


Review

# Non-IID Medical Imaging Data on COVID-19 in the Federated Learning Framework: Impact and Directions

Fatimah Saeed Alhafiz \* and Abdullah Ahmad Basuhail \*

Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia

\* Correspondence: fabdullahalhafiz@stu.kau.edu.sa or fshafed@ub.edu.sa (F.S.A.);  
abasuhail@kau.edu.sa (A.A.B.)

**Abstract:** After first appearing in December 2019, coronavirus disease 2019 (COVID-19) spread rapidly, leading to global effects and significant risks to health systems. The virus's high replication competence in the human lung accelerated the severity of lung pneumonia cases, resulting in a catastrophic death rate. Variable observations in the clinical testing of virus-related and patient-related cases across different populations led to ambiguous results. Medical and epidemiological studies on the virus effectively use imaging and scanning devices to help explain the virus's behavior and its impact on the lungs. Varying equipment resources and a lack of uniformity in medical imaging acquisition led to disorganized and widely dispersed data collection worldwide, while high heterogeneity in datasets caused a poor understanding of the virus and related strains, consequently leading to unstable results that could not be generalized. Hospitals and medical institutions, therefore, urgently need to collaborate to share and extract useful knowledge from these COVID-19 datasets while preserving the privacy of medical records. Researchers are turning to an emerging technology that enhances the reliability and accessibility of information without sharing actual patient data. Federated learning (FL) is a technique that learns distributed data locally, sharing only the weights of each local model to compute a global model, and has the potential to improve the generalization of diagnosis and treatment decisions. This study investigates the applicability of FL for COVID-19 under the impact of data heterogeneity, defining the lung imaging characteristics and identifying the practical constraints of FL in medical fields. It describes the challenges of implementation from a technical perspective, with reference to valuable research directions, and highlights the research challenges that present opportunities for further efforts to overcome the pitfalls of distributed learning performance. The primary objective of this literature review is to provide valuable insights that will aid in the formulation of effective technical strategies to mitigate the impact of data heterogeneity on the generalization of FL results, particularly in light of the ongoing and evolving COVID-19 pandemic.



**Citation:** Alhafiz, F.S.; Basuhail, A.A. Non-IID Medical Imaging Data on COVID-19 in the Federated Learning Framework: Impact and Directions. *COVID* **2024**, *4*, 1985–2016. <https://doi.org/10.3390/covid4120140>

Academic Editors: Somenath Chakraborty and Simone Brogi

Received: 5 November 2024

Revised: 5 December 2024

Accepted: 10 December 2024

Published: 16 December 2024

**Keywords:** COVID-19 lung medical image; federated learning; data heterogeneity; non-IID type; generalization; personalization



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

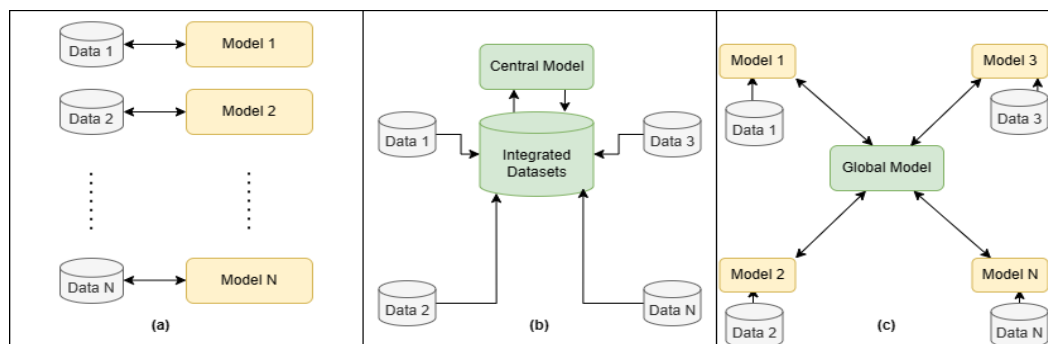
## 1. Introduction

COVID-19 is a worldwide pandemic first reported in December 2019 and continues to affect people all over the world, with the virus transforming into multiple strains [1]. The number of infected cases has increased exponentially, and the number of reported deaths reached more than 7 million in 2024, according to the World Health Organization (WHO) [2]. The pulmonary system serves as the principal target of infection for severe acute respiratory syndrome (COVID-19), with profound hypoxemia identified as the leading cause of mortality in the most severe instances.

COVID-19 exhibits extremely heterogeneous clinical manifestations regarding its severity, clinical presentation, and, significantly, its worldwide prevalence [3]. While most individuals diagnosed with this acute infection ultimately achieve recovery [4], a substantial

number experience long-term complications that impact multiple organ systems, including the lungs [5]. The precise pathobiological mechanisms underlying the pulmonary vascular complications associated with COVID-19, particularly in both the acute and chronic phases of the disease, remain inadequately elucidated and are not fully comprehended within the current scientific literature [3]. The lack of reliable information about the behavior of COVID-19 (e.g., how it is spread, variants of individual symptoms, and unstable responses to treatment strategies) has created a need for effective collaboration to collect more data about the virus.

Imaging equipment could be applied to understand the virus behaviors and help diagnose lung damage in the initial phase of the infection, thus providing justifiable relationships between different populations. In addition, screening for uncertain, risky cases or reducing the time and complexity of manual reverse transcription–polymerase chain reaction (RT-PCR) tests which reduced the misdiagnosis rate of manual RT-PCR tests by 30% [6]. However, due to the continuous increase in the number of COVID-19 infections, many medical images have been produced, which leaves hospitals with two challenges. First, interpreting medical images requires radiology experts for manual labeling, segmenting, and annotating, and the limited number of radiologists in hospitals creates challenges for accumulating these data. There is, therefore, a need to assist radiologists working alone to develop individual deep learning (DL) models locally as embedded software in computer-aided diagnosis (CAD) systems to interpret these images promptly for the diagnosis, treatment, and prognosis of COVID-19 during lung testing, as shown in Figure 1a [7].



**Figure 1.** Training techniques for distributed data: (a) individual training technique, (b) centralizing technique, and (c) federated learning technique.

The second challenge is the reported results of an individual training model of data from single sources, which lack a diversity of infected cases and produce bias due to the locality of the patient population. Hence, the generalization measure of that model will be inefficient for out-of-sample testing of new data [8]. The alternative solution of collecting and standardizing medical images at a single point may be effective for improving diversity and generalization problems, as shown in Figure 1b. However, this requires a large storage capacity, computational resources, communication bandwidth, and security management for accessing and retrieving data at the central data point [9]. Furthermore, data governance policies preserve patient privacy and prevent the sharing of medical data that may reveal sensitive information about patients, even with anonymization technology, which limits the efforts of medical and health institutions to benefit from DL systems [10].

Federated learning (FL) is a paradigm that enables distributed points to train their data locally without the need to share the actual data, as shown in Figure 1c. It collects only the weights (i.e., learning parameters/gradients) from different parties and computes the global model from distributed training. FL techniques are successful for medical image applications, such as disease diagnosis [11], segmentation [8], and treatments [12], and they result in higher accuracy than centralized learning [13]. The distributed processing

of sensitive information has motivated researchers to utilize FL to overcome the lack of information about COVID-19 [14].

The actual implementation of an FL framework presents several open technical issues, such as data and system heterogeneity and privacy and security challenges. This study focuses on the issues relating to data heterogeneity and reviews the potential opportunities for using FL on the available COVID-19 lung medical imaging data for different applications. Additionally, it describes the open issues of FL, available solutions in the medical imaging field, and other suitable solutions in medical applications. The significant contributions of this study are the following:

- This study identifies the applicability and benefits of implementing FL to process and train distributed COVID-19 lung data using various imaging modalities and equipment, identifying the imaging types and modes available in distributed hospitals and medical institutions.
- It provides an overview of the FL system and describes the variables of implementation and the practical constraints in medical fields. It also investigates the progress made in developing FL frameworks to train medical images and identifies areas that require further effort to overcome the pitfalls of distributed learning performance.
- This article provides detailed descriptions of the data heterogeneity issue, identifies the metrics that might be affected by that issue, and offers a mathematical description of the problem for each type of skewness, along with valuable research directions to mitigate the impact of data heterogeneity.
- It emphasizes other prevalent FL issues in a concise manner to offer a comprehensive perspective for research on the FL environment.
- This study uses imaging data to outline potential avenues for future research to explore how COVID-19 affects the lung and internal organs, referencing ongoing studies that consider relevant factors from a medical and radiology standpoint.

The organizational structure of this manuscript is as follows: Section 2 describes the methods that were used to collect published studies. Section 3 lists related works and describes the main contributions of this review. Section 4 identifies the opportunities for using the FL technique for COVID-19 images to overcome the current FL issues. Section 5 provides an overview of the FL technique and the available COVID-19 medical images. Section 6 provides a comprehensive perspective of the data heterogeneity impact, skewness types, bias resources, and directions of current solutions to fix them and where the research has reached. Section 7 describes the current common issues of FL implementation for medical imaging. Section 8 provides the results of this literature review and discusses the investigations the results provided. Section 9 outlines the open directions and provides recommendations to improve the performance of FL in the medical image field. Finally, Section 10 includes a brief summary of this study.

## 2. Procedure

This review commenced with a comprehensive search across various scientific databases in the English language, including Google Scholar, Wiley, MedRxiv, IEEE, Springer, and other academic resources. The search utilized case-insensitive keywords such as “federated learning”, “medical images”, “non-IID”, “data heterogeneity”, and “COVID-19”. The initial search returned 28 papers. Subsequently, the search criteria were expanded to include broader terms like “federated learning” and “medical data”, which resulted in 33 papers and 11 surveys.

After reviewing the technical challenges associated with federated learning (FL) in medical imaging, we conducted individual searches for each issue to provide both high-level and detailed insights into the proposed solutions. Keywords such as “federated learning”, “distributed learning”, “data issues”, “non-IID”, “data heterogeneity”, “domain shift image”, “lung imaging”, “FL bias”, “medical data”, and “COVID-19” were used. This search returned 101 research papers, including both journal articles and conference papers, and 27 review papers.

These were excluded after full-text review due to failure to meet five quality criteria: clarity of research objectives, focus on human medical images, use of machine learning models in a federated learning environment, sufficient methodological details, and technical value to the medical imaging field. After reviewing these papers, only those proposing solutions to data heterogeneity in lung imaging data were included, resulting in 41 research papers and 8 reviews.

### 3. Related Works

Creating data silos for medical imaging achieves high generalizability and valuable observations as part of the Enhancing Neuro Imaging Genetics through Meta-analysis (ENIGMA) project. By integrating medical image data from 70 distributed sites, ENIGMA discovered factors related to brain disease that individual sites could not reach [15]. However, the process necessitated a high level of computations and a security budget to guarantee processing safety and generate generalizable models. Additionally, this violates privacy rules found in most governance data, which focus on preventing the disclosure of patient information.

Federated learning provides a framework for researchers in the medical field using large-scale data while eliminating the risks and costs of data centralization. This encourages hospitals and clinical institutions to develop or utilize available FL software. There are few studies that have reviewed FL frameworks using COVID-19 medical images, which relates to the results of guaranteed privacy-preserving methods in distributed training, which have recently been provided [10], which ensures that hospitals maintain patient privacy, as required by governance privacy laws. This article, therefore, delves into studies that examine the concept, design, and challenges of implementing federated learning (FL) on COVID-19 medical images, medical imaging, and the broader medical field.

We categorized the related works into three review study perspectives: technical, privacy-preserving and security, and FL for COVID-19 data. Technical-perspective-related work began with the work of by Rieke et al., one of the most cited reviews about FL in the medical field [16]. They described the applicability of FL for the digital health sector and discussed the benefits of FL software for medical stockholders and patients. They also expressed belief in a promising future for FL in the medical field by overcoming the technical issues referred to in their study. Erfan et al. conducted a two-part review study [17,18] that identified various types of FL algorithms in the health domain, particularly for medical image processing. They also reviewed the technical challenges associated with data heterogeneity, model bias, lack of standardization, privacy and security, and system architecture.

Xu et al. [19] conducted a similar review and described a set of proposed solutions for three types of technical implementation issues in the field of general medicine: statistical, communications, and privacy. However, they surveyed solutions from different fields outside of the medical area, handling different types of data that may have lower sensitivity to privacy and security. Hun Yoo et al. [20] conducted a technical survey discussing FL issues in medical data and provided detailed approaches toward resolving them. They provided an overview of the specific challenges in both medical data and FL scenarios to identify technical and security issues.

From a privacy-preserving and security perspective, Kaissis et al. [10] categorized FL attacks into two types: model and data. Due to the availability of medical image-sharing protocols in a standard Digital Imaging and Communication in Medicine (DICOM) format, their study utilized medical images, providing a comprehensive review of the FL scenarios and privacy-preserving and security methods applicable to medical images. However, they did not consider the overhead costs of the FL framework when implementing multiple security methods against the attacks they mentioned.

The research that reviewed methods using a COVID-19 dataset are more relevant to this study. Peifer-Smadja et al. [21] strongly advocated for collaborative efforts and the use of medical data related to COVID-19 to expedite the advancement of prognosis,

diagnosis, and treatment applications. Shuja et al. [22] also provided a comprehensive review of COVID-19 open-source datasets and their AI applications. Both studies highly recommend taking advantage of FL to mitigate the limitations of COVID-19 data. In depth, Mondal et al. [23] reviewed various datasets on medical imaging, including X-rays, computed tomography (CT) scans, and ultrasound images, considering the number of images, COVID-19 samples, and classes within the datasets. The authors comprehensively discussed the proposed FL methods that used pretrainable CNN models and compared reported results. Furthermore, Hwang et al. [24] provided insights into FL's fundamental concepts and discussed the key challenges related to the medical domain and adversarial attacks. They highlighted the promising applications of FL in the medical domain, using the recent COVID-19 pandemic as a use case. However, these review works did not consider the heterogeneity issue of medical data.

Shyu et al. [12] provides a comprehensive analysis of recent challenges in FL from a data-centric perspective, addressing issues such as data partitioning, distribution patterns, protection mechanisms, and benchmark datasets for healthcare applications. Additionally, for medical imaging data, non-IID issues are thoroughly discussed by [25] and categorized into data imbalance and heterogeneous imaging datasets. The authors further reviewed and evaluated the proposed solutions to address these issues, as well as highlighting privacy and communication challenges in FL.

Naz et al. [26] surveyed the proposed FL methods for detecting COVID-19 in patients using chest images but did not consider other applications of the FL model, such as segmentation, quantities, and treatments. This creates a need to review the efforts of researchers to utilize FL for COVID-19 medical images in different model applications. Numerous researchers have reviewed various sub-algorithms supporting FL for medical data, each from a unique perspective. They have also reviewed the general issues of using FL in distributed environments, such as privacy concerns [27] and communication methods [28].

This study explores what makes COVID-19 medical image data unique, how non-IID issues affect the performance of the global model, and the local behaviors of distributed sites from training steps all the way through to the fitting of the aggregative model per round. In this study, we categorize the type of data heterogeneity based on the different types of skewness and then analyze and review valuable and applicable solutions that mitigate this issue, thereby identifying potential research directions. Finally, we present our findings and recommendations, which may require further technical investigations.

#### 4. FL Opportunities for COVID-19 Lung Imaging

In the medical field, DL has reduced the time and cost of making clinical decisions regarding diagnosis, prognosis, and treatment [21], thus providing sufficient accuracy and high confidence in the results. Big data is an essential requirement for DL models because it provides high volumes and variant samples, and such volumes and variant cases of medical imaging data are required to compose multiple datasets from different hospitals or clinical institutions around the world.

Practically speaking, COVID-19 medical imaging data have special characteristics that may provide misleading DL results. In one study, a large-scale deep learning model for COVID-19 medical imaging showed promising results, but it also revealed numerous biases [29]. In this section, we outline the unique characteristics of COVID-19 medical image data that should be considered before implementing the DL model. This is because these characteristics can lead to a lack of justification for the results, which the FL framework can potentially address. The following points describe these features and define the FL framework's applicability.

##### 4.1. Data Availability

The rapid spread of the virus in many countries, with no clear cause of infection and a wide range of symptoms in different infected populations [3,30], led to the compilation

of COVID-19 medical data from scattered data collections. This phenomenon may have precipitated the misclassification of annotated data, particularly during the initial phase of the pandemic [31]. Additionally, the testing kit model was to archive and label images solely for verified COVID-19 infections due to the substantial volume of patients combined with restricted storage capacity in some medical facilities [26]. This complicates the differentiation between the biomarkers of COVID-19 pathology and other ambiguous infections, such as pneumonia and SARS, at a localized level.

Therefore, training the deep learning model on a small but high-quality local hospital dataset is insufficient for generalization, as it contains unmeasurable instances of bias due to the model's incomplete understanding of co-features [10]. In practice, local models encounter overfitting challenges and present elevated accuracy with decreased reliability. On the other hand, several studies have integrated more than one source medical imaging dataset for COVID-19 to train their own deep learning models, with the goal of improving the model's generalizability [31,32]. They compiled these datasets from the published literature or radiology lecture notes; however, their poor quality led to underperforming systems [22] and produced outcomes with significant bias. This is because deep learning models are unable to identify COVID-19 characteristics, basing their knowledge on the common image features in the same dataset [33].

Federated learning, also referred to as distributed learning, is a methodology that instructs a model founded on all co-features that differ from one demographic to another. The methodology aims to train a model by utilizing multiple image acquisitions, rare instances, and diverse attributes in an acceptable ratio to refine the overall model. This approach alleviates the bias of local models by utilizing a high-quality dataset without the need to exchange actual data.

#### 4.2. Cold-Start Problem

Over the centuries, as epidemiological diseases have plagued humanity, the observed diseases within a specific geographical locale do not initiate catastrophic events during the initial stages of viral propagation. However, as time progresses, these infections tend to manifest more uniformly with identified symptoms and modes of transmission, thereby highlighting the evolution of that disease. However, the COVID-19 infection took on a different situation when the World Health Organization (WHO) declared COVID-19 a global pandemic on 11 March 2020. The WHO made this announcement approximately 2.5 months after reporting the first cases in Wuhan, China, in December 2019 [34]. Moreover, the rates of confirmed cases and deaths from COVID-19 vary across different countries, possibly due to their distinct social, political, and healthcare circumstances during the epidemic's spread. For example, while the virus has had terrible effects on China, other countries, like North Korea, may see impacts more like those of a seasonal virus [35]. Across countries, there are different levels of access to epidemiological data, so the high availability of data in some countries and lower availability in others necessitates collaboration between them to benefit from analyzing these scattered data without actually sharing patient records. Identifying the problem on a larger scale can thus lead to faster diagnosis, treatment, and other healthcare services, and FL is the paradigm for accumulating knowledge about COVID-19 and sharing insights globally. This highlights how important it is for those countries to work together, as it would make it easier to obtain useful insights from the disparate datasets without having to directly share private patient records, which can raise ethical and privacy concerns. Addressing the challenges of infectious diseases on a broader scale can significantly expedite the processes of diagnosis, treatment, and overall provision of healthcare services, thereby improving public health outcomes and responses to health crises. Federated learning (FL) has emerged as the most practical paradigm for generalizable results regarding COVID-19 while enabling global collaboration between heterogeneous data sources, ultimately fostering a more collaborative approach to tackling this unprecedented public health challenge [36].

#### 4.3. Time and Cost of Processing

Commercial fields place a high value on patient information, which leads hospitals and medical institutions to process and analyze medical data locally to prevent unexpected leakage. Since hospitals and clinical institutions manage patient data under data governance rules, local processing incurs significant costs in terms of human and hardware resources [30].

Labeling represents an additional cost since population density is unequal over geographical areas, and larger towns have large, busy hospitals with higher rates of screening tests. The disparity between hospitals and medical institutions leads to variations in the radiology department's peak working hours. As a result, radiologists must spend more time and effort analyzing and labeling medical images produced by crowded hospitals. The rapid spread of COVID-19 infections means that numerous screening tests are thus still pending labeling, we can use these unlabeled data as a testing set, training a global model on distributed datasets to classify infections [37].

#### 4.4. Security and Privacy

In most countries, health data governance has strict rules to prevent the leakage of patient information. It is highly protected, and access is controlled by administrators of hospitals and clinical institutions. Commonly, privacy concerns are still the largest obstacle to sharing patient information, as regulated by the United States Health Insurance Portability and Accountability Act (HIPAA) [38] and the European General Data Protection Regulation (GDPR) [39] rules. They set rigorous conditions that prevent the sharing of patient information or the use of smart technologies to extract patient information, even under research justifications. Privacy-preserving methods have been shown to be insufficient, with multiple attacks on anonymized patient information, which are easily targeted by re-identification attacks (i.e., linkage attacks). Recent studies have also developed machine learning (ML) approaches for predicting an individual's face, age, gender, and name from chest images [40].

FL offers a local processing framework under the ownership of medical data control [10]. Based on the aforementioned characteristics of COVID-19 medical image data, collaborative research efforts are necessary to provide consistent and reliable facts about the virus. FL has shown highly promising results for managing these data efficiently [41], and it could be generalized as opportunities for other medical imaging data in similar diseases.

### 5. COVID-19 Medical Imaging Data

Since the pandemic began in late 2019, ongoing collective scientific efforts have aimed to build reliable and consistent information about COVID-19, creating many medical applications from textual [34], tabular [42], audio [43], and medical image datasets [22]. This section provides a brief description of the types of common imaging modalities, how AI systems identify the biomarkers in the lungs, and the selection criteria of available databases that DL uses for diagnosing, treating, extracting, and analyzing information about the infected lung. The research focuses on the lung medical image datasets for COVID-19 patients.

From a technical perspective, Junaid Shuja et al. [22] classified the imaging modalities used by learning models for COVID-19 applications into CT images and X-ray images, mentioning the low availability and quality of ultrasound datasets. Both X-ray images and CT images can capture COVID-19 symptoms in the lungs, but CT images use a multifocal view that more clearly distinguishes COVID-19 symptoms from other types of viral pneumonia [44]. This provides DL models with high applicability for diagnosing COVID-19 in the early stages, around days 2–4 of onset [6]. Different hospitals and clinical institutions store and label X-ray and CT images in various formats, leading to heterogeneity in shared medical image data. Hospitals and clinical institutions identified heterogeneity issues and took early steps to overcome this by developing a standard format for CT images, X-rays, and other medical image modalities. DICOM is a medical technical

tool that combines medical images with structured patient information reports. A study reported the successful collection, visualization, and diagnosis of COVID-19 from four hospitals remotely over the DICOM network [45].

Unfortunately, according to radiology experts, most open-source DICOM datasets collected from published works or educational sites do not follow DICOM standards [4] and reflect low-quality acquisition [46] with weak standardization. Data quality plays a vital role in the analysis of information about COVID-19 and patients, and to avoid the pitfalls of FL applications, the garbage in, garbage out theory must be considered. To be useful for technical applications, each collection of medical images must pass evaluation and validation checks performed by more than one radiology expert. In addition, a dataset must have annotated labels, segmentation, and reports based on AI application needs, such as the RSNA dataset [47] and BIMCV dataset [48]. The findings advise medical institutions and hospitals to collaborate with ML researchers to access high-quality datasets. Researchers note that the DICOM standard provides a step forward for FL because it eliminates the challenges of data heterogeneity in distributed environments [10]. One important factor is the variety of cases in the training sites with enough volume to provide more generalizable results by simply averaging weights, even when evaluating the model on a new external dataset [49]. Regrettably, studies that integrate multiple datasets from multiple devices still report the heterogeneity issue; Section 7.1 thus delves into a detailed discussion of data heterogeneity, a major issue in FL.

### 6. Federated Learning Overview

FL is a technology for implementing DL over distributed data in multiple distinct sites without the need to share data from hospitals or medical institutions. To compute a global model, nodes upload only the local weights, also known as learning parameters or theta values, to a central point and then download them again [50]. FL takes one of two common design architectures: central aggregation or peer-to-peer architecture (P2P). In centralized aggregation, a central server coordinates the participant nodes, computes the global model, and communicates with the nodes, resulting in higher performance and flexibility, as illustrated in Figure 2. P2P proposes more a reliable and secure architecture to prevent a central point of failure in aggregation. It is a fully distributed architecture in which each P2P node coordinates itself and communicates with other nodes to compute the global model locally [51], as shown in Figure 3.

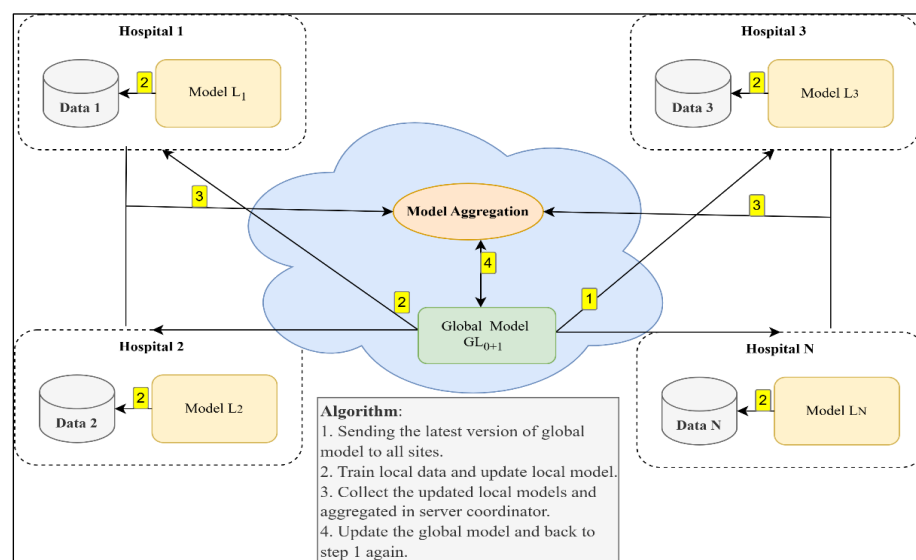
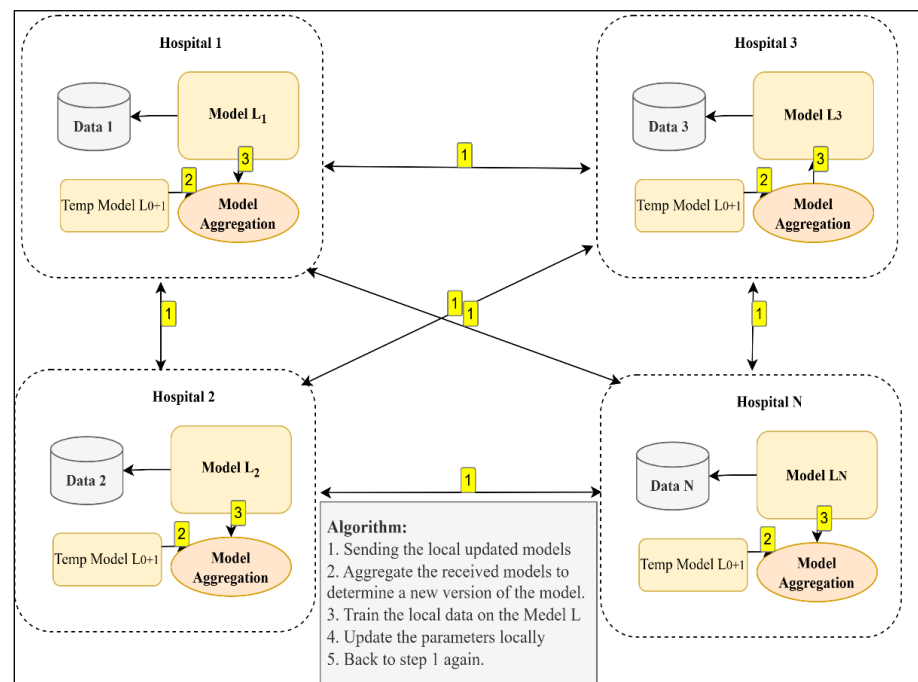


Figure 2. The algorithm of central FL architecture.





**Figure 3.** The algorithm of peer-to-peer architecture.

Depending on the federation’s goals, FL categorized into horizontal federated learning (HFL) and vertical federated learning (VFL) methods in distributed data contexts. If the objective of the federation is to expand the data sample to provide a greater volume and variety of training datasets, this necessitates the use of HFL. HFL is satisfied when the distributed data sample contains similar features, particularly in medical settings where different patient populations share the same disease, such as COVID-19 and cancer. On the other hand, if the goal is to increase the feature dimensions for identical records in distinct datasets, VFL is appropriate [52]. That is good enough when the distributed data have different sets of features for the same sample or records have the same ID, like when researchers need to work together between obstetrics and gynecology clinics and pulmonary clinics to investigate how COVID-19 affects pregnancy. Practically speaking, VFL is rarely used in the medical field because of the need for multiple datasets that must be identical, arranged, and standard. This is unrealistic due to the complex nature of medical image datasets and the lack of standardization management [17]. This study concentrates on the HFL, a commonly utilized format for medical images [53,54].

For the FL scenario in a medical image application, suppose the dataset  $D_h$  for a hospital or a medical institution where  $h$  is  $D(S_e)$ , where  $h = 0, 1, 2, \dots, N, N \in \mathbb{R}$ , and  $\mathbb{R}$  is real number.  $N$  is the number of participant nodes for a smart medical application (e.g., diagnosing, segmenting, or other application) for a disease  $S$  by teaching a model  $L_k(S)$ . Initially, the central server system offers a predefined learning model  $L_0(S_e)$  and sends to all or a subset of nodes in parallel. Each node updates the model weights or gradients value  $\theta$  based on its own image data if the number of training images in the local dataset is  $d$  and  $d = size(D(S_e))$ :

$$\theta_j := \theta_j + \alpha \frac{1}{d} \sum_{i=1}^d (L_0(S_e(x)) - \hat{y})x_j \tag{1}$$

The value of each  $\theta_j$  is updated with the learning rate  $\alpha$  for each image  $j$  until acceptable results are retrieved by the node  $h$ . The next local model,  $L_{0+1}(S_e)$ , is then computed and returned to the central server to compute the global model at the end of round number  $T$  using Equation (2).

$$L_{g(j=T)}(S_e(x)) = \frac{1}{N} \sum_{i=0}^N L_i(S_e) \tag{2}$$

Equation (2) aggregates the global model by averaging each weight over the number of sites, a process known as the FedAvg method [50]. Each round follows the same process to compute a new version of the global model. Notably, two types of proposed aggregation strategies exist: sequential aggregation and parallel aggregation. Sequential aggregation trains the distributed data site by site, updating the global model after each site completes its local training and moves on to the next. You can either repeat the training iteration until the model converges, like in cycling weight transfer (CWT), or train it at a site for a predetermined duration or number of epochs, like in single weight transfer (SWT). It is efficient to perform these steps in a fully distributed FL framework. The second aggregation strategy, known as a parallel aggregation, shares the same initial model with the distributed sites, enabling them to train data simultaneously. It then collects the local updated models from all sites before aggregating a global model. In HFL, averaging the weights in the global model is a vanilla method because it supports the convergence of updated weights. There are variant settings in the computations of local sides to preserve the personality of each hospital, such as when computed with the proxy local model in FedProx [55].

The following points describe the practical differences between FL for medical image applications and the standard processes of other fields, such as IOT devices, and explain the reasons behind each.

- Medical imaging data management is costly. Many medical institutions lack the infrastructure necessary to manage their imaging data according to standard management requirements. This is an emerging challenge in implementing federated learning for research: the limitation of the number of data sources that can be selected for training data.
- Only medium-to-large hospitals or medical research institutions own the repositories of standardized medical images. This enables deep learning to concentrate on valuable features, avoid incorporating weak features from low-quality data, and identify trustworthy participants. As a result, the local update models received from distributed sites might be more reliable [52], which leads to improving the global model's convergence in a lower number of rounds to achieve satisfactory accuracy.
- Datasets of medical images contain highly sensitive patient information. However, if the application of FL ignores privacy-preserving methods such as differential privacy, then the homomorphic encryption of the sharing weights may result in the leakage of patient or institutional privacy. At the same time, it greatly increases the computational overheads of training models because the medical image models exceed 10 million weights [56].

Considering these scenarios, the practical implementation of FL provides promising results for medical image analysis. Brain segmentation has demonstrated FL's applicability for MRI images [57], as well as biomarkers detected at four fMRI sites [58]. In previous projects, research has recognized the value of secure collaboration, and FL has advantages for fighting COVID-19. Several studies have shown that FL can provide accurate information about COVID-19 in a range of situations, such as automatically diagnosing COVID-19 in chest images [54,59–62], identifying the amount and location of lung damage [63], and estimating oxygen needs [64]. FL in medical imaging provides a feasible collaborative technique to analyze sensitive data securely with limited and trusted participants, a lower number of rounds, and synchronized communication with high bandwidth to achieve reliable results. However, when designing an FL solution, there are challenges that arise from a distributed framework that have to be considered by the model designer.

## 7. Data Heterogeneity Issue in Medical Imaging

FL allows each site to train its local model on its own data and optimize it according to the features that arise internally. The model weights are adjusted to align with the local results. This implies that we update each local model to align with its unique features. In the aggregation phase, the global model tries to balance the received weights by combining these local updates and averages. This makes the optimization very unstable and leads to a

high degree of divergence. Moreover, it produces an optimization drift between the local and global models, which leads the global model to bias [65]. Additionally, the distributed sites assign the global model to train local data for the next round, leading to a deficiency in personalization metrics and degradation of the local performance model.

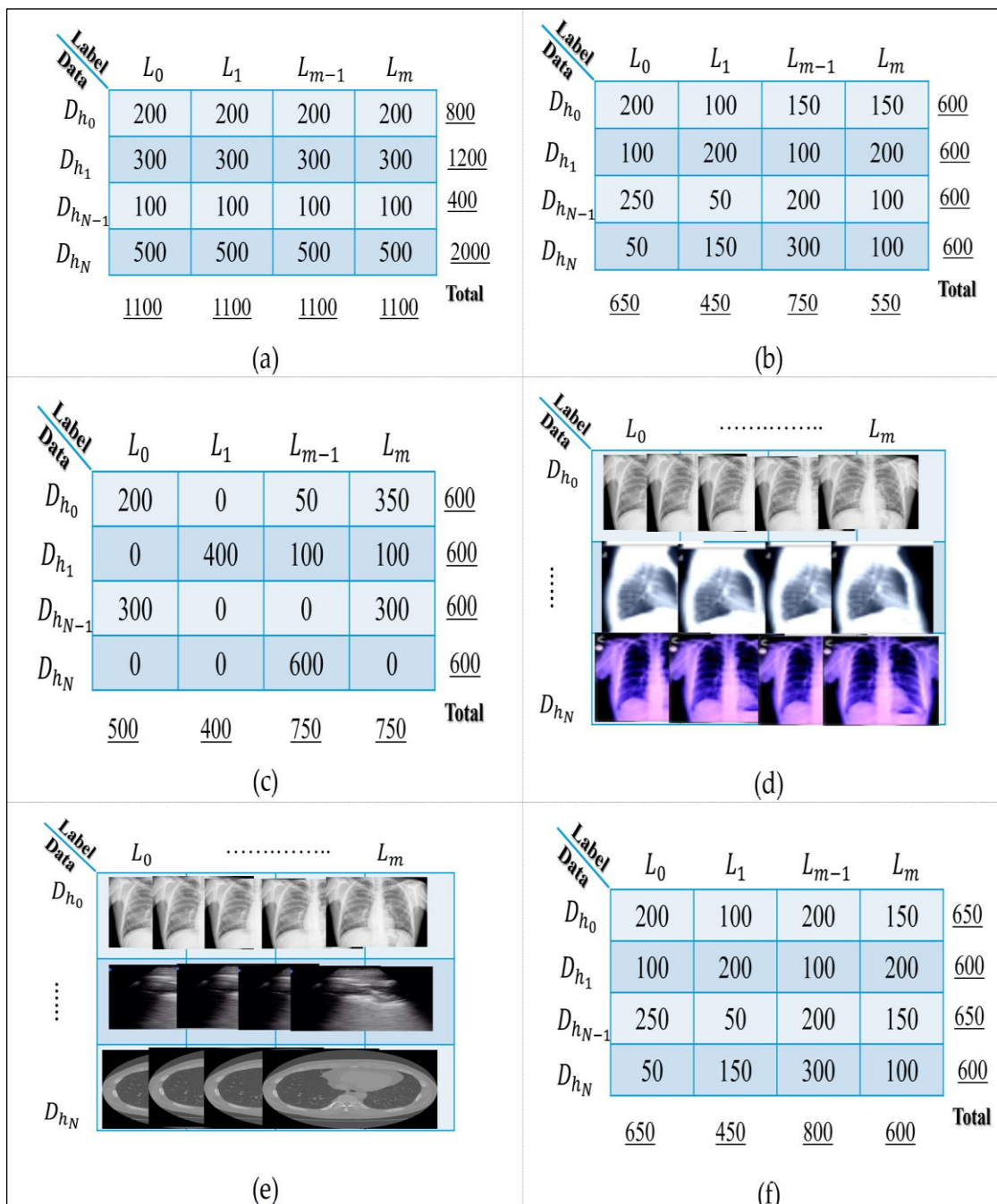
Bias, a commonly reported issue in FL, negatively impacts both the generalizability of the global model and the personality of local models. The generalization measure refers to the global model's ability to make more comprehensive decisions about out-sampling or external training data, which may contain new features and come from a variety of sources. Each round's aggregation step progresses to collect additional weights from local updates. Therefore, suggested solutions to address the generalizability metric include preventing participant poisoning, measuring local contributions, and sharing metadata and computational resources. Conversely, federated learning personalization refers to the global model's ability to effectively validate local data after each round. The global model performance for the site that has the lowest size of data achieved the lowest accuracy due to the averaging process of weights biased to the sites having a larger training sample [66]. Further computations are suggested to adapt each local model depending on the specifications corresponding to the participant sites (e.g., PPML [67] and FedAMP [49]). However, this measure required more investigation to specify effective results and findings about variant non-IID situations.

Specifically, controlling a larger divergence between these two measures becomes more challenging in FL when dealing with more heterogeneous data. The large variation between distributed datasets across FL participants is called the non-identical independent distribution (non-IID) data issue [63]. By investigating related works that have considered the non-IID data, we initially mentioned the variant process of partitioning experimented data before viewing their proposed solutions to mitigate this issue. By default, the type of skewness or shift in the data determines the partitioning process of data across FL participants. In certain proof-of-concept experiments, the partitioning process may accurately reflect the real-world data heterogeneity or sharing model used by distributed hospitals to evaluate their methods in clinical practice [59]. This section therefore aims to identify the situations that had been investigated for a distribution skew or shift in COVID-19 lung imaging data and estimate the negative impact of each type depending on the selected studies that have compared the FL performance in IID situations and in any one of the non-IID types. Depending on the reviewed studies, we categorized the skewness of datasets into the following six distinct types of data and provided examples for each in Figure 4. Furthermore, Figure 5 shows the number of investigations per skewness and their impact on the performance of FL frameworks.

### 7.1. Non-IID Types

#### 7.1.1. Quantity Skew

Quantity skew occurs when larger hospitals have larger datasets based on the number of patients, and the other participants have small datasets with a large difference between the size of each. The different number of medical images affects the model, with an extreme increase or decrease in data quantity [52]. The aggregation strategies may provide a larger data size and a higher ratio of changed weights (e.g., FedAvg). However, these changes may come from a limited variance of data or may only be cached based on trivial features that are unrelated to COVID-19 but only depend on the protocol of acquisition data in each local site. Sites with a small data quantity may lead to data ignorance due to the averaging of the model weights over sampling data per training site. As a result, the existing emphasis on COVID-19 features would fade with smaller datasets and bias the global model toward a higher quantity of data.



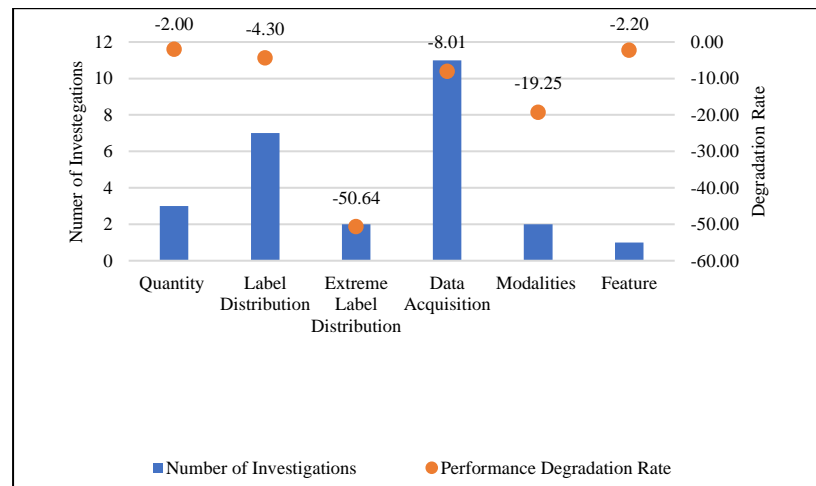
**Figure 4.** Skewness type examples including (a) quantity skew example, (b) label distribution skew example, (c) extreme label skew example, (d) acquisition protocol skew example, (e) modality skew, and (f) feature skew example.

By assuming N hospitals have the same image modality (i.e., X-ray, CT, and ultrasound) regardless of the captured equipment version, there is a dataset  $D_i$  of hospital i that does not have a size equal to that of the other dataset, where all the datasets have the same set of labels  $L(D)$ , and the sum of the lengths of each label k in all datasets is identical for all labels m. An example is shown in Figure 4a, where each cell represents the number of images for label  $L_m$  that are included in the corresponding dataset  $D_j$ . The mathematical expression of quantity skew is given via conditional Equation (3):

$$\exists i, j \in \{1, 2, \dots, N\} : |D_i| \neq |D_j| \wedge L(D_i) \equiv L(D_j) \wedge \sum_i^N |L_k(D_i)| \equiv \text{constant for all } k \in \{1, 2, \dots, m\} \quad (3)$$

A proposal suggests limiting the accepted ratio of quantity skew between two sites to 1:100 [52]. Experimentally, the FL performance is not significantly affected by quantity skew, as shown in Figure 5; it only degrades FL performance by approximately < 2% [68]. The researchers proposed the following methods to alleviate quantity skew across distributed data:

- Using the augmentation method to expand the size of the image dataset is a simple and common solution for highly training the model on the same data features, which can be achieved by changing various scales such as transformation, zooming, and rotation. However, the transformation methods used for generating data are not always effective in training, which may degrade the model’s performance [69].
- Alternatively, a generative adversarial network (GAN) [59] can be utilized [60]. It offers small improvements with a high computational time.
- In such aggregation strategies, quantity skew issues are improved by assigning a learning rate or batch number to each client variant based on the quantity of data [70].
- The FedAMP model exhibits resistance to quantity skew because its aggregate weights are adaptively learned throughout the training process [49].



**Figure 5.** The number of investigations of skewness types and the impact of each on the FL performance (collected from considered papers, as referred to in each skewness-type section).

### 7.1.2. Label Distribution Skew

Label distribution skew could be a combination of two types of skewness: quantity and label. Label distribution skews occur when datasets have the same labels but different quantities for each site. By assuming N hospitals have the same image modality (i.e., X-ray, CT, or ultrasound) regardless of the captured equipment version, there is a dataset  $D_i$  of hospital  $i$  with a size equal to that of dataset  $D_j$ , where all the datasets have the same set of labels  $L(D)$ , and the summation of the lengths of each label  $k$  over all datasets is variable for other labels in the set. An example is shown in Figure 4b, where each cell represents the number of images for label  $L_m$  that are included in the corresponding dataset  $D_i$ . The mathematical expression of label distribution skew is given via conditional Equation (4):

$$\exists i, j \in \{1, 2, \dots, N\} : |D_i| \equiv |D_j| \wedge L(D_i) \equiv L(D_j) \wedge \sum_i^N |L_k(D_i)| \neq \sum_i^N |L_{k+1}(D_i)| \text{ for all } k \in \{1, 2, \dots, m\} \quad (4)$$

This skew has a negative impact on FL generalizability, personality, and the computational time [60]. FedAvg [50] is proposed to address this type of skewness, but it fails when the distribution skewness ratio is high. It is the second skewness investigated, as shown in Figure 5, and it reduced accuracy approximately by 4.30% within a low ratio of label distribution skew [68].

To mitigate this issue, researchers have followed one of three directions.

- A first-direction solution is implemented before training data by preprocessing data to ensure a uniform distribution across sites using the local augmentation method [51], GANs [52], and the synthetic minority oversampling technique (SMOTE) [71]. However, these solutions require more communication between parties to fine-tune the number of labels and the distribution of images in each. This may also lead to information leakage from participant data with slightly improvements in accuracy, approximately around 1.6% [68].
- The second direction is to improve convergence between updated models locally, which focuses on the monitoring of local updates per batch in FedBN [72], per round in FedProx [55], or by normalizing both local and global updates in HarmoFL [65]. These methods report efficient bias mitigation and improve the global model’s generalizability. However, they may have a negative impact on the model’s personality. In other words, the model may yield poor results due to its incompatibility with local population data.
- The third direction is examining the local updates on the server side before accepting the updated models. This may rely on various calculations of acceptance priority [30], the use of voting methods [7], and the implementation of smart contracts in blockchain-based systems [50,73]. However, these methods have additional computational overheads.

### 7.1.3. Extreme Label Skew

This type of extreme skewness occurs when one or more labels completely disappear from one or more sites; this phenomenon is known as extreme label distribution skew. This situation is common because, during the COVID-19 pandemic, some hospitals stored only confirmed infection cases due to storage capacity limitations [26]. Additionally, certain medical image datasets categorize their data into two [74], three [54], or more [13] labels. By assuming N hospitals have the same image modality (i.e., X-ray, CT, or ultrasound) regardless of the captured equipment version, there exists a dataset  $D_i$  of hospital  $i$  with a different size for dataset  $D_j$ ,  $L(D)$  is the set of labels in all the datasets, and the summation of the lengths of each label  $k$  over all datasets is extremely variable for other labels in the set. An example is shown in Figure 4c, where each cell represents the number of images for label  $L_m$  that are included in the corresponding dataset  $D_i$ . The mathematical expression of extreme label skew is given via conditional Equation (5):

$$\exists i, j \in \{1, 2, \dots, N\} : |D_i| \neq |D_j| \wedge L(D_i) \neq L(D_j) \wedge \sum_i^N |L_k(D_i)| \neq \sum_i^N |L_{k+1}(D_i)| \text{ for all } k \in \{1, 2, \dots, m\} \quad (5)$$

The existence of a unique data label in one site and its absence in another are considered extreme non-IID cases, which results in high divergence between the global aggregation model and the local model performance. Even with more rounds, the global model may fail to preserve that label’s features when averaging over larger data with variant labels. This leads to poor convergence and degrades global accuracy by 50.64% in the central testing node of the FL framework for COVID-19 X-ray images [75]. Figure 5 shows that the experiments that have participants with one data label reported the lowest generalizability accuracy compared to other types of skew. However, the global averaging model can enhance the local model’s performance for participants trained on a single specific label or class. Because the variety of their own data is very limited, overfitting during the training process leads to high bias in the DL model. This type of skewness necessitates further investigation into the behavior of updated models at each site, both before and after aggregate models.

The following are valuable solutions proposed by different studies for this type of skew:

- The authors used a semi-supervised method to label unlabeled data and reported satisfactory accuracy [63]. As a recommendation to the uniform label name in the FL framework, their method could be useful in this situation.
- To address the word variants issue, radiologists could also analyze meta-data using natural language processing (NLP) [37].

#### 7.1.4. Data Acquisition Protocol Skew

Under this skew type, the same modality can be considered, but with image datasets with different attributes, such as X-ray modality but with different scanning devices, resolutions, light conditions, protocol for capturing (e.g., patient with upper/lower hands posed), and protocol for storing (e.g., with/without compression and with/without radiologist annotations).

By assuming N hospitals have the same image modality (i.e., X-ray, CT, or ultrasound) regarding the captured equipment version and the protocol of collecting the local data, where  $P = \{\text{volume, resolution, format, equipment, patient pose, etc.}\}$ , there exists a dataset  $D_i$  of hospital  $i$  with a size almost equal to that of dataset  $D_j$ .  $L(D)$  is an identical set of labels in each the datasets, and the summation of the lengths of each label  $k$  over all datasets approximately equal for other labels in the set. An example is shown in Figure 4d, where each cell represents the number of images for label  $L_m$  included in the corresponding protocol  $P_i$  of  $D_i$ . The mathematical expression of the data acquisition protocol skew is given via conditional Equation (6):

$$\exists i, j \in \{1, 2, \dots, N\} : |D_i| \cong |D_j| \wedge P(D_i) \neq P(D_j) \wedge L(D_i) \equiv L(D_j) \wedge \sum_i^N |L_k(D_i)| \cong \sum_i^N |L_{k+1}(D_i)| \text{ for all } k \in \{1, 2, \dots, L_m\} \quad (6)$$

As Figure 5 shows, this type is the most commonly considered skew. Many studies have applied the same modality of acquisition in several hospitals, and it has a moderately decreased impact on FL performance (around 8.01%) [74]. A previous study investigated X-Ray datasets with a different image volume thickness in each site of FL. The authors reported results with satisfactory performance, which could be collaborated to identify COVID-19 features, even with a variety of image attributes [67].

The researchers followed one of the following directions to fix acquisition skew:

- A self-adaptive hyperparameter is proposed to adopt the variability of data between distributed sites [70].
- In one study, the authors used a self-adaptive aggregation function with meta-transfer modules to manage the settings of training data locally [74].

In two further studies, the authors made distributed datasets uniform by incorporating regularization methods and leveraging ensemble techniques such as normalize image intensity, resizing, and geometric transforming methods [59,64].

The most common preprocessing methods are indicated in Table 1 under the skew type and FL architecture.

#### 7.1.5. Modality Skew

Medical image equipment varies between single sites and different sites. Different devices produce different imaging modalities (e.g., X-ray, CT, and ultrasound), with variations in volume, resolution, color intensity, type, and amount of noise. Different image modalities require different models and analysis procedures, which are difficult to unify [63]. It is a rare situation in an FL framework (as Figure 6 shows), where most studies considered single-modality X-ray data [53,68,74,76] or CT data [13,30,63,77,78], some studies combined both and then redistributed randomly [79] or assumed variant modalities in clustered edge nodes [80], and one study employed emerging X-ray imaging with simple vital signs [64].

By assuming N hospitals have different image modalities ( $M = \{\text{X-ray, CT, ultrasound, etc.}\}$ ), there exists a dataset  $D_i$  of hospital  $i$  with a size almost equal to that of dataset  $D_j$ ,  $L(D)$  is an identical set of labels in each the datasets, and the summation of the lengths of each label  $k$  over all datasets is approximately equal for other labels in the set. An example

is shown in Figure 4e, where each cell represents example images of modality for label  $L_m$  that are included in the corresponding M for data  $D_j$ . The mathematical expression of modality skew is given via conditional Equation (7):

$$\exists i, j \in \{1, 2, \dots, N\} : |D_i| \cong |D_j| \wedge M(D_i) \neq M(D_j) \wedge L(D_i) \equiv L(D_j) \wedge \sum_i^N |L_k(D_i)| \cong \sum_i^N |L_{k+1}(D_i)| \text{ for all } k \in \{1, 2, \dots, L_m\} \quad (7)$$

In Figure 5, this skew is the second highest degradation rate in FL, reducing the accuracy approximately by 19.25, as per the investigated study [80]. Practically, Qayyum et al. [80] conducted an experiment using the same model for X-ray, CT, and ultrasound images for COVID-19 diagnosis and fixing the impact of modality skew by clustering sites of FL that have a similar modality. They described the model’s ability to recognize COVID-19 features without prior knowledge of the image modality. However, this type of skewness is challenging the performance of the model that applied for training, which comparing features per pixel. Researchers thus need to further investigate this skew type and define the impact of varying image qualities on FL accuracy [69].

One feasible approach worth further exploration involves implementing an additional layer inside a neural network. This layer would be capable of determining the most suitable FL for deployment based on medical imaging and clinical similarities between variant dataset features [49].

### 7.1.6. Feature Skew

The feature skew of datasets relates each participant’s training sample to distinct features like age range, patient gender, smoking patients, and dead patients. The hospital or medical institution may acquire data for an intended feature, or it may depend on the virus’s behavior in their population, such as the ability of COVID-19 to infect aging patients more rapidly in some regions than others.

By assuming N hospitals have the same image modality (i.e., X-ray, CT, and ultrasound) regardless of the captured equipment version, there exists a dataset  $D_i$  of hospital i with a size almost equal to that of the other dataset, where all the datasets have the same set of labels  $L(D)$ , and the sum of the lengths of each label k in all datasets is identical for all labels m. An example is shown in Figure 4f, where each cell represents the number of images for label  $L_m$  in the corresponding feature F for  $D_i$  (where  $F = \{\text{age, gender, country, etc.}\}$ ). The mathematical expression of feature skew is given via conditional Equation (8):

$$\exists i, j \in \{1, 2, \dots, N\} : |D_i| \equiv |D_j| \wedge F(D_i) \neq F(D_j) \wedge L(D_i) \equiv L(D_j) \wedge \sum_i^N |L_k(D_i)| \equiv \sum_i^N |L_{k+1}(D_i)| \text{ for all } k \in \{1, 2, \dots, L_m\} \quad (8)$$

This type of feature skew is applicable to horizontal FL, where the non-overlapping sample at each site differs from the one defined in vertical FL. However, it is a real situation and rare in vertical experiments studies. A unique study investigated the FL performance by distributing the dataset into four sites, each with a range of ages. They reported a satisfactory degradation performance in the central node, ranging from 82% to 80% [81], as shown in Figure 5.

As mentioned above, a real medical data situation may consist of one or many skewness types in the training or testing data in the FL framework. It is crucial to examine the data specifications for each participant, as this information can be quantified and shared as meta-data prior to the start of training. These meta-data were used to identify the degree of non-IID per skewness, which provides the manager with the ability to assess the source of bias and help them to take advantage of FL, depending on their goal.

### 7.2. Bias Generation Factors

Many factors vary in the FL framework, which may be influenced by researchers’ considerations rather than the nature of distributed data. The model architecture that exchanges between participant sites and updates based on the training of each local dataset is a vital factor with an essential impact on the FL framework. Additionally, the strategy used to accept and aggregate these updated models and then compute the global model for

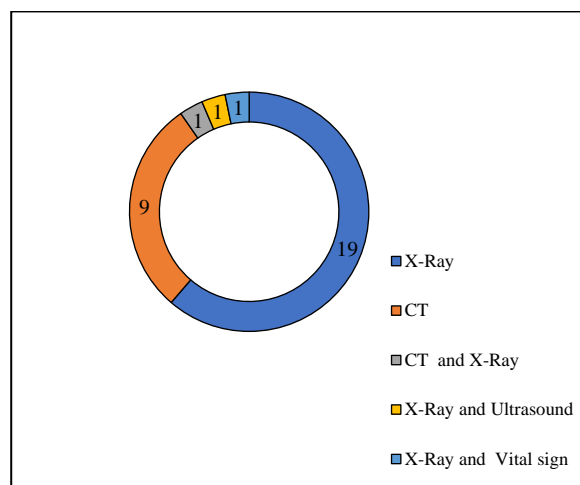


the next round is a crucial factor that could either be inherent in the bias of non-IID data or overcome it, as briefly described below.

### 7.2.1. Training Model

Identifying a model for training medical image data for all applications is difficult. However, researchers can conduct tests to identify an appropriate model that aligns with the system’s size, available bandwidth, and resource capabilities. It is possible to train large amounts of data effectively with pre-trained models based on convolutional neural networks. These models can also solve the overfitting problem and are very flexible and scalable in distributed settings [82]. Additionally, training the federated model from scratch yields a lower generalizable performance compared to transfer learning models, which have been shown to deliver higher improvements [59].

As there are many studies in the literature comparing how well common models worked [30,53,68,69,76,83], we found the average accuracy of the most common models reported in those studies based on the type of skewness studied. Researchers commonly use pretrained-CNN models (as mentioned in Figure 7) to train COVID-19 image data within the FL framework. Meanwhile, the residual network (ResNet) has gained widespread adoption in various medical imaging applications, such as diagnosis [54,59–62], segment and quantity [63], and oxygen needs prediction [64]. It is used by researchers to investigate feature skewness, label distribution, and acquisition, and may relate to the bottleneck architecture of residual blocks that augment network performance. In addition, it protects against vanishing gradient issues during training by utilizing identity connection.



**Figure 6.** Type of lung imaging dataset modalities used in FL framework for COVID-19.

The second most common model used in FL for COVID-19 imaging is the DenseNet model, which investigates acquisition, quantity, and feature skewness types. The VGG model is the third most popular used in the investigation of label, acquisition, quantity, and modality skewness experiments. ResNet50 and DenseNet, on the other hand, have higher and similar performance models and more stable convergence in comparison to the state-of-the-art ML models used for medical image classification in FL environments with an equal data distribution across sites [82]. Based on our literature review analysis, we identified other pre-trained CNN or simple CNN models that could fit into the FL architecture for the IID case or acquisition skew. Furthermore, the literature review revealed that the CNN’s architecture, which includes batch normalization layers and dropout parameters, negatively impacts the model’s convergence [56]. The literature review demonstrated that these elements could potentially deteriorate the overall FL performance [64].

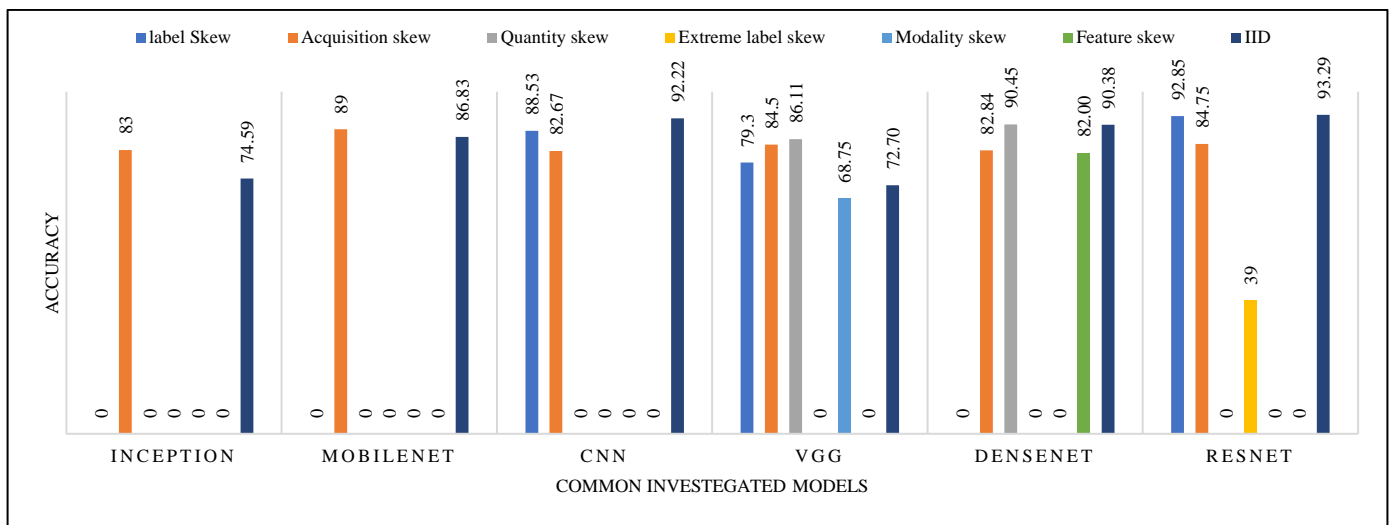


Figure 7. The average accuracy of the models that correspond to the skewness type.

### 7.2.2. Aggregative Strategy

Aggregation strategies are methods used for collecting updated local models from participants and computing the global model. Selecting an adopted aggregation strategy is a vital phase that could include the blockade bias or breadth for the distributed sites. FedAvg [50] is a commonly used strategy for aggregation, as it effectively prevents bias from data heterogeneity and enhances the fairness contribution of global models from sites using their trained data. However, it can present challenges in certain data heterogeneity distribution scenarios. Some improvements have been proposed to address the bias of aggregated weights resulting from non-IID skewness, which was achieved by collecting meta-data from participants, including computational resources and the training time [30]. Based on this information, the hyperparameters for each local model were then adopted.

Darzi et al. [84] performed a comparative experiment on different aggregative strategies and found that more stable convergence strategies required greater computational costs and communication rounds. Therefore, defining the goal metric is essential for identifying the aggregate strategy. Researchers can choose the appropriate aggregation strategy for their situation by managing the goals based on the generalization and personalization metrics, the specified FL architecture, and the application aim. Table 1 provides a summary of the proposed aggregation strategy for non-IID solutions within the FL framework for COVID-19 lung imaging, considering the metric, FL architecture, and training applications. This could guide researchers in identifying the existing solutions under investigation and suggesting ways to enhance them. Further investigation is required to identify the dependent factors and independent variables that need tuning to mitigate non-IID issues. Additionally, efforts must be made to create a benchmark for COVID-19 medical image datasets for FL simulation, considering all the challenges inherent in real datasets. This would help researchers study the relationships between these factors and system entities and interpret the results more clearly.

**Table 1.** Summary of the researchers’ directions for managing non-IID skewness types using an aggregative strategy and preprocessing.

Application	FL Architecture	Measured Metrics	Skewness Type	Aggregative Strategy	Preprocessing
Classification of lung diseases	Central	Generalization	Data acquisition	FedAvg	Used CycleGAN method [69]
			Quantity/label distribution	FedAvg	SMOTE [71]
		Extreme label	FedAvg	GAN with augmentation method [60]	
	P2P	Personalization	Data acquisition	FedBN	Lung segmentation, image normalization, and data augmentation [49]
			Quantity/label distribution	FedAvg under smart contract	The size of training sample is computed based on the ratio of class in the test set [70]
		Generalization	FedAvg	Augmentation [68] Change of the setting of FL hyperparameters [75]	
Segmentation of lung infections	Central	Personalization	Data acquisition	FedAvg with local adoption epoch	Spatial normalization and scaling [67]
	P2P	Generalization		FedAvg with weights of computational cost	Spatial and signal normalization with segmentation [30]
Boundary box of lung lesions	Central	Generalization vs. personalization	Data acquisition	FedAvg	Normalization and data augmentation [59]
Labeling and annotating data	Central	Generalization	Quantity/label distribution	Self-adaptive aggregation method	CLAHE parameter on data and transferring meta data [74]
			Data acquisition	FedAvg	Augmentation [63]
Oxygen prediction	Central	Generalization and personalization	Data acquisition	FedAvg	Normalization and augmentation of distributed data [64]
	Central	Generalization	IID	FedAvg	Not mentioned [13]
Severity diagnosing	P2P	Generalization and personalization	Data acquisition	FedAvg with timer for generated ledger	Capsule network for segmentation and classification with blockchain technology [73]

### 8. Common FL Challenges

The FL framework, a distributed system that remotely analyzes sensitive data, emphasizes the consideration of several challenges to achieve a satisfactory result. Model hyperparameters, as well as varying software and hardware, are variables that could be difficult to tune efficiently without in-depth investigation and practical experimentation. This section briefly discusses the common issues of communication costs, privacy and security attacks, and system resources, followed by a critical analysis of the available proposed solutions.

#### 8.1. Communication Issues

The communication links between servers and participating hospitals were used to share the weights of DL models for local medical image training. In the COVID-19 situation, reliable/synchronized communication is required to ensure the ability of participants to measure each contribution to the global model computation. Therefore, improvements in communication might adjust the links based on three factors: the size of the model, the number of rounds, and the available bandwidth. Medical image training typically involves a large model with millions of weights, necessitating the upload and download of numerous megabytes. For example, ResNet models are the most successful for COVID-19 images, and 3D-ResNet101 is 85.21 MB in size [69]. Similarly, size presents a challenge to FL performance, especially with the encryption overheads of privacy-preserving methods. The fusion model successfully minimizes the classification model, yielding satisfactory outcomes [79]. However, further research is required to understand its influence on various

non-IID skewness. The integration of FL and knowledge distillation boosts the volume of computational data, leading to an improvement in communication efficiency and training throughput. When evaluated on two sets of medical image segmentation datasets, their results demonstrate data privacy, reduced communication costs, and improved accuracy when using TransUNet and ResUNet as teacher models [86].

The number of rounds is essential to consider, as mentioned in Section 5. The COVID-19 medical situation requires a lower number of rounds than FL implementation in another domain. However, many experimental studies [54,61,62,64,69,83,87] provided results after 15 rounds as a minimum for their evaluation, with some studies requiring up to 200 rounds.

Observing the ratio of improvements in global model accuracy with increasing numbers of rounds revealed that it became insignificant after a few rounds in non-IID situations. However, in the IID situation, the global model convergence significantly improved after a few rounds, starting from the first 5–10 training rounds. Chowdhury et al. [76] observed stable convergence after the third round. To determine the stopping point of systems with low-performance improvements, further experiments are required.

### 8.2. Privacy and Security Issues

FL frameworks provide partial privacy and security. Guaranteed privacy is required to avoid reverse engineering attacks against local model updates, and numerous attacks have been reported in the FL system, as illustrated in Table 2. It is possible to implement privacy-preserving methods over three entities in FL systems, namely, images, local updates and global models, and communication channels, which can prevent any re-identification and reconstructor attacks. Reducing image quality by adding an amount of differential privacy (DP) noise significantly degrades model performance [56]. Finding methods to preserve patient privacy requires more effort.

**Table 2.** Types of security and privacy attacks.

Attack Name	Description of Impact	Methods
<b>Reconstructor attacks</b>	The image features are retrieved from the local updated weights.	DP/HE.
<b>Poisoning model</b>	a local model trained on fake labels or irrelevant datasets aimed at harming a global model is uploaded.	Measure the quality of local updates, which is still an open door in FL systems.

This method works for both local and global models. It adds noise to the model's weights when it collects models from sites or uses homomorphic encryption (HE) on the weights that are shared (step 3 of Figure 2 or step 1 of Figure 3). HE has a lower impact on model accuracy because it depends on the encryption on the local side and the decryption on the server side. However, it has higher computational overheads than DP [8]. The Secure Multi-Party Channel (SMPC) method encrypts communication channels between servers and participant nodes. Reports have indicated that it boasts a high number of communication channels [56], while blockchain-based systems are also used to maintain secure communication channels [73] and improve the accountability and accessibility of FL frameworks [62,88]. For further details about FL attacks in the medical image domain, Kessie et al. [10] provided a comprehensive review of secure and private methods for FL in medical imaging.

### 8.3. System Resource Issues

For hospitals and medical institutions, computational resources are usually limited. High computational resources, GPUs, bandwidth, and secure storage are essential for an efficient FL framework. The complexity of tuning hyperparameters in an FL setting and detecting errors in deployment and configuration over various resources may necessitate the involvement of numerous technical experts [49]. In FL, it is crucial to understand that

every single training round functions completely independent of the participating sites. The selecting of optimal local epoch, batch size, learning rate, training model, and number of rounds are factors significantly influence on the robustness performance [26]. Abdul Salam et al. [54] identified which factors affect model accuracy and loss, such as activation function, model optimizer, learning rate, number of rounds, and data size in IID settings. However, in the context of non-IID medical imaging, further investigation is required. We suggest classifying the configuration factors into high impact and low impact, with the aim of reducing negative performance.

## 9. Results and Discussion

FL provides promising solutions to mitigate privacy risks that hinder the full utility of medical data. It significantly enhances the applicability of various hospitals and medical institutions worldwide, enabling the construction of a global model. This provides the medical field with more generalizable results, emphasizing the relationships between the various COVID-19 variables. This is crucial for ensuring the generalizability of the achieved results across a larger population. Furthermore, it is crucial for the FL facility to maintain updated models with newly generated data from various sources, ensuring they align with current trends and strains, particularly given the evolving nature of COVID-19 [89].

This study discussed one of the main obstacles challenging FL's applicability, which is the lack of standardization of medical imaging data, and this rise in heterogeneous data across distributed hospitals and medical institutions is called non-IID. The non-IID issue arises when data distributions across different sites are heterogeneous, leading to challenges in model training and generalization. In this review, the types of skewness in data heterogeneity are described as the real situation mentioned in the radiologist field [46]. As described, the types and sources of data skewness can hinder the aggregated model's performance. This variety can lead to unfair model updates, where some hospital site data have a greater impact on the overall model than they should, making the model less useful across different patient populations and medical settings. Non-IID data can also worsen overfitting problems because models may become too specific to the features of a few medical resources, and not enough attention is paid to the wider range of variations in the smaller resources.

Table A1 shown the summarization of original papers that used FL framework to analysis and training COVID-19 lung medical imaging. Several research studies on the non-IID problem focused on comparing the performance of different models and the model's ability to achieve stable accuracy results over the training rounds. Other studies looked at how different hyperparameter settings affected the results of federated learning for different types of data skewness and tried to find a logical relationship between these factor settings in different situations of data heterogeneity. In addition, those research studies focused on examining the various types of distributed data skewness that are mentioned in Section 6 and providing solutions to mitigate their impacts. Specifically, the proposed solutions were diverse and followed one or more of three methods, which include expanding data, dynamic parameter adaptation, and the aggregation strategy.

The multiple reviewed experiments indicate that an increase in the homogeneity of the distributed data leads to a tendency toward the least effective type of skewness, known as the quantity skew. Conversely, as the data homogeneity decreases, there is a tendency toward extreme label skew and modality skew in the distributed medical imaging data. These have a significant negative impact on the performance of the federated learning model, as shown in Figure 5. This literature survey observed the label distribution skew as the most prevalent phenomenon in the distributed reality of medical imaging data, having attracted the attention of most researchers. It also revealed a paucity of studies on the three types of skewness, namely, feature, modality, and extreme label skews. This emphasizes the necessity of studying these types and determining the impact they have on both generalization and personalization metrics. This literature review demonstrates the pretrainable-based CNN model architecture, specifically the ResNet and DenseNet models

within the FL framework, which yield more consistent results. It interprets these models' appropriateness for training medical imaging in distributed environments.

This review discusses other common issues within the FL framework, such as privacy and security, communication, and system resources. To ensure FL's success in various medical applications beyond diagnosing and segmenting COVID-19 infection from lung imaging data, specialists in those fields must consider and further investigate these issues. Besides the technical issues mentioned in this review, clinical efforts are required to reflect the theoretical and simulation results in real-life situations. However, validation by independent medical researchers is crucial for predictive analytics [24]. Clinical recommendations should advocate publicly available and verified algorithms, and adequate and in-depth analysis of data complexity is necessary. For example, in survival cases, some prognostic results should be tracked for 30 days. Monitoring patients and analyzing their updates, whether they have recovered or died within that period, could be necessary as part of the risk management framework.

## 10. Recommendations and Directions

FL in medical imaging succeeded in creating a private environment to process these sensitive data. However, it suffers from inconsistent results because the training relies on insufficient medical imaging data for COVID-19. Many open-source datasets are published in technical communities such as Kaggle and GitHub [20]; unfortunately, these datasets have low quality and are not confirmed by validation or testing radiology methods to support measurement experiments.

Identifying the FL framework with effective hyperparameter settings for medical images needs further work. It is important to study the trade-off settings in the case of non-IID data, such as the number of local epochs against GPU consumption, the number of rounds against accuracy, the number of participants, and the convergence rate against the batch size. Furthermore, more experiments are needed to define the proper aggregation strategy, activation function, and generalizability of the global model without compromising the personality of the local model.

Moreover, researchers are required to provide further investigations about the solutions to data heterogeneity by studying the mentioned skewness issues. In addition, more efforts are required to reduce the communication cost and find a recovery strategy in case of lost connections or dropping one or more communication sites during the training round.

Improving FL frameworks and overcoming these challenges can provide promising results for valuable studies in the medical field. This requires researchers to broaden the benefits of federated learning for COVID-19 data beyond just identifying the disease or extracting the affected part of the lung to encompass a wide range of applications, especially studies identifying the relationship between COVID-19 and patient immunity, genetic impact, and chronic diseases. FL enables studies on larger samples to gain reliable results.

We highly recommend broadening the contact between researchers from both the medical and technical fields to facilitate studies about COVID-19 virus behaviors. The FL method could effectively process the variety of COVID-19 viral sequences, which aids medical studies in tracing the origins and transmission pathways of infections [3]. These data are crucial for improving global contact tracking, guiding epidemiological studies, and supporting broader public health campaigns to stop the virus's spread. The distributed samples that FL trains are also useful for studying and understanding how quickly COVID-19 sequences change. This has caused a lot of worry about the appearance of new variants that might be more virulent or more easily spread than the strains that are already out there. Additionally, it could provide an answer to a crucial question in the context of viral evolution: could specific mutations within the viral RNA sequence potentially undermine the effectiveness of existing vaccines? This could potentially aid in the development of new immune strategies to combat the virus.

### 11. Conclusions and Future Work

Federated learning utilizes local data to learn without revealing private information or granting access to the data. As described in this literature review, the FL paradigm provides a secure distributed environment for training sensitive medical data, with the aim of providing a useful context for unresolved issues that necessitate further research in data heterogeneity. We started by discussing how FL can address the global shortage of COVID-19 information. Next, we identified the available medical images for COVID-19 and discussed how DL could efficiently read COVID-19 symptoms from various chest image modalities, depending on the radiologist’s perspective. We discussed FL scenarios and the unique characteristics of the system for medical image training. Lastly, we summarized the recently suggested FL systems for the non-IID distribution of COVID-19 data variants. Furthermore, other common challenges were discussed in detail, and effective solutions were described under each issue.

Despite significant technical efforts to harness the benefits of AI, its application in combating the COVID-19 pandemic alone may be limited. Technical issues such as bias, data heterogeneity, compromised privacy and security, and a lack of resources have been reported, but these are not the only obstacles to achieving an efficient outcome in real-world scenarios. Technical professionals must collaborate to read, interpret, and offer comments and recommendations to enhance the design of the FL system. Tools must be developed to tune hyperparameters remotely according to local resources and assist radiology experts and doctors in reading and evaluating FL findings on COVID-19 data from a healthcare perspective. Collaboration between medical and technical fields is thus essential to gaining the advantages afforded by FL.

**Author Contributions:** Conceptualization, F.S.A. and A.A.B.; methodology, F.S.A.; software, F.S.A.; validation, A.A.B.; formal analysis, F.S.A.; investigation, F.S.A.; resources, F.S.A.; data curation, F.S.A.; writing—original draft preparation, F.S.A.; writing—review and editing, A.A.B.; visualization, A.A.B.; supervision, A.A.B.; project administration, A.A.B.; funding acquisition, not applicable. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data from the present study are available from the corresponding author for private use only.

**Conflicts of Interest:** The authors declare no conflicts of interest.

### Appendix A

**Table A1.** A summary of proposed FL framework for COVID-19 lung medical imaging.

Study	Aim	Application	Contributions	Limitations
Xu et al., 2020 [13]	Compares the accuracy of FL models with six radiologists in a diagnosing task. Fixes the lack of generalization for local models.	Diagnosing CT lung images with four infection labels: COVID-19, viral pneumonia, bacterial pneumonia, and healthy.	They achieved a comparable FL model in terms of sensitivity-specificity for classification results compared with six radiologists. They conducted real FL experiments with data from three hospitals in Wuhan.	There was a trade-off between performance and communication because 16 hours were required to finish 200 training rounds.

Table A1. Cont.

Study	Aim	Application	Contributions	Limitations
Zhang et al., 2021 [79]	Improves communication efficiency using dynamic fusion-based federated learning.	Diagnosing X-ray and CT lung images with three infection labels: COVID-19, pneumonia and healthy.	They were able to reduce communication overheads by scaling down the uploaded model to 1/16 of the time needed by galaxy FL in complicated models with satisfactory accuracy.	They did not consider reversing engineering abilities in their solution.
Feki et al., 2021 [68]	Investigates properties and specificities of FL settings, including non-IID and unbalanced data distribution.	Diagnosing chest X-ray images with two infection labels: COVID-19 and healthy.	They found the following: Increasing the number of rounds could improve the accuracy of models. More participants led to fast convergence rates and reduced the need for more rounds. - Labeled distribution skew led to worse performance than quantity skewness.	They reported the results on a small dataset containing only 108 chest X-ray images of positive COVID-19.
Liu et al., 2021 [53]	Compares the performance in FL of four DL models on COVID-19 X-ray images: COVIDNet, ResNeXt, MobileNet-v2, and ResNet18.	Diagnosing X-ray lung images with three infection labels: COVID-19, pneumonia, and healthy.	They found ResNeXt has the best performance in images with COVID-19 labels.	Models were trained on data containing only 2% COVID-19 labels, which may provide unreliable results without considering non-IID issues.
Jabłecki et al., 2021 [83]	Measures the impact of the non-IID issue on the accuracy of FL models.	Diagnosing chest X-ray images with three infection labels: COVID-19, pneumonia, and healthy.	They found the following: More local epochs increase GPU time without significant impact on accuracy. Non-IID degraded accuracy from 0.923 to 0.39. - EfficientNetB0 achieved the best performance.	The time needed for the first round was longest due to the construction of the execution graph in the TensorFlow framework at the beginning of training. However, they neglected the impact of low availability of GPU resources on Google Collab cloud.
Dou et al. n.d. 2021 [59]	Improves generalizability of and automated estimation of the lesion progression using data from 4 different hospitals in Germany and China in testing with comparison to radiologists' report.	Quantifying lesions from COVID-19 CT images.	They found the following: - Increasing data size is important to mitigate model bias and improve generalizability of diverse training data associated with imaging scanners and annotation protocols.	The time required was 40 ms per round to test one CT image, but they did not consider reverse engineering attacks.



Table A1. Cont.

Study	Aim	Application	Contributions	Limitations
Qayyum et al., 2022 [80]	Attempts to fix heterogeneity of imaging modalities and improves computational overheads by using edges to cluster each type of modality with different models for automatic diagnosis of COVID-19.	Diagnosing chest X-rays and ultrasound images with binary classification of COVID-19 and normal.	They found that the same result can be reported by sharing the same model with different modalities. The generalizability of the global model can be improved, even with limited hospital resources, and they could benefit from this collaborative learning method.	They did not mention how the data were distributed across clients and clusters in their experiments. Privacy was not guaranteed. They mentioned improving the low latency of FL as an aim of the study, but there were no results about it.
Yang et al., 2021 [63]	Evaluates FL performance with heterogeneity of data acquisition skew and unlabeled data by training on data from China, Italy, and Japan.	Segmenting and annotating lesions on lungs infected by COVID-19 using CT images.	They reported the importance of data augmentation strategies for computing consistency loss, which improves the generalizability of model. They described the need to tune the trade-off between aggregation frequency and communication cost based on the applications.	They did not solve the problem of how to improve models with non-IID issues and mitigate or detect bias during FL.
Bai et al., 2021 [69]	Aims to improve generalizability by collecting data from 5 hospitals and challenging the FL method with high heterogeneity of data.	Diagnosing chest X-ray images with three infection labels: non-COVID-19 viral and bacterial pneumonia, COVID-19, and healthy.	They provided the results of computational cost FLOPS with different models.	They mentioned the lack of bias in their study and dropping of participants during training rounds.
Kumar et al., 2021 [77]	Attempts to overcome the problem of a central point using a fully decentralized blockchain and HE.	Segmentation and classification to detect the COVID-19.	They introduced a new dataset containing 34,006 CT scan slices for 89 patients and 28,395 CT positive scans. The accuracy of the global model was $84.21 \pm 0.43$ .	They did not report the latency of the blockchain or minimize the cost of the solution.
Abdul et al., 2021 [54]	They studied the impact of the FL hyperparameters during testing on the accuracy and loss of the global model.	Diagnosing chest X-ray images with three infection labels: COVID-19, pneumonia, and healthy.	They found the following: - Softmax activation function and SGD optimizer gave the best prediction accuracy and loss.	They reported the limited impact of increasing data size and number of rounds. However, the results cannot be generalized because they are incompatible with other studies [50,73].

Table A1. Cont.

Study	Aim	Application	Contributions	Limitations
Zhang et al., 2021 [60]	Attempts to fix data availability and data privacy issues by using generative adversarial networks to generate fake chest X-ray images and DP to determine the gradient’s weights.	Diagnosing chest X-ray images with three infection labels: COVID-19, pneumonia, and healthy.	They demonstrated that the impact of generating fake images improves global accuracy by 0.84% and reduces loss by 3.0%. They achieved high performance with a low ratio of noise. They reported satisfactory results, even with non-IID.	They reported results of non-IID with label distribution skew only and did not consider other types of skewness.
Kumar et al., 2021 [30]	Proposes a normalization method for uniform data to fix heterogeneity of data using a blockchain-based method.	Segmentation and classification to detect COVID-19.	Their method achieved the highest sensitivity and lowest specificity. They reported the negative impact of communication costs when increasing the number of participants.	The configuration procedure was not explained clearly.
Dong et al., 2021 [74]	Attempts to annotate unlabeled data with a federated contrastive learning framework with two modules: metadata transfer module and self-adaptive aggregation module.	Labeling unlabeled data with two infection labels: COVID-19 and healthy.	They reduced annotation costs while utilizing only 3% of labeled data in training to achieve 90% accuracy. Their aggregation module outperformed the FedAvg method consistently, even with non-IID issues, while metadata transfer improved performance.	They did not apply any privacy-preserving method to guarantee privacy.
Dayan et al., 2021 [64]	Uses data from 20 distributed sites to predict outcomes at 24 and 72 h from time of initial presentation to the emergency room and predicts mechanical ventilation treatment or death at 24 h for symptomatic patients with COVID-19 using inputs of vital signs, laboratory data, and chest X-ray images.	Predicting future oxygen requirements.	FL provided comparable performance even when only 25% of weight updates were shared. Personalization could be improved by fine-tuning local parameters. Participant diversity improved generalizability by 38%.	They did not refer to the time/cost of computations.
Nguyen et al., 2021 [61]	Attempts to fix data availability and data privacy by using generative adversarial networks to generate fake chest X-ray images in edge cloud computing.	Diagnosing chest X-ray images with three infection labels: COVID-19, pneumonia, and healthy.	They improved the generalizability of models.	They did not apply any privacy-preserving methods to guarantee privacy.

Table A1. Cont.

Study	Aim	Application	Contributions	Limitations
Lo et al., 2021 [70]	Attempts to enhance the accountability and fairness of FL by using a blockchain-based smart contract system and a weighted fair data sample algorithm.	Diagnosing chest X-ray images with four infection labels: COVID-19, pneumonia, lung opacity, and healthy.	They found the following: More stable and faster convergence rate than ResNet50 models. Blockchain-based smart contracts provided satisfying performance with accountability. Weighted fair data improved performance in cases of distribution skew.	They did not apply any privacy-preserving methods to guarantee privacy.
Bhattacharya et al., 2022 [66]	Uses three different sources of data to maintain non-IID nature.	Diagnosing chest X-ray images with two infection labels: COVID-19 and healthy.	They found that personality was improved while each client's models performed well on the test data belonging to the same source. However, they found that generalizability could be improved by averaging the weight on a global model.	They did not apply any privacy-preserving method to avoid privacy attacks. They did not mention the configuration process or HW of the system.
Ho et al., 2022 [75]	Aims to improve the privacy and accuracy of COVID-19 detection models using an FL model with X-ray image and symptom data.	Diagnosing chest X-ray images with three infection labels: COVID-19, pneumonia, and healthy.	They found that SPP-CNN with 3X3 had higher accuracy because it extracts more spatial details. The accuracy was reduced with non-IID data from 14% to 24%. A larger batch size achieved faster convergence. The accuracy was only reduced by 0.17% with DP noise.	They did not fix the lack of data quantity using any preprocessing method. Their dataset contained only 3616 COVID-19 positives against 10,192 normal images.
Durga et al., 2022 [87]	Combines a model of capsule networks and extreme learning machines (ELMs) to improve the accuracy of segmentation and COVID-19 detection.	Segmentation and classification to detect COVID-19.	The ensemble of capsule networks and ELMs produced the best accuracy in detecting COVID-19 from multiple datasets and was superior to other algorithms.	In the first phase, each hospital uploads image datasets for collaborative learning. In the second phase, hospitals share the locally trained model weights with the blockchain and use FL to aggregate all local models into a global model. Uploading images to the BC involves high costs and threatens privacy.

Table A1. Cont.

Study	Aim	Application	Contributions	Limitations
Chowdhury et al., 2023 [76]	Proposes a web application to help users detect COVID-19 in a few seconds by uploading a single chest X-ray image.	Diagnosing chest X-ray images with three infection labels: COVID-19, pneumonia, and healthy.	They found that the Xception model outperforms other models.	They did not apply any privacy-preserving method to avoid privacy attacks. Also, they did not consider non-IID.
Kumar et al., 2022 [73]	Attempts to improve fully decentralized FL by using distributed blockchain ledgers that share weights with HE.	Diagnosing chest X-ray images with three infection labels: COVID-19, pneumonia, and healthy.	They proposed a method to ensure the quality of the model and the learned data. The dropping of any FL participant may affect the performance of the model due to divergence of weights in the local models from the global model. HE provided lower reduction in accuracy than DP.	They mentioned the limitation of latency caused the blockchain and encryption computations.
Wang et al., 2022 [62]	Attempts to fix the third-party dependence of FL on blockchain technology.	Diagnosing CT lung images with two infection labels: COVID-19 and healthy.	They found that the asynchronous method in the FL process achieved similar performance to using non-IID datasets. They reported results with different link capacities and found that increasing link capacity may decrease iteration delay time.	They reported difficulty in ensuring the quality of the local updated model because the operation was consistent for each local node. However, it was measured by Kumar et al. [62].
Kandati and Gadekallu [90] 2023	Aims to address the issue of communication cost using swarm optimization algorithm.	Diagnosing X-ray images into three labels: Normal, COVID, and Viral Pneumonia.	They found swarm optimization has effective results only with small datasets and lower number of participants.	Their algorithm took longer to convert global model and required huge search space.

## References

1. Coronavirus Disease (COVID-19). Available online: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019> (accessed on 25 May 2024).
2. WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus (COVID-19) Dashboard with Vaccination Data. Available online: <https://covid19.who.int/> (accessed on 3 May 2024).
3. Halawa, S.; Pullamsetti, S.S.; Bangham, C.R.M.; Stenmark, K.R.; Dorfmueller, P.; Frid, M.G.; Butrous, G.; Morrell, N.W.; de Jesus Perez, V.A.; Stuart, D.I.; et al. Potential Long-Term Effects of SARS-CoV-2 Infection on the Pulmonary Vasculature: A Global Perspective. *Nat. Rev. Cardiol.* **2022**, *19*, 314–331. [CrossRef] [PubMed]
4. Li, R.; Pei, S.; Chen, B.; Song, Y.; Zhang, T.; Yang, W.; Shaman, J. Substantial Undocumented Infection Facilitates the Rapid Dissemination of Novel Coronavirus (SARS-CoV-2). *Science (1979)* **2020**, *368*, 489–493. [CrossRef]
5. Williamson, E.J.; Walker, A.J.; Bhaskaran, K.; Bacon, S.; Bates, C.; Morton, C.E.; Curtis, H.J.; Mehrkar, A.; Evans, D.; Inglesby, P.; et al. Factors Associated with COVID-19-Related Death Using OpenSAFELY. *Nature* **2020**, *584*, 430–436. [CrossRef]
6. Aljondi, R.; Alghamdi, S. Diagnostic Value of Imaging Modalities for COVID-19: Scoping Review. *J. Med. Internet Res.* **2020**, *22*, e19673. [CrossRef]
7. Bahadur, T.; Verma, K.; Kumar, B.; Jain, D. Coronavirus Disease (COVID-19) Detection in Chest X-Ray Images Using Majority Voting Based Classifier Ensemble. *Expert Syst. Appl.* **2021**, *165*, 113909.

8. Sarma, K.V.; Harmon, S.; Sanford, T.; Roth, H.R.; Xu, Z.; Tetreault, J.; Xu, D.; Flores, M.G.; Raman, A.G.; Kulkarni, R.; et al. Federated Learning Improves Site Performance in Multicenter Deep Learning without Data Sharing. *J. Am. Med. Inform. Assoc.* **2021**, *28*, 1259–1264. [[CrossRef](#)]
9. Shen, M.; Deng, Y.; Zhu, L.; Du, X.; Guizani, N. Privacy-Preserving Image Retrieval for Medical IoT Systems: A Blockchain-Based Approach. *IEEE Netw.* **2019**, *33*, 27–33. [[CrossRef](#)]
10. Kaissis, G.A.; Makowski, M.R.; Rückert, D.; Braren, R.F. Secure, Privacy-Preserving and Federated Machine Learning in Medical Imaging. *Nat. Mach. Intell.* **2020**, *2*, 305–311. [[CrossRef](#)]
11. Li, L.; Qin, L.; Xu, Z.; Yin, Y.; Wang, X.; Kong, B.; Bai, J.; Lu, Y.; Fang, Z.; Song, Q.; et al. Using Artificial Intelligence to Detect COVID-19 and Community-Acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy. *Radiology* **2020**, *296*, E65–E71. [[CrossRef](#)]
12. Raisaro, J.L.; Marino, F.; Troncoso-Pastoriza, J.; Beau-Lejdstrom, R.; Bellazzi, R.; Murphy, R.; Bernstam, E.V.; Wang, H.; Bucalo, M.; Chen, Y.; et al. SCOR: A Secure International Informatics Infrastructure to Investigate COVID-19. *J. Am. Med. Inform. Assoc.* **2020**, *27*, 1721–1726. [[CrossRef](#)]
13. Xu, Y.; Ma, L.; Yang, F.; Chen, Y.Y.; Ma, K.; Yang, J.; Yang, X.; Chen, Y.Y.; Shu, C.; Fan, Z.; et al. A Collaborative Online AI Engine for CT-Based COVID-19 Diagnosis. *medRxiv* **2020**. [[CrossRef](#)]
14. Mbunge, E.; Akinnuwesi, B.; Fashoto, S.G.; Metfula, A.S.; Mashwama, P. A Critical Review of Emerging Technologies for Tackling COVID-19 Pandemic. *Hum. Behav. Emerg. Technol.* **2021**, *3*, 25–39. [[CrossRef](#)]
15. Thompson, P.M.; Stein, J.L.; Medland, S.E.; Hibar, D.P.; Vasquez, A.A.; Renteria, M.E.; Toro, R.; Jahanshad, N.; Schumann, G.; Franke, B.; et al. The ENIGMA Consortium: Large-Scale Collaborative Analyses of Neuroimaging and Genetic Data. *Brain Imaging Behav.* **2014**, *8*, 153–182. [[CrossRef](#)]
16. Rieke, N.; Hancox, J.; Li, W.; Milletari, F.; Roth, H.R.; Albarqouni, S.; Bakas, S.; Galtier, M.N.; Landman, B.A.; Maier-Hein, K.; et al. The Future of Digital Health with Federated Learning. *NPJ Digit. Med.* **2020**, *3*, 1–7. [[CrossRef](#)]
17. Darzidehkalani, E.; Ghasemi-rad, M.; van Ooijen, P.M.A. Federated Learning in Medical Imaging: Part II: Methods, Challenges, and Considerations. *J. Am. Coll. Radiol.* **2022**, *19*, 975–982. [[CrossRef](#)]
18. Darzidehkalani, E.; Ghasemi-rad, M.; van Ooijen, P.M.A. Federated Learning in Medical Imaging: Part I: Toward Multicentral Health Care Ecosystems. *J. Am. Coll. Radiol.* **2022**, *19*, 969–974. [[CrossRef](#)]
19. Xu, J.; Glicksberg, B.S.; Su, C.; Walker, P.; Bian, J.; Wang, F. Federated Learning for Healthcare Informatics. *J. Healthc. Inform. Res.* **2021**, *5*, 1–19. [[CrossRef](#)]
20. Yoo, J.H.; Jeong, H.; Lee, J.; Chung, T.M. Federated Learning: Issues in Medical Application. In *Future Data and Security Engineering, Proceedings of the 8th International Conference, FDSE 2021, Virtual Event, 24–26 November 2021*; Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Berlin/Heidelberg, Germany, 2021; Volume 13076 LNCS, pp. 3–22. [[CrossRef](#)]
21. Peiffer-Smadja, N.; Maatoug, R.; Lescure, F.-X.; D’Ortenzio, E.; Pineau, J.; King, J.-R. Machine Learning for COVID-19 Needs Global Collaboration and Data-Sharing. *Nat. Mach. Intell.* **2020**, *2*, 293–294. [[CrossRef](#)]
22. Shuja, J.; Alanazi, E.; Alasmary, W.; Alashaikh, A. COVID-19 Open Source Data Sets: A Comprehensive Survey. *Appl. Intell.* **2021**, *51*, 1296–1325. [[CrossRef](#)]
23. Mondal, M.R.H.; Bharati, S.; Podder, P.; Kamruzzaman, J. Deep Learning and Federated Learning for Screening COVID-19: A Review. *BioMedInformatics* **2023**, *3*, 691–713. [[CrossRef](#)]
24. Hwang, S.O.; Majeed, A. Analysis of Federated Learning Paradigm in Medical Domain: Taking COVID-19 as an Application Use Case. *Appl. Sci.* **2024**, *14*, 4100. [[CrossRef](#)]
25. Hernandez-cruz, N.; Saha, P.; Sarker, M.K.; Noble, J.A. Review of Federated Learning and Machine Learning-Based Methods for Medical Image Analysis. *Big Data Cogn. Comput.* **2024**, *8*, 99. [[CrossRef](#)]
26. Naz, S.; Phan, K.T.; Chen, Y.P.P. A Comprehensive Review of Federated Learning for COVID-19 Detection. *Int. J. Intell. Syst.* **2022**, *37*, 2371–2392. [[CrossRef](#)]
27. Banabilah, S.; Aloqaily, M.; Alsayed, E.; Malik, N.; Jararweh, Y. Federated Learning Review: Fundamentals, Enabling Technologies, and Future Applications. *Inf. Process. Manag.* **2022**, *59*, 103061. [[CrossRef](#)]
28. Lim, W.Y.B.; Luong, N.C.; Hoang, D.T.; Jiao, Y.; Liang, Y.C.; Yang, Q.; Niyato, D.; Miao, C. Federated Learning in Mobile Edge Networks: A Comprehensive Survey. *IEEE Commun. Surv. Tutor.* **2020**, *22*, 2031–2063. [[CrossRef](#)]
29. Roberts, M.; Driggs, D.; Thorpe, M.; Gilbey, J.; Yeung, M.; Ursprung, S.; Aviles-Rivero, A.I.; Etmann, C.; McCague, C.; Beer, L.; et al. Common Pitfalls and Recommendations for Using Machine Learning to Detect and Prognosticate for COVID-19 Using Chest Radiographs and CT Scans. *Nat. Mach. Intell.* **2021**, *3*, 199–217. [[CrossRef](#)]
30. Kumar, R.; Khan, A.A.; Kumar, J.; Golilarz, N.A.; Zhang, S.; Ting, Y.; Zheng, C.; Wang, W. Blockchain-Federated-Learning and Deep Learning Models for COVID-19 Detection Using CT Imaging. *IEEE Sens. J.* **2021**, *21*, 16301–16314. [[CrossRef](#)]
31. Loddo, A.; Pili, F.; di Ruberto, C. Deep Learning for COVID-19 Diagnosis from CT Images. *Appl. Sci.* **2021**, *11*, 8227. [[CrossRef](#)]
32. Frid-Adar, M.; Amer, R.; Gozes, O.; Nassar, J.; Greenspan, H. COVID-19 in CXR: From Detection and Severity Scoring to Patient Disease Monitoring. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 1892–1903. [[CrossRef](#)]
33. Tartaglione, E.; Barbano, C.A.; Berzovini, C.; Calandri, M.; Grangetto, M. Unveiling COVID-19 from Chest x-Ray with Deep Learning: A Hurdles Race with Small Data. *Int. J. Environ. Res. Public Health* **2020**, *17*, 6933. [[CrossRef](#)]

34. World Health Organization. *A Timeline of WHO's COVID-19 Response in the WHO European Region: A Living Document (Version 3.0, from 31 December 2019 to 31 December 2021)*; Licence: CC BY-NC-SA 3.0 IGO; World Health Organization: Geneva, Switzerland, 2022.
35. Mortality Analyses—Johns Hopkins Coronavirus Resource Center. Available online: <https://coronavirus.jhu.edu/data/mortality> (accessed on 8 May 2024).
36. Pang, J.; Huang, Y.; Xie, Z.; Li, J.; Cai, Z. Collaborative City Digital Twin for the COVID-19 Pandemic: A Federated Learning Solution. *Tsinghua Sci. Technol.* **2021**, *26*, 759–771. [[CrossRef](#)]
37. Ng, D.; Lan, X.; Yao, M.M.S.; Chan, W.P.; Feng, M. Federated Learning: A Collaborative Effort to Achieve Better Medical Imaging Models for Individual Sites That Have Small Labelled Datasets. *Quant. Imaging Med. Surg.* **2021**, *11*, 852–857. [[CrossRef](#)]
38. Privacy | HHS.Gov. Available online: <https://www.hhs.gov/hipaa/for-professionals/privacy/index.html> (accessed on 8 November 2022).
39. Processing—General Data Protection Regulation (GDPR). Available online: <https://gdpr-info.eu/issues/processing/> (accessed on 8 November 2022).
40. Yi, P.H.; Wei, J.; Kim, T.K.; Shin, J.; Sair, H.I.; Hui, F.K.; Hager, G.D.; Lin, C.T. Radiology “Forensics”: Determination of Age and Sex from Chest Radiographs Using Deep Learning. *Emerg. Radiol.* **2021**, *28*, 949–954. [[CrossRef](#)]
41. Qian, F.; Zhang, A. The Value of Federated Learning during and Post-COVID-19. *Int. J. Qual. Health Care* **2021**, *33*, mzab010. [[CrossRef](#)]
42. Banda, J.M.; Tekumalla, R.; Wang, G.; Yu, J.; Liu, T.; Ding, Y.; Artemova, E.; Tutubalina, E.; Chowell, G. A Large-Scale COVID-19 Twitter Chatter Dataset for Open Scientific Research—An International Collaboration. *Epidemiologia* **2021**, *2*, 315–324. [[CrossRef](#)] [[PubMed](#)]
43. Xia, T.; Spathis, D.; Brown, C.; Chauhan, J.; Grammenos, A.; Han, J.; Hasthanasombat, A.; Bondareva, E.; Dang, T.; Floto, A.; et al. COVID-19 Sounds: A Large-Scale Audio Dataset for Digital Respiratory Screening. In Proceedings of the Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2), Virtual-only Conference, August 2021; pp. 1–13.
44. Kvak, D.; Bendik, M.; Chromcova, A. Towards Clinical Practice: Design and Implementation of Convolutional Neural Network-Based Assistive Diagnosis System for COVID-19 Case Detection from Chest X-Ray Images. *arXiv* **2022**, arXiv:2203.10596.
45. Golubev, A. Dicom Network Implementation and Usage in the Context of the Covid-19 Pandemic. *Arch. Balk. Med. Union* **2021**, *56*, 80–87. [[CrossRef](#)]
46. Aiello, M.; Esposito, G.; Pagliari, G.; Borrelli, P.; Brancato, V.; Salvatore, M. How Does DICOM Support Big Data Management? Investigating Its Use in Medical Imaging Community. *Insights Imaging* **2021**, *12*, 164. [[CrossRef](#)]
47. Tsai, E.B.; Simpson, S.; Lungren, M.P.; Hershman, M.; Roshkovan, L.; Colak, E.; Erickson, B.J.; Shih, G.; Stein, A.; Kalpathy-Cramer, J.; et al. The RSNA International COVID-19 Open Radiology Database (RICORD). *Radiology* **2021**, *299*, E204–E213. [[CrossRef](#)] [[PubMed](#)]
48. Vayá, M.d.I.I.; Saborit, J.M.; Montell, J.A.; Pertusa, A.; Bustos, A.; Cazorla, M.; Galant, J.; Barber, X.; Orozco-Beltrán, D.; García-García, F.; et al. BIMCV COVID-19+: A Large Annotated Dataset of RX and CT Images from COVID-19 Patients. *arXiv* **2020**, arXiv:2006.01174.
49. Peng, L.; Luo, G.; Walker, A.; Zaiman, Z.; Jones, E.K.; Gupta, H.; Kersten, K.; Burns, J.L.; Harle, C.A.; Magoc, T.; et al. Evaluation of Federated Learning Variations for COVID-19 Diagnosis Using Chest Radiographs from 42 US and European Hospitals. *J. Am. Med. Inform. Assoc.* **2023**, *30*, 54–63. [[CrossRef](#)]
50. McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; y Arcas, B.A. Communication-Efficient Learning of Deep Networks from Decentralized Data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, Ft. Lauderdale, FL, USA, 22 April 2017; Volume 54, p. 10.
51. Darzidehkalani, E. Federated Learning in Medical Image Analysis. *Pattern Recognit.* **2024**, *151*, 110424.
52. Shyu, C.; Putra, K.T.; Chen, H.; Tsai, Y.; Hossain, K.S.M.T.; Jiang, W.; Shae, Z. A Systematic Review of Federated Learning in the Healthcare Area: From the Perspective of Data Properties and Applications. *Appl. Sci.* **2021**, *11*, 11191. [[CrossRef](#)]
53. Liu, B.; Yan, B.; Zhou, Y.; Yang, Y.; Zhang, Y. Experiments of Federated Learning for COVID-19 Chest X-Ray Images. *arXiv* **2020**, arXiv:2007.05592.
54. Abdul, M.; Id, S.; Taha, S.; Ramadan, M. COVID-19 Detection Using Federated Machine Learning. *PLoS ONE* **2021**, *16*, e0252573. [[CrossRef](#)] [[PubMed](#)]
55. Li, T.; Sahu, A.K.; Zaheer, M.; Sanjabi, M.; Talwalkar, A.; Smith, V. Federated Optimization In Heterogeneous Networks. *Proc. Mach. Learn. Syst.* **2020**, *2*, 429–450.
56. Kaissis, G.; Ziller, A.; Passerat-Palmbach, J.; Ryffel, T.; Usynin, D.; Trask, A.; Lima, I.; Mancuso, J.; Jungmann, F.; Steinborn, M.M.; et al. End-to-End Privacy Preserving Deep Learning on Multi-Institutional Medical Imaging. *Nat. Mach. Intell.* **2021**, *3*, 473–484. [[CrossRef](#)]
57. Guha Roy, A.; Siddiqui, S.; Pölsterl, S.; Navab, N.; Wachinger, C. BrainTorrent: A Peer-to-Peer Environment for Decentralized Federated Learning. *arXiv* **2019**, arXiv:1905.06731.
58. Li, X.; Gu, Y.; Dvornek, N.; Staib, L.H.; Ventola, P.; Duncan, J.S. Multi-Site fMRI Analysis Using Privacy-Preserving Federated Learning and Domain. *Med. Image Anal.* **2020**, *65*, 101765. [[CrossRef](#)] [[PubMed](#)]

59. Dou, Q.; So, T.Y.; Jiang, M.; Liu, Q.; Vardhanabhuti, V.; Kaissis, G.; Li, Z.; Si, W.; Lee, H.H.C.; Yu, K.; et al. Federated Deep Learning for Detecting COVID-19 Lung Abnormalities in CT: A Privacy-Preserving Multinational Validation Study. *NPJ Digit. Med.* **2021**, *4*, 60. [[CrossRef](#)] [[PubMed](#)]
60. Zhang, L.; Shen, B.; Barnawi, A.; Xi, S.; Kumar, N.; Wu, Y. FedDPGAN: Federated Differentially Private Generative Adversarial Networks Framework for the Detection of COVID-19 Pneumonia. *Inf. Syst. Front.* **2021**, *23*, 1403–1415. [[CrossRef](#)]
61. Nguyen, D.C.; Ding, M.; Member, S.; Pathirana, P.N.; Member, S. Federated Learning for COVID-19 Detection with Generative Adversarial Networks in Edge Cloud Computing. *IEEE Internet Things J.* **2021**, *9*, 10257–10271. [[CrossRef](#)]
62. Wang, Z.; Cai, L.; Zhang, X.; Choi, C.; Su, X. Research Article A COVID-19 Auxiliary Diagnosis Based on Federated Learning and Blockchain. *Comput. Math. Methods Med.* **2022**, *2022*, 7078764. [[PubMed](#)]
63. Yang, D.; Xu, Z.; Li, W.; Myronenko, A.; Roth, H.R.; Harmon, S.; Xu, S.; Turkbey, B.; Turkbey, E.; Wang, X.; et al. Federated Semi-Supervised Learning for COVID Region Segmentation in Chest CT Using Multi-National Data from China, Italy, Japan. *Med. Image Anal.* **2021**, *70*, 101992. [[CrossRef](#)]
64. Dayan, I.; Roth, H.R.; Zhong, A.; Harouni, A.; Gentili, A.; Abidin, A.Z.; Liu, A.; Costa, A.B.; Wood, B.J.; Tsai, C.S.; et al. Federated Learning for Predicting Clinical Outcomes in Patients with COVID-19. *Nat. Med.* **2021**, *27*, 1735–1743. [[CrossRef](#)]
65. Jiang, M.; Wang, Z.; Dou, Q. HarmoFL: Harmonizing Local and Global Drifts in Federated Learning on Heterogeneous Medical Images. *Proc. AAAI Conf. Artif. Intell.* **2022**, *36*, 1087–1095. [[CrossRef](#)]
66. Bhattacharya, A.; Gawali, M.; Seth, J.; Kulkarni, V. Application of Federated Learning in Building a Robust COVID-19 Chest X-Ray Classification Model. *arXiv* **2022**, arXiv:2204.10505.
67. Zhou, J.; Zhou, L.; Wang, D.; Xu, X.; Li, H.; Chu, Y.; Han, W.; Gao, X. Personalized and Privacy-Preserving Federated Heterogeneous Medical Image Analysis with PPPML-HMI. *Comput. Biol. Med.* **2024**, *169*, 107861. [[CrossRef](#)]
68. Feki, I.; Ammar, S.; Kessentini, Y.; Muhammad, K. Federated Learning for COVID-19 Screening from Chest X-Ray Images. *Appl. Soft Comput.* **2021**, *106*, 107330. [[CrossRef](#)]
69. Bai, X.; Wang, H.; Ma, L.; Xu, Y.; Gan, J.; Fan, Z.; Yang, F.; Ma, K.; Yang, J.; Bai, S.; et al. Advancing COVID-19 Diagnosis with Privacy-Preserving Collaboration in Artificial Intelligence. *Nat. Mach. Intell.* **2021**, *3*, 1081–1089. [[CrossRef](#)] [[PubMed](#)]
70. Lo, S.K.; Liu, Y.; Lu, Q.; Wang, C.; Xu, X.; Paik, H.-Y.; Zhu, L. Blockchain-Based Trustworthy Federated Learning Architecture. *arXiv* **2021**, arXiv:2108.06912.
71. Malik, H.; Naeem, A.; Naqvi, R.A.; Loh, W.K. DMFL\_Net: A Federated Learning-Based Framework for the Classification of COVID-19 from Multiple Chest Diseases Using X-Rays. *Sensors* **2023**, *23*, 743. [[CrossRef](#)]
72. Li, X.; Jiang, M.; Zhang, X.; Kamp, M.; Dou, Q. FedBN: Federated Learning on Non-IID Features via Local Batch Normalization. *arXiv* **2021**, arXiv:2102.07623.
73. Kumar, R.; Kumar, J.; Aman, A.; Ali, H.; Bernard, C.M.; Ullah, R.; Zeng, S. Blockchain and Homomorphic Encryption Based Privacy-Preserving Model Aggregation for Medical Images. *Comput. Med. Imaging Graph.* **2022**, *102*, 102139. [[CrossRef](#)]
74. Dong, N.; Voiculescu, I. Federated Contrastive Learning for Decentralized Unlabeled Medical Images. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021, Proceedings of the 24th International Conference, Strasbourg, France, 27 September–1 October 2021*; Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Berlin/Heidelberg, Germany, 2021; Volume 12903 LNCS, pp. 378–387. [[CrossRef](#)]
75. Ho, T.T.; Tran, K.D.; Huang, Y. FedSGDCOVID: Federated SGD COVID-19 Detection under Local Differential Privacy Using Chest X-Ray Images and Symptom Information. *Sensors* **2022**, *22*, 3728. [[CrossRef](#)]
76. Chowdhury, D.; Banerjee, S.; Sannigrahi, M.; Dey, A.; Dhar, A.; Chakraborty, A.; Das, A. Federated Learning Based Covid-19 Detection. *Expert Syst.* **2023**, *40*, e13173. [[CrossRef](#)]
77. Kumar, R.; Wang, W.; Yuan, C.; Kumar, J.; Zheng, C.; Aman, A. Blockchain Based Privacy-Preserved Federated Learning for Medical Images: A Case Study of COVID-19 CT Scans. *arXiv* **2021**, arXiv:2104.10903.
78. Florescu, L.M.; Streba, C.T.; Șerbănescu, M.S.; Mămuleanu, M.; Florescu, D.N.; Teică, R.V.; Nica, R.E.; Gheonea, I.A. Federated Learning Approach with Pre-Trained Deep Learning Models for COVID-19 Detection from Unsegmented CT Images. *Life* **2022**, *12*, 958. [[CrossRef](#)] [[PubMed](#)]
79. Zhang, W.; Zhou, T.; Lu, Q.; Wang, X.; Zhu, C.; Sun, H.; Wang, Z.; Lo, S.K.; Wang, F.-Y. Dynamic Fusion-Based Federated Learning for COVID-19 Detection. *IEEE Internet Things* **2021**, *8*, 15884–15891. [[CrossRef](#)] [[PubMed](#)]
80. Qayyum, A.; Ahmad, K.; Ahsan, M.A.; Al-Fuqaha, A.; Qadir, J. Collaborative Federated Learning For Healthcare: Multi-Modal COVID-19 Diagnosis at the Edge. *IEEE Open J. Comput. Soc.* **2022**, *3*, 1–10. [[CrossRef](#)]
81. Adhikari, R.; Settles, C. Secure Federated Learning Approaches to Diagnosing COVID-19. *arXiv* **2024**, arXiv:2401.12438.
82. Kareem, A.; Liu, H.; Velisavljevic, V. A Federated Learning Framework for Pneumonia Image Detection Using Distributed Data. *Healthc. Anal.* **2023**, *4*, 100204. [[CrossRef](#)]
83. Jablęcki, P.; Ślęzyk, F.; Malawski, M. Federated Learning in the Cloud for Analysis of Medical Images—Experience with Open Source Frameworks. In *Clinical Image-Based Procedures, Distributed and Collaborative Learning, Artificial Intelligence for Combating COVID-19 and Secure and Privacy-Preserving Machine Learning, Proceedings of the 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, 27 September and 1 October 2021*; Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Berlin/Heidelberg, Germany, 2021; Volume 12969 LNCS, pp. 111–119. [[CrossRef](#)]

84. Darzi, E.; Sijtsema, N.M.; van Ooijen, P.M.A. A Comparative Study of Federated Learning Methods for COVID-19 Detection. *Sci. Rep.* **2024**, *14*, 3944. [[CrossRef](#)] [[PubMed](#)]
85. Sun, G.; Shu, H.; Shao, F.; Racharak, T.; Kong, W.; Pan, Y.; Dong, J.; Wang, S.; Nguyen, L.M.; Xin, J. FKD-Med: Privacy-Aware, Communication-Optimized Medical Image Segmentation via Federated Learning and Model Lightweighting Through Knowledge Distillation. *IEEE Access* **2024**, *12*, 33687–33704. [[CrossRef](#)]
86. Balachandar, N.; Chang, K.; Kalpathy-Cramer, J.; Rubin, D.L. Accounting for Data Variability in Multi-Institutional Distributed Deep Learning for Medical Imaging. *J. Am. Med. Inform. Assoc.* **2020**, *27*, 700–708. [[CrossRef](#)] [[PubMed](#)]
87. Durga, R.; Poovammal, E. FLED-Block: Federated Learning Ensembled Deep Learning Blockchain Model for COVID-19 Prediction. *Front. Public Health* **2022**, *10*, 892499. [[CrossRef](#)]
88. Jothimurugesan, E.; Hsieh, K.; Wang, J.; Joshi, G.; Gibbons, P.B. Federated Learning under Distributed Concept Drift. In Proceedings of the 26th International Conference on Artificial Intelligence and Statistics, Valencia, Spain, 25–27 April 2023; Volume 206, pp. 5834–5853.
89. Chetoui, M.; Akhloufi, M.A. Federated Learning Approach for Early Detection Federated Learning for COVID-19 Detection. *Computers* **2023**, *12*, 106. [[CrossRef](#)]
90. Kandati, D.R.; Gadekallu, T.R. Federated Learning Approach for Early Detection of Chest Lesion Caused by COVID-19 Infection Using Particle Swarm Optimization. *Electronics* **2023**, *12*, 710. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.