DIFFERENTIAL PRIVACY BY DESIGN: ENSURING COVID-19 DATA PRIVACY ON FEDERATED LEARNING SYSTEMS

_____

Anwaar Ulhaq
Machine Vision and Digital Health Research Group,
Charles Sturt University, Port Macquarie, NSW, Australia
aulhaq@csu.edu.au

Mike Horry
University of Technology Sydney, NSW, Australia

Ammar Haider
Charles Sturt University, Wagga Wagga, NSW, Australia

Oliver Burmeister
Charles Sturt University, Bathurst, NSW, Australia

ABSTRACT

To address COVID-19 healthcare challenges, we need frequent sharing of health data,

knowledge and resources at a global scale. However, in this digital age, data privacy is a big

concern that requires the secure embedding of privacy assurance into the design of all

technological solutions that use health data. In this paper, we introduce differential privacy by

design (dPbD) framework and discuss its embedding into the federated machine learning

system. To limit the scope of our paper, we focus on the problem scenario of COVID-19

imaging data privacy for disease diagnosis by computer vision and deep learning approaches.

We discuss the evaluation of the proposed design of federated machine learning systems and

discuss how differential privacy by design (dPbD) framework can enhance data privacy in

federated learning systems with scalability and robustness. We argue that scalable

differentially private federated learning design is a promising solution for building a secure,

private and collaborative machine learning model required to combat challenges like COVID-19.

Keywords: Differential privacy, COVID-19, Privacy by design (PbD), Federated learning systems

1 Introduction

COVID-19 pandemic has changed our world and its challenges (Gasser, Ienca, Scheibner, Sleigh, & Vayena, 2020). Today, more than ever, we require a centralised platform and collective approach to facilitate our collaborative research efforts in different scientific disciplines (Ienca & Vayena, 2020; Keesara, Jonas, & Schulman, 2020). Artificial intelligence, especially computer vision, has responded strongly to this challenge (Naudé, 2020). Various imaging modalities are being processed and analysed for COVID-19 control (Horry, Paul, et al., 2020; Ulhaq, Khan, Gomes, & Pau, 2020). These approaches vary from disease diagnosis and prognosis to disease prevention and management based on different imaging modalities like digital chest x-ray radiography (CXR), chest computed tomography (CT) and Lung ultrasound (LUS) (Gomes et al., 2020; Horry, Chakraborty, et al., 2020; Ulhaq, Born, et al., 2020). An ethical study of imaging data requires the privacy, confidentiality and integrity throughout data analysis. However, it seems a tremendous task due to the existing vulnerabilities of traditional machine learning systems that heavily rely on shared datasets for their training. Similarly, new data protection regulations like General Data Protection Regulation (GDPR) (Voigt & Von dem Bussche, 2017) in the European Union put restrictions on the moving data outside of their regional territories. This situation requires a fundamental change in the ways machine learning systems can work collaboratively.

Privacy by design (PbD) approach (Cavoukian et al., 2009) ensures that privacy assurance is embedded into a system design lifecycle by default from the beginning to the end. Therefore, any system built without including privacy considerations as a core part of their design process often behaves poorly in terms of privacy control. A post-GDPR world requires a focus on privacy by design with data privacy at the core of system design (Rubinstein, 2011). PbD has a special place in the design of machine learning systems, especially deep learning systems that are going to share our digital future. Google introduced a collaborative machine learning system (Konečný, McMahan, & Ramage, 2015; Konečný, McMahan, Ramage, & Richtárik, 2016; Yang, Liu, Chen, & Tong, 2019) with embedded PbD named as federated learning in 2016 that nullified the need for a centralised training data. The fundamental idea is to use different clients or nodes to train local machine learning models on local data samples without any exchange, and sharing of model parameters (e.g. the weights and biases) between these local nodes at some frequency to generate a global model by a process called federated averaging. All clients or nodes then share this global model. As no data sharing takes place between the nodes, federated design simplicity ensures data privacy. These systems are ideally suited for training machine learning algorithms for COVID-19 control as sensitive COVID-19 health data will not be shared and would remain in the custody of their subjects. Figure 1 illustrates privacy by design approach to machine learning with a federated model design for COVID-19 detection using CXR.
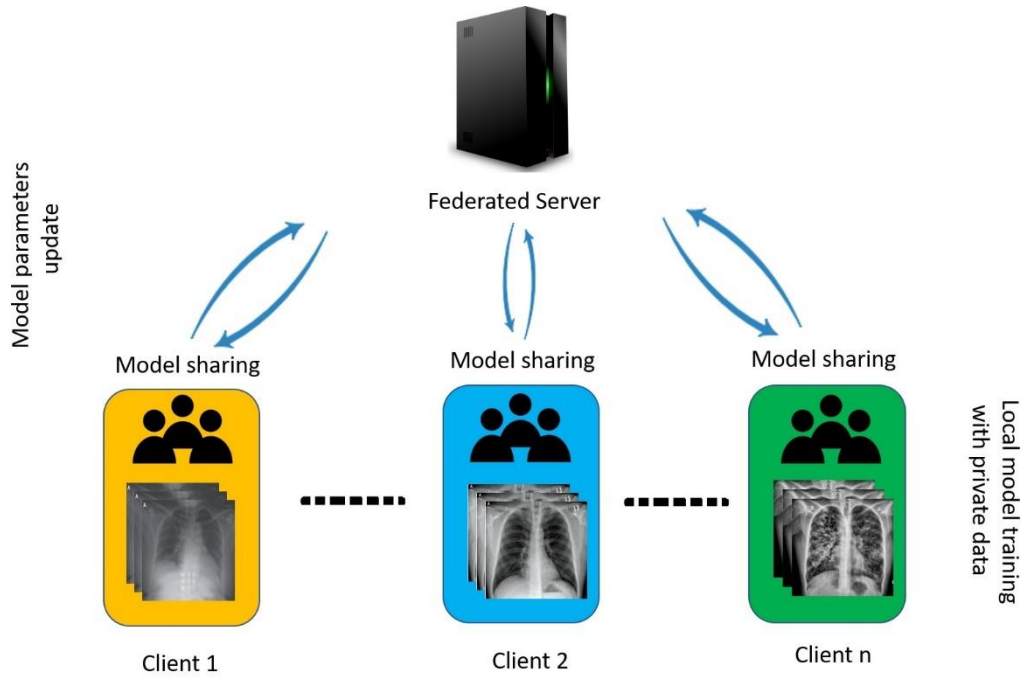
Figure 1: A federated learning system designed for COVID-19 X-ray imaging Data is shown. It illustrates how diagnosis based on data from different clients can be performed without any sharing of data.

However, this process does not guarantee data protection and system robustness. Attackers can figure out individual data based on the global model (Bhagoji, Chakraborty, Mittal, & Calo, 2019; Li, Sahu, Talwalkar, & Smith, 2020). Differential privacy adds random noise to an individual's model, obscuring the results (Dwork, 2008; Ji, Lipton, & Elkan, 2014). This random data can be added before the model is shared with the server, without revealing the actual data. This process preserves the individual's privacy. However, this design approach has limited scalability and robustness. Additionally, this integration is hard to grasp for system designers due to its unreasonable complexity. Inspired from the preliminary empirical works on differentially-private federated learning, in this paper, we introduce a theoretical

framework called differential privacy by design (dPbD) that can help to design scalable and robust federating learning systems for COVID-19 data privacy. We address the following research questions in this paper:

- How can we devise a theoretical framework that underpins all salient and impeding factors that impact the design of differentially-private federated learning systems with scalability and robustness?
- How can differentially-private federated learning systems ensure the privacy of COVID-19 imaging data without adversely harming the accuracy?

The contributions of this paper are as follows: We extend privacy by design framework in the context of differential privacy and federated learning. We propose a theoretical framework called differential privacy by design (dPbD) that underpins salient and impeding factors that impact the design of differentially-private federated learning systems with scalability and robustness. We discuss how the proposed framework can be used to model privacy of COVID-19 imaging data for training differentially-private federated learning systems.

We have organised the paper as follow: Section 2 describes the related work. A brief introduction to preliminary techniques that have inspired the development of this theoretical framework is discussed in section 3. The proposed framework is presented in section 4. We discuss seven principles to implement this design in section 5. Embedding of the proposed principles into federated learning system design for COVID-19 data is discussed in section 6. It is followed by a discussion and concluding remarks.

2 Related Work

At an early stage of COVID-19, authorities recognised the importance of privacy and public trust to fight against the pandemic. Some data privacy experts also dubbed it as an unusual scenario like the European Data Protection Board that highlighted article 9 of the General Data Protection Regulation. This article allows the processing of personal data "for reasons of public interest in the area of public health, such as protecting against serious cross-border threats to health (Ienca & Vayena, 2020). However, careful data-management practices in our data-intensive world require giving critical consideration to data privacy aspects. COVID-19 poses various challenges to the Biomedical Imaging (BI), Artificial Intelligence (AI), Data Analytics (DA), Computer Vision (CV) and Machine learning (ML) as these disciplines require adequate access to big data to detect, diagnose, and predict the spread of the infection (Gomes et al., 2020; Ulhaq, Born, et al., 2020; Ulhaq, Khan, et al., 2020). This article restricts its scope to data privacy of COVID-19 medical imaging data required by computer vision and machine learning approaches for COVID-19 control.

Privacy by design is considered as an approach to embedding privacy directly into system design and introduced by Ann Cavoukian in 2009 (Cavoukian et al., 2009; Wahlstrom, Ulhaq, Burmeister, et al., 2020). This design is incorporated into the European GDPR that shows its importance (Voigt & Von dem Bussche, 2017). Privacy-preserving machine learning uses privacy by design approach in its framework (Mohassel & Zhang, 2017). Most recently, Google introduced a federated machine learning framework for preserving the privacy of the training data (Konečnỳ et al., 2015, 2016). Federated learning systems use distributed framework where model training is performed locally at the client-side, then the updated model parameters are sent to a central server for aggregation. Contrary to the idea of

data fusion (Mirza, Qamar, et al., 2005), a global model aggregation or fusion is used. Several works (Li et al., 2020; Yang et al., 2019) discussed design, challenges and future direction in the field of federated learning. We refer the interested readers to these articles. Federated learning systems inherently support privacy by design.

However, PbD is criticised for being unclear, complex to enforce its implementation, and difficult to adapt to certain disciplines (Rubinstein, 2011). Another line of work in the privacy domain takes a huge interest in leveraging differential privacy into machine learning design for privacy-preserving AI. We refer the interested readers to review these articles (Dwork, 2008; Ji et al., 2014). With the increased popularity and enhanced performance of deep learning, differential privacy is proposed for deep learning models and federated learning (Abadi et al., 2016; Kumar et al., 2020; Wei et al., 2020). The integration of differential privacy and federated learning leads to future research directions in privacy preserving AI. However, this marriage is incredibly complex and lacks clarity in its design due to various competing factors. The complexity of the preliminary work in this direction inspired us to propose a theoretical framework called differential privacy by federated design. It is specially designed to provide differential privacy to COVID-19 imaging data using a federated machine learning system.

3 Preliminaries:

3.1 Federated Machine Learning Systems

Federated learning is a powerful framework that is devised for machine learning scientists to work collaboratively with decentralised data with privacy by default setting. Google generated the initial idea as part of the series of works in 2015 and 2016 (Konečnỳ et

al., 2015, 2016). The initial focus of the federated design was on on-device federated learning tailored to distributed mobile-user interactions. A sample example is Gboard app that predicts the next word to make typing effortless based on typed text using a Federated Recurrent neural network (RNN) model (Xu et al., 2020).

Consider a scenario of n machine learning scientists $\{s_1, \ldots, s_n\}$ working on COVID-19 imaging data, all of whom wish to collaborate to train a machine-learning model by sharing their respective imaging data $\{d_1, \ldots, d_n\}$. A federated-learning system defines a distributed machine learning process in which the ML scientists collaboratively train a shared model M_f. However, this process ensures that any ML scientist s_i does not expose its data d_i to any other scientist s_j during the local training process. Each of the data scientists, however, shares its local model m_i to the server (centralised architecture) (Liu, Yan, Zhou, Yang, & Zhang, 2020) or on a blockchain (decentralised architecture) (Kumar et al., 2020). A federating averaging is used for model aggregation to get a shared model M_f that is then sent back to each ML scientist.

Various improvements are proposed in the literature after the initial design (Konečnỳ et al., 2015, 2016). It includes horizontal and vertical federated learning architectures, federated transfer learning, federated domain adaptation, federated adversarial learning, improving communication protocols and security, and making federated learning more personalisable. It is emerging as a promising research topic in machine learning. Interested readers, please refer to the topic reviews (Li et al., 2020; Yang et al., 2019).

3.2 Differential Privacy in Machine Learning

Differential privacy (DF) defines a formal assurance of anonymity and indistinguishability in terms of a privacy budget (ϵ)—the smaller the budget, the stronger the confidence on privacy (Dwork, 2008; Ji et al., 2014). The topic has its roots in aggregate or adjacent databases. In the case of COVID-19 imaging datasets, each COVID-19 imaging training dataset is a set of image-label pairs for supervised learning. Any two of these datasets are adjacent if they vary in a single entry, that is, one image-label pair is present in one image dataset and absent in the other. In other words, two databases X and Y are neighbours or adjacent if $A(X, Y) = 1$ where A is the distance. A randomised mechanism $K: D \rightarrow R$ with domain D and range R preserves (ϵ, δ) differential privacy, if for any pair of adjacent databases (X, Y) belonging to D and set S of possible outputs: $Pr[K(X) \in S] \leq \epsilon \, Pr[K(Y) \in S] + \delta$.

A randomised mechanism like Gaussian mechanism (GM) (Geyer, Klein, & Nabi, 2017) approximates a real-valued function F for these neighbouring datasets, and differential privacy can be enforced by adding noise to the model to the scale depending on the sensitivity of this function function F. The global sensitivity GS of a function F is defined as:

$$GS_F = max_{X,Y:D(X,Y)=1}|F(X) - F(Y)|$$

DF provides a mathematically provable guarantee of privacy protection against a wide range of privacy attacks (include differencing attack, linkage attacks, and reconstruction attacks). Decreasing epsilon leads to a decrease in inaccuracy, where accuracy is defined to be the closeness of the output of DP algorithms to pure production. ϵ is a metric of privacy loss at a differentially change in data (adding, removing one entry). The smaller the value is, the better the privacy protection. $F(X)$ can be released accurately when F is insensitive to individual

models (Pillutla, Kakade, & Harchaoui, 2019). Many natural functions have low GS like sample mean and covariance matrix. To achieve a small global sensitivity, the ideal condition is that all the clients use sufficient local datasets for training (Wei et al., 2020).

4 The proposed Differential Privacy by Design (dPbD) framework

In privacy-preserving machine learning, we search for an algorithm that takes as input a dataset (sampled from some distribution), and then privately output a hypothesis h that with high probability has low error over the distribution.

Our proposed theoretical framework is extending privacy by design framework (Cavoukian et al., 2009) for federating machine learning design by using the notion of embedding differential privacy into its system design from end to end. Figure 2 provides a visual description of our proposed differential privacy by design (dPbD) framework.

This framework is partially inspired by empirical studies (Abadi et al., 2016; Geyer et al., 2017; Kumar et al., 2020; Liu et al., 2020; Wei et al., 2020) on the use of differential privacy for machine learning. From these empirical studies, two major dimensions consistently emerged. One dimension differentiates between privacy and utility and highlights their trade-off—utility in machine learning can be taken as model accuracy or decrease in testing loss. The other dimensions are design scalability and robustness. Here scalability

means scalability of federated learning model and number of clients. Robustness indicates system performance against attacks.
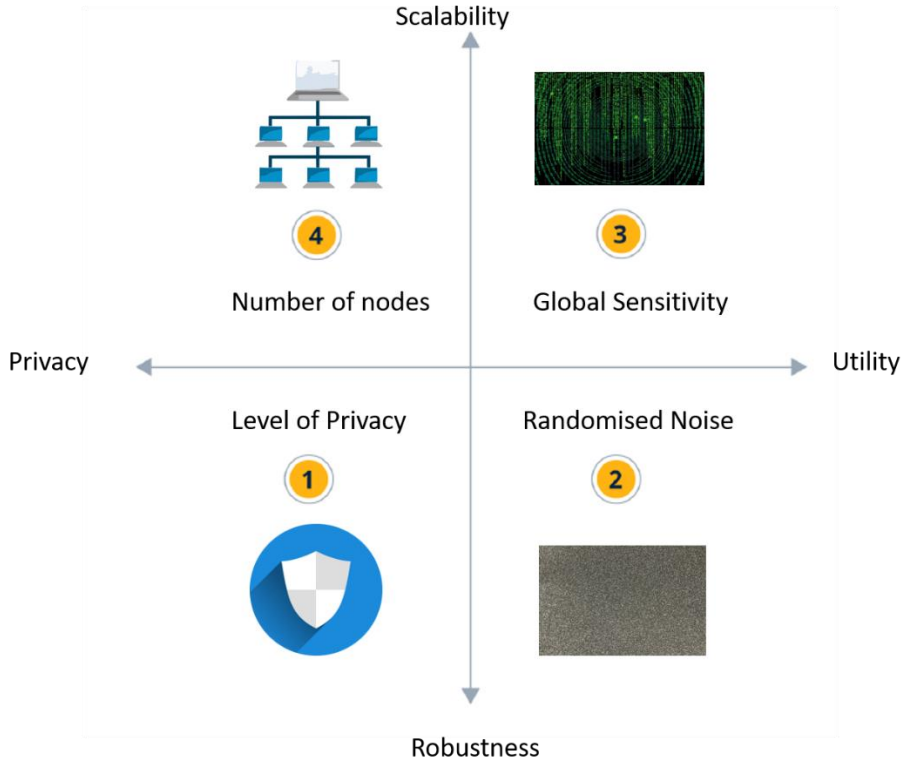

Figure 2: The proposed differential privacy by design (dPbD) framework.

There are four quadrants and thus four important junctions of the proposed framework that define most important factors in defining differential privacy in federated learning design. These factors include the level of privacy, level of randomised noise, global sensitivity and number of clients or nodes in a differentially private federated learning system.

As differential privacy guarantees anonymity and indistinguishability in terms of a privacy budget (epsilon)—the smaller the budget, the stronger the confidence on privacy, and we call it a level of privacy. It is significant characteristics of differential privacy by design as it guarantees and provides a quantitative notion of privacy compared to the unclear concept of privacy in privacy by design framework. The level of privacy defines system robustness

against attacks. Similarly adding more noise as part of differential privacy can reduce system utility and robustness. The amount of randomised noise is thus an important factor to balance the trade-off between privacy and utility.

On the other hand, scalability requires an increase in the number of nodes and the effectiveness of robust aggregation. Robust aggregation depends on global sensitivity. A small value of global sensitivity requires that all the clients use sufficient local datasets for training, and the type of aggregation function. Similarly, functions having low global sensitivity are preferred for better performance and utility like sample mean and covariance matrix.

5 The 7 Foundational Principles

We advise these seven principles to accomplish proposed differential privacy by design framework for system design.

- Privacy can be guaranteed in design: The system must ensure a meaningful privacy guarantee. For instance, choosing a smaller epsilon produces noisier results and better privacy guarantees in differential privacy. Privacy guarantee during system design will build trust.

- Privacy can be quantified in design: Differential privacy can be used to quantify privacy. The strategy of using budgets, expenses and losses in terms of privacy is known as privacy accounting. The maximum privacy loss is called the privacy budget. This quantification leads to better privacy-preserving design.

- Privacy by Modularity: Modularity is the key to ensure privacy as it reduces complexity (Bryson & Stein, 2001). Modules can be removed, replaced, or upgraded

without affecting other components. Privacy of the system should not be affected by removing, replacing, or upgrading any system component.

- Privacy Embedded into Federated Design: Privacy is preserved in model aggregation with low global sensitivity. The aggregated model should also be insensitive to local data at different nodes.

- Privacy with Scalability: Privacy notion is unaffected by scaling the system. Any change in scalability for Federated Learning system and asynchronous or synchronous training algorithms should not degrade privacy.

- Anonymity in Design Lifecycle: Original data is never shared. Only modified function or model parameters are transmitted. Data anonymity must be ensured in complete Design Lifecycle.

- Optimising Privacy-Utility Trade-off in Design: we refer utility to certain system properties (Makhdoumi & Fawaz, 2013). Increase in privacy often causes a decrease in utility. A design should optimise this trade-off by ensuring privacy without adversely harming system utility.

6 A Scenario of COVID-19 Imagaing Data Privacy By Design

The privacy of COVID-19 imaging data on a collaborative machine learning network (federated machine learning) can be taken as a case study of differential privacy by design framework. Its seven foundation principles can be applied to ensure privacy while designing collaborative machine learning systems. Here, we describe a brief view of the implementation of these seven principles in the scenario of the training machine learning model on federated learning systems.

- Privacy-preserving federated learning design must control the level of privacy to control perturbation and noise for robustness. However, the design must provide adequate level of privacy to prevent patients' personal information embedding or related to their COVID-19 imaging data.

- Privacy level must be quantified in federated design. There must be understanding and consensus between teams involved in federated learning setup about the minimum and maximum level of privacy required for collaborative research.

- Adequate privacy must be enforced during design at all system components like clients, federated server and communication channels. Addition or removal of clients and change in client-side learning model should not affect data privacy promised by the system.

- A robust model aggregation strategy must be adopted. A function with low global sensitivity be chosen while designing any federated system for COVID-19. The robust aggregation should be insensitive to local changes and changes in data on client nodes.

- COVID-19 imaging data is never shared throughout the training lifecycle. Model sharing is protected and maintains the anonymity of patients' information through the model upgradation and communication rounds.

- The system design must be flexible, and adding and removing clients and teams or increasing or decreasing COVID-19 data size should not adversely harm patient data and shared model privacy. Horizontal and vertical scalability be supported during design with features of robustness and resilience.

- Accuracy is an important consideration in COVID diagnosis, and it should not be harmed. It is critical to reduce or sacrifice it for the sake of privacy as it can adversely affect system utility. An optimised trade-off be sought during the design.

7 Discussion and Concluding Remarks

The proposed differential privacy by design framework is focused on the design of privacy-preserving federated machine learnng systems. This theoretical framework is developed by getting inspiration from various empirical studies about the use of differential privacy in federated learning. However, we discovered that while the majority of the proposed systems emphasise on the trade-off between privacy and utility, they often ignore scalability and robustness of the system. Our proposed framework fills up that gap and proposes a comprehensive framework that covers the majority of the design concepts. We argue that following the footsteps of privacy by design framework, differential privacy must be embedded throughout the design lifecycle, and it should provide overall coverage of protection and privacy to any proposed federated machine learning system. To pave the way of this embedding in design lifecycle, we defined seven foundational principles similar to seven principles of privacy by design framework.

We use a case study of COVID-19 imaging data. However, our proposed framework should be applicable to all types of data on a federated learning system. In our future work, we would implement and validate these principles for designing a computer vision-based COVID-19 diagnosis system based on pathological imaging data available from various teams around the world. We hope that the proposed framework with impact designing a privacy-preserving federated learning system with reduced complexity and sufficient data protection for collaborative research to combat COVID-19 challenge.

References

Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security* (pp. 308–318).

Bhagoji, A. N., Chakraborty, S., Mittal, P., & Calo, S. (2019). Analysing federated learning through an adversarial lens. In *International conference on machine learning* (pp. 634–643).

Bryson, J. J., & Stein, L. A. (2001). Modularity and design in reactive intelligence. In *International joint conference on artificial intelligence* (Vol. 17, pp. 1115–1120).

Cavoukian, A., et al. (2009). Privacy by design: The 7 foundational principles. *Information and privacy commissioner of Ontario, Canada*, *5*.

Dwork, C. (2008). Differential privacy: A survey of results. In *International conference on theory and applications of models of computation* (pp. 1–19).

Gasser, U., Ienca, M., Scheibner, J., Sleigh, J., & Vayena, E. (2020). Digital tools against covid-19: taxonomy, ethical challenges, and navigation aid. *The Lancet Digital Health*.

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

Gomes, D. P., Ulhaq, A., Paul, M., Horry, M. J., Chakraborty, S., Saha, M., … Rahaman, D. (2020). Potential features of ICU admission in x-ray images of covid-19 patients. *arXiv preprint arXiv:2009.12597*.

Horry, M. J., Chakraborty, S., Paul, M., Ulhaq, A., Pradhan, B., Saha, M., & Shukla, N. (2020). Covid-19 detection through transfer learning using multimodal imaging data. *IEEE Access*, *8*, 149808–149824.

Horry, M. J., Paul, M., Ulhaq, A., Pradhan, B., Saha, M., Shukla, N., et al. (2020). X-ray image based covid-19 detection using pre-trained deep learning models.

Ienca, M., & Vayena, E. (2020). On the responsible use of digital data to tackle the covid-19 pandemic. *Nature medicine*, *26*(4), 463–464.

Ji, Z., Lipton, Z. C., & Elkan, C. (2014). Differential privacy and machine learning: a survey and review. *arXiv preprint arXiv:1412.7584*.

Keesara, S., Jonas, A., & Schulman, K. (2020). Covid-19 and health care's digital revolution. *New England Journal of Medicine*, *382*(23), e82.

Konečnỳ, J., McMahan, B., & Ramage, D. (2015). Federated optimisation: Distributed optimisation beyond the datacenter. *arXiv preprint arXiv:1511.03575*.

Konečnỳ, J., McMahan, H. B., Ramage, D., & Richtárik, P. (2016). Federated optimisation: Distributed machine learning for on-device intelligence. *arXiv preprint arXiv:1610.02527*.

Kumar, R., Khan, A. A., Zhang, S., Wang, W., Abuidris, Y., Amin, W., & Kumar, J. (2020). Blockchain-federated-learning and deep learning models for covid-19 detection using CT imaging. *arXiv preprint arXiv:2007.06537*.

Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine*, *37*(3), 50–60.

Liu, B., Yan, B., Zhou, Y., Yang, Y., & Zhang, Y. (2020). Experiments of federated learning for covid-19 chest x-ray images. *arXiv preprint arXiv:2007.05592*.

Makhdoumi, A., & Fawaz, N. (2013). Privacy-utility tradeoff under statistical uncertainty. In *2013 51st annual allerton conference on communication, control, and computing (allerton)* (pp. 1627–1634).

Mirza, A., Qamar, S., et al. (2005). An optimised image fusion algorithm for night-time surveillance and navigation. In *Proceedings of the IEEE symposium on emerging technologies, 2005.* (pp. 138–143).

Mohassel, P., & Zhang, Y. (2017). Secureml: A system for scalable privacy-preserving machine learning. In *2017 IEEE symposium on security and privacy (sp)* (pp. 19–38).

Naudé, W. (2020). Artificial intelligence vs covid-19: limitations, constraints and pitfalls. *AI & Society*, 1.

Pillutla, K., Kakade, S. M., & Harchaoui, Z. (2019). Robust aggregation for federated learning. *arXiv preprint arXiv:1912.13445*.

Rubinstein, I. S. (2011). Regulating privacy by design. *Berkeley Tech. LJ*, *26*, 1409.

Ulhaq, A., Born, J., Khan, A., Gomes, D., Chakraborty, S., & Paul, M. (2020). Covid-19 control by computer vision approaches: A survey. *IEEE Access*.

Ulhaq, A., Khan, A., Gomes, D., & Pau, M. (2020). Computer vision for covid-19 control: A survey. *arXiv preprint arXiv:2004.09420*.

Voigt, P., & Von dem Bussche, A. (2017). The EU general data protection regulation (GDPR). *A Practical Guide, 1st Ed., Cham: Springer International Publishing*.

Wahlstrom, K., Ul-haq, A., Burmeister, O., et al. (2020). Privacy by design. *Australasian Journal of Information Systems*, *24*.

Wei, K., Li, J., Ding, M., Ma, C., Yang, H. H., Farokhi, F., … Poor, H. V. (2020). Federated learning with differential privacy: Algorithms and performance analysis. *IEEE Transactions on Information Forensics and Security*.

Xu, D., Li, T., Li, Y., Su, X., Tarkoma, S., & Hui, P. (2020). A survey on edge intelligence. *arXiv preprint arXiv:2003.12172*.

Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *10*(2), 1–19.