software tools analysis, the Sherpa. ai FL framework and methodological guidelines for preserving data privacy. Information Fusion. 2020 Dec 1, 64:270-92.

[116] Kairouz P, McMahan HB, Avent B, Bellet A, Bennis M, Bhagoji AN, Bonawitz K, Charles Z, Cormode G, Cummings R, D'Oliveira RG. Advances and open problems in federated learning. Foundations and Trends® in Machine Learning. 2021 Jun 22, 14(1–2):1-210.

[117] Nyangaresi VO. A Formally Validated Authentication Algorithm for Secure Message Forwarding in Smart Home Networks. SN Computer Science. 2022 Jul 9, 3(5):364.

[118] Liu W, Zhang Y, Han G, Cao J, Cui H, Zheng D. Secure and Efficient Smart Healthcare System Based on Federated Learning. International Journal of Intelligent Systems. 2023 Feb 27, 2023.

[119] Xiao Y, Lin W, Zhao Y, Cui C, Cai Z. A high-speed elliptic curve cryptography processor for teleoperated systems security. Mathematical Problems in Engineering. 2021 Jan 22, 2021:1-8.

[120] Liu YX, Yang CN, Wu CM, Sun QD, Bi W. Threshold changeable secret image sharing scheme based on interpolation polynomial. Multimedia Tools and Applications. 2019 Jul 15, 78:18653-67.

[121] Liew SP, Hasegawa S, Takahashi T. Shuffled check-in: Privacy amplification towards practical distributed learning. arXiv preprint arXiv:2206.03151. 2022 Jun 7.

[122] Liu K, Hu S, Wu SZ, Smith V. On privacy and personalization in cross-silo federated learning. Advances in Neural Information Processing Systems. 2022 Dec 6, 35:5925-40.

[123] Panigrahi A, Nayak AK, Paul R. HealthCare EHR: A Blockchain-Based Decentralized Application. International Journal of Information Systems and Supply Chain Management (IJISSCM). 2022 Jul 1, 15(3):1-5.

[124] Nyangaresi VO. ECC based authentication scheme for smart homes. In2021 International Symposium ELMAR 2021 Sep 13 (pp. 5-10). IEEE.

[125] Erlingsson Ú, Pihur V, Korolova A. Rappor: Randomized aggregatable privacy-preserving ordinal response. InProceedings of the 2014 ACM SIGSAC conference on computer and communications security 2014 Nov 3 (pp. 1054-1067).

[126] Dwork C, Smith A, Steinke T, Ullman J. Exposed! a survey of attacks on private data. Annual Review of Statistics and Its Application. 2017 Mar 7, 4:61-84

[127] Danger R. Differential Privacy: What is all the noise about?.arXiv preprint arXiv:2205.09453. 2022 May 19.

[128] Zhao J, Yang M, Zhang R, Song W, Zheng J, Feng J, Matwin S. Privacy-Enhanced Federated Learning: A Restrictively Self-Sampled and Data-Perturbed Local Differential Privacy Method. Electronics. 2022 Dec 2, 11(23):4007.

[129] Nyakomitta SP, Omollo V. Biometric-Based Authentication Model for E-Card Payment Technology. IOSR Journal of Computer Engineering (IOSRJCE). 2014, 16(5):137-44.

[130] Basu P, Roy TS, Naidu R, Muftuoglu Z, Singh S, Mireshghallah F. Benchmarking differential privacy and federated learning for bert models. arXiv preprint arXiv:2106.13973. 2021 Jun 26.

[131] Yao Y, Wei J, Liu J, Zhang R. Efficiently secure multiparty computation based on homomorphic encryption. In2016 4th International Conference on Cloud Computing and Intelligence Systems (CCIS) 2016 Aug 17 (pp. 343-349). IEEE.

[132] Acar A, Aksu H, Uluagac AS, Conti M. A survey on homomorphic encryption schemes: Theory and implementation. ACM Computing Surveys (Csur). 2018 Jul 25, 51(4):1-35.

[133] Song J, Gu T, Fang Z, Feng X, Ge Y, Fu H, Hu P, Mohapatra P. Blockchain meets COVID-19: A framework for contact information sharing and risk notification system. In2021 IEEE 18th international conference on mobile ad hoc and smart systems (MASS) 2021 Oct 4 (pp. 269-277). IEEE.

[134] Nyangaresi VO. Terminal independent security token derivation scheme for ultra-dense IoT networks. Array. 2022 Sep 1, 15:100210.

[135] Abouelmehdi K, Beni-Hessane A, Khaloufi H. Big healthcare data: preserving security and privacy. Journal of big data. 2018 Dec, 5(1):1-8.

[136] Wang L. Heterogeneous data and big data analytics. Automatic Control and Information Sciences. 2017 Oct, 3(1):8-15.vacy. Journal of big data. 2018 Dec, 5(1):1-8.

[137] Malin BA, Emam KE, O'Keefe CM. Biomedical data privacy: problems, perspectives, and recent advances. Journal of the American medical informatics association. 2013 Jan 1, 20(1):2-6.

[138] Karakus M, Durresi A. A survey: Control plane scalability issues and approaches in software-defined networking (SDN). Computer Networks. 2017 Jan 15, 112:279-93.

[139] Papageorgiou L, Eleni P, Raftopoulou S, Mantaiou M, Megalooikonomou V, Vlachakis D. Genomic big data hitting the storage bottleneck. EMBnet. journal. 2018, 24.

[140] Nyangaresi VO. Provably secure protocol for 5G HetNets. In2021 IEEE International Conference on Microwaves, Antennas, Communications and Electronic Systems (COMCAS) 2021 Nov 1 (pp. 17-22). IEEE.

[141] Das SK, Bebortta S. Heralding the future of federated learning framework: architecture, tools and future directions. In2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) 2021 Jan 28 (pp. 698-703). IEEE.

[142] Budrionis A, Miara M, Miara P, Wilk S, Bellika JG. Benchmarking pysyft federated learning framework on mimic-iii dataset. IEEE Access. 2021 Aug 18, 9:116869-78.

[143] Beutel DJ, Topal T, Mathur A, Qiu X, Fernandez-Marques J, Gao Y, Sani L, Li KH, Parcollet T, de Gusmão PP, Lane ND. Flower: A friendly federated learning research framework. arXiv preprint arXiv:2007.14390. 2020 Jul 28.

[144] Liu Y, Fan T, Chen T, Xu Q, Yang Q. Fate: An industrial grade platform for collaborative learning with data protection. The Journal of Machine Learning Research. 2021 Jan 1, 22(1):10320-5.

[145] Su W, Li L, Liu F, He M, Liang X. AI on the edge: a comprehensive review. Artificial Intelligence Review. 2022 Mar 21:1-59.

[146] Qammar A, Ding J, Ning H. Federated learning attack surface: taxonomy, cyber defences, challenges, and future directions. Artificial Intelligence Review. 2022 Jun 1:1-38.

[147] Lombardi A. WebSocket: lightweight client-server communications. " O'Reilly Media, Inc.", 2015 Sep 9.

[148] Nyangaresi VO. Hardware assisted protocol for attacks prevention in ad hoc networks. InEmerging Technologies in Computing: 4th EAI/IAER International Conference, iCETiC 2021, Virtual Event, August 18–19, 2021, Proceedings 4 2021 (pp. 3-20). Springer International Publishing.

[149] Sun J, Khan F, Li J, Alshehri MD, Alturki R, Wedyan M. Mutual authentication scheme for the device-to-server communication in the Internet of medical things. IEEE Internet of Things Journal. 2021 May 10, 8(21):15663-71.

[150] Tang J, Tang X, Yuan J. Optimizing inter-server communication for online social networks. In2015 IEEE 35th International Conference on Distributed Computing Systems 2015 Jun 29 (pp. 215-224). IEEE.

[151] Chen H, Jin H, Wu S. Minimizing inter-server communications by exploiting self-similarity in online social networks. IEEE Transactions on Parallel and Distributed Systems. 2015 Apr 28, 27(4):1116-30.

[152] Basin D, Radomirovic S, Schläepfer M. A complete characterization of secure human-server communication. In2015 IEEE 28th Computer Security Foundations Symposium 2015 Jul 13 (pp. 199-213). IEEE.

[153] Nyangaresi VO. Lightweight key agreement and authentication protocol for smart homes. In2021 IEEE AFRICON 2021 Sep 13 (pp. 1-6). IEEE.

[154] Araújo M, Maia M, Rego P, de Souza J. Performance analysis of computational offloading on embedded platforms using the gRPC framework. In8th International Workshop on ADVANCEs in ICT Infrastructures and Services (ADVANCE 2020) 2020 Jan 27 (pp. 1-8).

[155] Indrasiri K, Kuruppu D. gRPC: up and running: building cloud native applications with Go and Java for Docker and Kubernetes. O'Reilly Media, 2020 Jan 23.

[156] Bolanowski M, Żak K, Paszkiewicz A, Ganzha M, Paprzycki M, Sowiński P, Lacalle I, Palau CE. Eficiency of REST and gRPC realizing communication tasks in microservice-based ecosystems. arXiv preprint arXiv:2208.00682. 2022 Aug 1.

[157] Bansal S, Singh M, Bhadauria M, Adalakha R. Federated Learning Approach towards Sentiment Analysis. In2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS) 2022 Oct 10 (pp. 717-724). IEEE.

[158] Kholod I, Yanaki E, Fomichev D, Shalugin E, Novikova E, Filippov E, Nordlund M. Open-source federated learning frameworks for IoT: A comparative review and analysis. Sensors. 2020 Dec 29, 21(1):167.

[159] van Rooij SB, Bouma H, van Mil J, ten Hove JM. Federated tool for anonymization and annotation in image data. InCounterterrorism, Crime Fighting, Forensics, and Surveillance Technologies VI 2022 Oct 28 (Vol. 12275, pp. 90-99). SPIE.

[160] Beutel DJ, Topal T, Mathur A, Qiu X, Fernandez-Marques J, Gao Y, Sani L, Li KH, Parcollet T, de Gusmão PP, Lane ND. Flower: A friendly federated learning research framework. arXiv preprint arXiv:2007.14390. 2020 Jul 28.

[161] Nyangaresi VO, Petrovic N. Efficient PUF based authentication protocol for internet of drones. In2021 International Telecommunications Conference (ITC-Egypt) 2021 Jul 13 (pp. 1-4). IEEE.

[162] Kim SH, Kim NU, Chung TM. Attribute relationship evaluation methodology for big data security. In2013 International conference on IT convergence and security (ICITCS) 2013 Dec 16 (pp. 1-4). IEEE.

[163] El Arass M, Souissi N. Data lifecycle: From big data to smartdata. In2018 IEEE 5th international congress on information science and technology (CiSt) 2018 Oct 21 (pp. 80-87). IEEE.

[164] Jung K, Park S, Park S. Hiding a needle in a haystack: privacy preserving apriori algorithm inmapreduce framework. InProceedings of the First International Workshop on Privacy and Secuirty of Big Data 2014 Nov 7 (pp. 11-17).

[165] Abouelmehdi K, Beni-Hessane A, Khaloufi H. Big healthcare data: preserving security and privacy. Journal of big data. 2018 Dec, 5(1):1-8.

[166] Xu L, Jiang C, Wang J, Yuan J, Ren Y. Information security in big data: privacy and data mining. Ieee Access. 2014 Oct 9, 2:1149-76.

[167] Nyangaresi VO, Mohammad Z. Privacy preservation protocol for smart grid networks. In2021 International Telecommunications Conference (ITC-Egypt) 2021 Jul 13 (pp. 1-4). IEEE.

[168] Eke CI, Norman AA, Mulenga M. Machine learning approach for detecting and combating bring your own device (BYOD) security threats and attacks: a systematic mapping review. Artificial Intelligence Review. 2023 Jan 17:1-44.

[169] Lu R, Zhu H, Liu X, Liu JK, Shao J. Toward efficient and privacy-preserving computing in big data era. IEEE Network. 2014 Jul 24, 28(4):46-50.

[170] Zhang X, Yang LT, Liu C, Chen J. A scalable two-phase top-down specialization approach for data anonymization using mapreduce on cloud. IEEE Transactions on Parallel and Distributed Systems. 2013 Feb 25, 25(2):363-73.

[171] Wazid M, Das AK, Hussain R, Succi G, Rodrigues JJ. Authentication in cloud-driven IoT-based big data environment: Survey and outlook. Journal of systems architecture. 2019 Aug 1, 97:185-96.

[172] Nyangaresi VO, Moundounga AR. Secure data exchange scheme for smart grids. In2021 IEEE 6th International Forum on Research and Technology for Society and Industry (RTSI) 2021 Sep 6 (pp. 312-316). IEEE.

[173] Fadi AT, Deebak BD. Seamless authentication: for IoT-big data technologies in smart industrial application systems. IEEE Transactions on Industrial Informatics. 2020 Apr 27, 17(4):2919-27.

[174] Yang C, Lin W, Liu M. A novel triple encryption scheme for hadoop-based cloud data security. In2013 Fourth International Conference on Emerging Intelligent Data and Web Technologies 2013 Sep 9 (pp. 437-442). IEEE.

[175] Vahidi S, Ghafouri M, Au M, Kassouf M, Mohammadi A, Debbabi M. Security of Wide-Area Monitoring, Protection, and Control (WAMPAC) Systems of the Smart Grid: A Survey on Challenges and Opportunities. IEEE Communications Surveys & Tutorials. 2023 Mar 8.

[176] Chen CM, Liu S, Li X, Islam SH, Das AK. A provably-secure authenticated key agreement protocol for remote patient monitoring IoMT. Journal of Systems Architecture. 2023 Mar 1, 136:102831.

[177] Valencia CV, Dove MS, Cummins SE, Kirby C, Zhu SH, Giboney P, Yee Jr HF, Tu SP, Tong EK. A proactive outreach strategy using a local area code to refer unassisted smokers in a safety net health system to a quitline: a Pragmatic Randomized Trial. Nicotine and Tobacco Research. 2023 Jan, 25(1):43-9.

[178] Nyangaresi VO, Morsy MA. Towards privacy preservation in internet of drones. In2021 IEEE 6th International Forum on Research and Technology for Society and Industry (RTSI) 2021 Sep 6 (pp. 306-311). IEEE.

[179] Bates DW, Levine DM, Salmasian H, Syrowatka A, Shahian DM, Lipsitz S, Zebrowski JP, Myers LC, Logan MS, Roy CG, Iannaccone C. The Safety of Inpatient Health Care. New England Journal of Medicine. 2023 Jan 12, 388(2):142-53.

[180] Mehmood A, Natgunanathan I, Xiang Y, Hua G, Guo S. Protection of big data privacy. IEEE access. 2016 Apr 27, 4:1821-34.

[181] Mohammadian E, Noferesti M, Jalili R. FAST: fast anonymization of big data streams. InProceedings of the 2014 international conference on big data science and computing 2014 Aug 4 (pp. 1-8).

[182] Xu K, Yue H, Guo L, Guo Y, Fang Y. Privacy-preserving machine learning algorithms for big data systems. In2015 IEEE 35th international conference on distributed computing systems 2015 Jun 29 (pp. 318-327). IEEE.

[183] Wei L, Zhu H, Cao Z, Dong X, Jia W, Chen Y, Vasilakos AV. Security and privacy for storage and computation in cloud computing. Information sciences. 2014 Feb 10, 258:371-86.

[184] Nyangaresi VO, Mohammad Z. Session Key Agreement Protocol for Secure D2D Communication. InThe Fifth International Conference on Safety and Security with IoT: SaSeIoT 2021 2022 Jun 12 (pp. 81-99). Cham: Springer International Publishing.

[185] Kotenko IV, Saenko I, Kushnerevich A. Parallel big data processing system for security monitoring in Internet of Things networks. J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl.. 2017 Dec, 8(4):60-74.

[186] De Pascale D, Cascavilla G, Tamburri DA, Van Den Heuvel WJ. Real-world K-Anonymity applications: The KGen approach and its evaluation in fraudulent transactions. Information Systems. 2023 May 1, 115:102193.

[187] Narayanan U, Paul V, Joseph S. A novel system architecture for secure authentication and data sharing in cloud enabled Big Data Environment. Journal of King Saud University-Computer and Information Sciences. 2022 Jun 1, 34(6):3121-35.

[188] Juma M, Alattar F, Touqan B. Securing Big Data Integrity for Industrial IoT in Smart Manufacturing Based on the Trusted Consortium Blockchain (TCB). IoT. 2023 Feb 6, 4(1):27-55.

[189] Lin G, Zhang H, Song X, Shibasaki R. Blockchain for location-based big data-driven services. InHandbook of Mobility Data Mining 2023 Jan 1 (pp. 153-171). Elsevier.

[190] Nyangaresi VO. A Formally Verified Authentication Scheme for mmWave Heterogeneous Networks. Inthe 6th International Conference on Combinatorics, Cryptography, Computer Science and Computation (605-612) 2021.

[191] Subbiah V. The next generation of evidence-based medicine. Nature Medicine. 2023 Jan 16:1-0.

[192] Abouelmehdi K, Beni-Hssane A, Khaloufi H, Saadi M. Big data security and privacy in healthcare: A Review. Procedia Computer Science. 2017 Jan 1, 113:73-80.

[193] Zhang H, McKenzie G. Rehumanize geoprivacy: from disclosure control to human perception. GeoJournal. 2023 Feb, 88(1):189-208.

[194] Zhang R, Wu X. Privacy Preservation Method Based on Clustering Interference Algorithm in Social Networks. Journal of Engineering Science & Technology Review. 2022 Mar 1, 15(2).

[195] Zhou H, Wen Q. Data security accessing for HDFS based on attribute-group in cloud computing. InInternational conference on logistics engineering, management and computer science (LEMCS 2014) 2014 May (pp. 1140-1143). Atlantis Press.

[196] Eid MM, Arunachalam R, Sorathiya V, Lavadiya S, Patel SK, Parmar J, Delwar TS, Ryu JY, Nyangaresi VO, Rashed AN. QAM receiver based on light amplifiers measured with effective role of optical coherent duobinary transmitter. Journal of Optical Communications. 2022 Jan 17.

[197] Chatterjee Y, Bourreau E, Rančić MJ. Solving various NP-Hard problems using exponentially fewer qubits on a Quantum Computer. arXiv preprint arXiv:2301.06978. 2023 Jan 17.

[198] Kaur H, Hooda N, Singh H. k-anonymization of social network data using Neural Network and SVM: K-NeuroSVM. Journal of Information Security and Applications. 2023 Feb 1, 72:103382.

[199] Orooji M, Rabbanian SS, Knapp GM. Flexible adversary disclosure risk measure for identity and attribute disclosure attacks. International Journal of Information Security. 2023 Jan 5:1-5.

[200] Rodríguez-Priego N, Porcu L, Peña MB, Almendros EC. Perceived customer care and privacy protection behavior: The mediating role of trust in self-disclosure. Journal of Retailing and Consumer Services. 2023 May 1, 72:103284.

[201] Sedayao J, Bhardwaj R, Gorade N. Making big data, privacy, and anonymization work together in the enterprise: experiences and issues. In2014 IEEE International Congress on Big Data 2014 Jun 27 (pp. 601-607). IEEE.