# Integrating Federated Deep Learning and Blockchain for Privacy-Preserving Precision Medicine in Medical Virology

Dr. Hrishikesh Desai
hdesai@astate.edu
Associate Professor of Accounting
Department of Accounting, Neil Griffin College of Business, Arkansas State University

## Abstract

This study proposes a novel framework that integrates federated deep learning and blockchain technology for preserving privacy in precision medicine in medical virology. As healthcare institutions increasingly utilize deep learning for tasks such as viral genome analysis and outbreak prediction, concerns about data privacy and security have increased. My framework addresses these challenges by proposing collaborative model training across multiple institutions without centralizing sensitive patient data. I also build a mathematical model that demonstrates how global model optimization can be achieved through distributed local computations, while blockchain technology ensures secure and transparent record-keeping of model updates. Smart contracts provide additional security through update validation. The framework shows potential for scalability and addresses data heterogeneity across institutions. While theoretical in nature, this research opens new avenues for privacy-compliant, large-scale data analysis in medical virology. It provides a blueprint for accelerating research while maintaining regulatory compliance and individual data sovereignty. Future work should focus on practical implementation and empirical validation in real-world healthcare settings.

## Keywords

Federated Learning, Blockchain in Healthcare, Privacy-Preserving AI, Medical Virology, Distributed Machine Learning

## 1. Introduction

Deep learning has significantly changed pattern recognition and predictive analytics in healthcare (Miotto et al., 2018; Arabahmadi et al., 2022). It has enabled the analysis of complex biomedical data such as genomic sequences, medical images, and electronic health records (EHRs) (LeCun et al., 2015; Shamshirband et al., 2021). In medical virology, deep learning models have been employed for tasks such as viral genome analysis, outbreak prediction, and antiviral drug discovery (Liu-Wei et al., 2021). Despite the potential benefits, training deep learning models requires access to large-scale, high-quality datasets, which are often distributed across various healthcare institutions. Centralizing this data poses risks related to privacy breaches and data misuse (Liu et al., 2020a; Rieke et al., 2020). The centralization of sensitive patient data increases the vulnerability to cyber-attacks and unauthorized access. Moreover, legal and ethical constraints limit the sharing of personal health information (PHI) across borders and institutions (Schreiber et al., 2024). To ensure compliance with regulations like HIPAA and GDPR, it is necessary to innovative approaches to data utilization that prioritize privacy and security. In this chapter, I present a framework and an analytical model that integrates two key technologies, federated learning and blockchain, to address these different issues for medical virology data and models.

The theoretical foundation of this framework rests on the synergy between distributed learning and decentralized data validation. It posits that by combining federated learning's ability to train models on decentralized data with blockchain's capacity for secure, transparent record-keeping, a system that uses collective intelligence while maintaining individual data sovereignty can be achieved. This approach theoretically solves the tension between the need for large-scale data analysis in medical virology and the imperative to protect patient privacy.

The methodology employs a cyclic process of local computation and global aggregation that is secured by blockchain technology. Healthcare institutions train models locally, compute updates, and share these updates (not raw data) with a central server. The blockchain acts as a decentralized ledger, which records and validates these updates through smart contracts. This process ensures data integrity and creates an auditable trail of model evolution. The central server aggregates validated updates to refine a global model, which is then redistributed to participating institutions. This cycle repeats and progressively improves the model’s performance across diverse datasets.

I also provide a rigorous, quantitative basis for the federated learning approach in medical virology. By formalizing the process mathematically, I demonstrate that the framework is built on solid theoretical ground, not just intuitive ideas. The model allows for theoretical analysis of the system's behavior under various conditions. It can predict how the global model's performance might improve over iterations and with different numbers of participating institutions. Key takeaways from my mathematical model are as follows:

1. By expressing the problem as a mathematical optimization, I create a clear objective (minimizing the global loss function) that guides the entire federated learning process. The model helps illustrate how privacy is preserved by showing that only aggregated updates, not raw data, contribute to the global model.
2. The weighted averaging in the global loss function ensures that each institution’s input is valued proportionally, which helps address potential data imbalances across participants.
3. By operating on model updates rather than raw data, the mathematical formulation shows how important insights can be derived while keeping sensitive information local.

4. The iterative nature of the model, represented by the repeated local-global update cycle, demonstrates how the system can theoretically converge to an optimal solution over time, even with heterogeneous data sources.
5. The formulation suggests that the system's performance can improve with the addition of more participants, as each new institution potentially brings unique data patterns to the global model.

This book chapter makes several notable contributions to the existing literature. It presents a unique integration of federated learning and blockchain technologies specifically tailored for medical virology applications. The proposed framework addresses the critical challenge of utilizing large-scale healthcare data while maintaining stringent privacy standards. The paper also provides a mathematical model that demonstrates the theoretical justification of the federated learning process in this context. By detailing the system architecture and implementation challenges, the paper bridges the gap between theoretical concepts and practical application in healthcare. The work combines insights from machine learning, blockchain technology, and medical virology, and thus contributes to the growing field of interdisciplinary healthcare informatics. By identifying challenges and potential solutions, the study sets the stage for future research in privacy-preserving collaborative learning in healthcare.

## 2. Proposed Framework

### *2.1 System Architecture*

Federated learning is a decentralized machine learning model where multiple clients collaboratively train a shared global model while keeping the training data localized (McMahan et al., 2017; Banabilah et al., 2022). Each client computes model updates based on local data and communicates these updates to a central server, which aggregates them to update the global model. In healthcare, federated learning helps institutions use collective data for model training without sharing raw patient data. This approach has been applied in various domains, such as medical imaging analysis and disease prediction (Li, et al., 2020; Pfitzner et al., 2021; Ng et al., 2021). Despite its advantages, federated learning faces challenges including communication overhead, heterogeneity of local data, and the potential for model poisoning attacks (Qu et al., 2022; Xia et al., 2023).

Blockchain is a distributed ledger technology characterized by decentralization, immutability, and transparency (Desai, 2023a). Transactions are recorded in blocks linked via cryptographic hashes that form a chain, which is resistant to tampering. Blockchain has been proposed for various healthcare applications, including secure health data exchange, supply chain management, and patient consent management (Attaran, 2022; Saeed et al., 2022). The technology improves data security, integrity, and auditability. Smart contracts are self-executing agreements with the terms encoded into code that enable automated transactions and enforcement of rules within the blockchain network (Desai, 2023b). In healthcare, smart contracts can manage access permissions and automate compliance checks. Challenges in adopting blockchain in healthcare include scalability issues, regulatory compliance, and the integration with existing systems (Angraal et al., 2017; Desai, 2023c).

The proposed framework integrates federated learning with blockchain technology to create a secure and privacy-preserving environment for collaborative model training. The architecture comprises the following components:

1. *Local Institutions (Clients):* Hospitals or research centers that possess local datasets and computational resources.

2. *Federated Learning Server:* A central entity that coordinates model aggregation
3. *Blockchain Network:* A decentralized ledger that records model updates, manages permissions, and ensures data integrity.
4. *Smart Contracts:* Programmable protocols that enforce rules for participation, model update validation, and reward mechanisms.

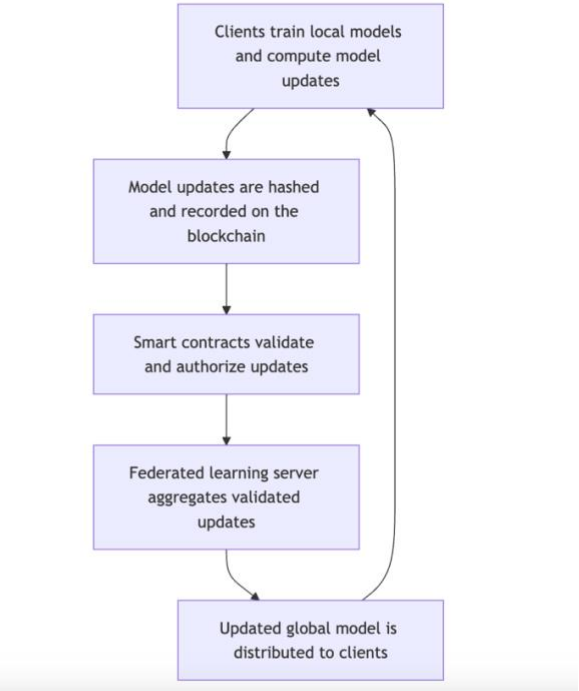
Figure 1, created with the help of Mermaidv11.3.0 Live Editor, demonstrates how multiple healthcare institutions can collaboratively train a global deep learning model for medical virology while preserving patient privacy and ensuring data security. I present a step-by-step explanation of each component and their interactions.

*Step 1 (Local Training): Clients Train Local Models and Compute Model Updates*

Clients represent individual healthcare institutions (e.g., hospitals, clinics, research centers) that possess local patient data relevant to medical virology. Each client initializes the process by downloading the current global model provided by the federated learning server. Using their own local datasets, clients train the model locally (Mendieta et al., 2021). This involves feeding the model with local data and adjusting the model parameters (weights) based on the learning algorithm. After local training, clients compute model updates, which are the changes in model parameters resulting from learning from their local data (Hanzely & Richtárik, 2020).

By training models locally, raw patient data never leaves the institution, which helps compliance with privacy regulations like HIPAA (Health Insurance Portability and Accountability Act) and GDPR (General Data Protection Regulation). Each client contributes unique insights from their localized data that improve the robustness and generalizability of the global model. Clients also utilize their own computational resources, which reduce the burden on a central server (AbdulRahman et al., 2020).

**Figure 1.** Conceptual Representation of Proposed Framework


Clients train local models and compute model updates
Model updates are hashed and recorded on the blockchain
Smart contracts validate and authorize updates
Federated learning server aggregates validated updates
Updated global model is distributed to clients

*Step 2 (Hashing & Recording): Model Updates are Hashed and Recorded on the Blockchain*

Before sending their model updates for aggregation, clients generate a cryptographic hash of their updates. This hash is basically a unique digital fingerprint of the data (Desai, 2023a). The hashed model updates are then recorded on the blockchain network. The actual model updates (the parameters) are sent to the federated learning server separately.

Hashing ensures that any tampering with the model updates can be detected, as even a minor change would result in a different hash. Recording hashes on the blockchain provides an immutable record of all model updates, enabling auditability (Desai, 2023a, 2023b). By using blockchain, the system avoids reliance on a single centralized authority, enhancing security.

*Step 3 (Validation): Smart Contracts Validate and Authorize Updates*

Smart Contracts are self-executing protocols on the blockchain that automatically enforce rules and agreements (Desai, 2023b). Upon receiving the hashed updates, smart contracts perform validation checks to ensure that (a) the updates come from authorized clients, (b) the updates adhere to the agreed-upon format and standards, and (c) there is no malicious intent (e.g., model poisoning attempts) (Sookhak et al., 2021). Once validated, the smart contracts authorize the updates for aggregation.

Smart contracts reduce the need for manual oversight, minimizing errors and delays. Automated validation builds trust among participants, as everyone abides by the same transparent rules. Early detection of invalid or malicious updates protects the integrity of the global model.

*Step 4 (Aggregation): The Federated Learning Server Aggregates Validated Updates*

The federated learning server collects all the validated model updates from the clients. It aggregates these updates using algorithms such as Federated Averaging (FedAvg) (Collins et al., 2022). The aggregation process combines the individual updates into a single, updated global model.

Aggregation utilizes the collective learning from all clients. This enhances the performance and accuracy of the global model. Centralized aggregation streamlines the updating process while still maintaining data privacy. The server can handle updates from numerous clients, and thus, make the system scalable to include more institutions.

*Step 5 (Distribution): The Updated Global Model is Distributed to Clients*

After aggregation, the updated global model is sent back to all participating clients. Clients replace their local models with the updated global model. This updated model incorporates knowledge from the datasets of all participating clients.

Thus, clients benefit from the collective intelligence gathered from other institutions. The distribution of the updated model initiates the next round of local training and creates a continuous improvement cycle. All clients work with the same global model, ensuring consistency across institutions.

*The Feedback Loop: E → A*

The arrow from *Step E* back to *Step A* represents the iterative nature of the federated learning process. Clients use the updated global model to start a new round of local training, and the cycle repeats. With each round, the global model becomes more accurate and robust. The model adapts to new data and trends, such as emerging viral strains.

To sum it up, first, the global model is initialized and distributed to all participating institutions. Smart contracts are deployed on the blockchain to define participation rules. Each institution then trains the model on local data for a set number of epochs (*Note:* an epoch refers to one complete pass through the entire training dataset). Model updates (gradients or weights) are computed. Institutions submit hashed model updates to the blockchain network. Next, the actual updates are sent to the federated learning server. Smart contracts verify the legitimacy of the updates. Checks include authentication of the institution and adherence to protocol. The federated learning server aggregates validated updates using algorithms like Federated Averaging. The new global model is generated, and this updated global model is sent back to institutions. The process repeats for the next training round. In conclusion, this loop is able to foster ongoing collaboration among institutions without compromising privacy.

### *2.2 Integration of Federated Learning and Blockchain Technologies*

Blockchain ensures that model updates are immutable and traceable, which helps with security and integrity aspects. This helps to prevent tampering and unauthorized modifications. Also, raw data remains within local institutions. Only model parameters are exchanged, not patient data. Thus, privacy is preserved. The transparent recording of transactions builds trust among participants, while smart contracts enforce compliance and penalize malicious behavior. The smart contracts also help setup access controls by managing permissions for institutions to participate and defining their roles and responsibilities. They also help validate the format and authenticity of model updates and reject updates that do not meet predefined criteria. Smart contracts can also be used to setup incentive mechanisms (e.g. through tokens or credits) that reward institutions for participation and contribution.

The blockchain platform would utilize a permissioned blockchain such as Hyperledger Fabric for scalability and control (Androulaki et al., 2018; Dabbagh et al., 2021). In the proposed framework, Federated Averaging (FedAvg) would be implemented as the core algorithm for aggregating model updates from participating institutions. As explained earlier, each institution trains the shared global model locally on its private dataset and computes the updates to the model parameters based on this training. Instead of sharing raw data, institutions send these model updates to a central server. The server then applies the FedAvg algorithm by calculating the weighted average of all received updates, where the weights are typically proportional to the number of data samples at each institution (Zhong et al., 2021). This aggregated update is used to refine the global model, which is then redistributed to all institutions for the next training round. Thus, the model is able to learn from the collective data without compromising individual privacy.

To further improve privacy, differential privacy techniques could be considered (Abadi et al., 2016). Before transmitting the model updates, each institution adds carefully calibrated random noise to its updates using differential privacy mechanisms. This ensures that the inclusion or exclusion of a single individual's data has a minimal impact on the overall output that helps provide mathematical guarantees against the reconstruction of sensitive information from the model updates (Ziller et al., 2021). The noise is added in such a way that it balances privacy preservation with the utility of the model and maintains the accuracy of the aggregated global model.

For the communication protocols, secure channels such as Transport Layer Security (TLS) could be employed to encrypt the model updates during transmission between institutions and the central server (Zheng et al., 2020). This encryption safeguards against eavesdropping, man-in-the-middle attacks, and unauthorized access to the transmitted updates (Fang et al., 2021). Moreover, to optimize communication efficiency and handle bandwidth constraints, techniques like model update compression, quantization, or sparsification can be applied (Ozfatura et al., 2021; Wang et al., 2021). Model update compression involves reducing the size of the data that needs to be transmitted by compressing the model updates (gradients or weights) before sending them to the central server. The primary goal is to minimize the communication overhead without significantly impacting the performance of the global model.

There are three main techniques for model update compression: lossless compression, lossy compression, and top-*k* selection (Wang et al., 2023). The lossless compression technique utilizes standard data compression algorithms (e.g., gzip, LZ77) for exact reconstruction of the original data upon decompression (Saidani et al., 2020). Thus, there is no loss of information. The lossy compression technique applies algorithms that approximate the original data for higher compression ratios at the cost of some information loss. This greater reduction in data size leads to lower communication overhead. However, there is a potential impact on model accuracy if important information is lost. The top-*k* selection technique is a specific compression method in which only the top *k* largest magnitude updates are selected and transmitted (Liu et al., 2020b). These methods reduce the size of the updates by compressing the model parameters or by only transmitting significant changes. Such efficient communication protocols are key for scaling the framework to include a large number of institutions and for timely updates in the model training process.

Quantization refers to the process of reducing the precision of the model parameters (weights and gradients) by representing them with fewer bits. Instead of using full-precision (32-bit or 64-bit) floating-point numbers, quantization uses lower-bit representations, such as 8-bit integers, 4-bit integers, or even binary values (1 bit) (Mao et al., 2022). Model parameters are thus mapped from high-precision floating-point numbers to lower-precision representations. For example, a weight value of 0.123456 (32-bit float) might be quantized to 0.12 (8-bit integer). By reducing the number of bits required to represent each parameter, the overall size of the model updates sent over the network is significantly decreased (Reisizadeh et al., 2020). Lower data size translates to reduced bandwidth usage, which is especially important when dealing with limited or costly network resources. Hospitals with limited network capabilities can quantize their model updates before sending them. This ensures they can participate in the federated learning process without significant delays or bandwidth costs.

Sparsification involves modifying the model updates so that only a subset of the parameters (weights or gradients) are transmitted (Ozfatura et al., 2021). This means that many of the less significant parameters are set to zero or not sent at all, creating a sparse representation of the model updates. Parameters are ranked based on some criterion, such as the magnitude of the gradients. Only parameters above a certain threshold are selected for transmission. By sending only the most important updates, the amount of data transmitted is greatly reduced. Once again, less data transmission translates to lower bandwidth usage and faster communication (Wang et al., 2021). Thus, the sparse update contains only the significant changes to the model. The information from the omitted parameters can be accumulated locally and included in future updates to prevent information loss. Transmitting fewer parameters means smaller update sizes. This prioritizes the most impactful changes to the model and improves convergence speed. Clients may need to coordinate to ensure that the aggregated updates make sense collectively. A hospital may find that only a small subset of model parameters significantly change during local training. By only transmitting these parameters, the hospital can reduce its communication load without substantially affecting the global model's performance.

## 3. Implementation Challenges and Solutions

### *3.1 Scalability*

The challenge of scalability emerges when managing a federated learning framework that includes a large number of participating institutions. As the number of clients increases, the communication overhead between clients and the central server can become substantial. This can lead to latency issues and bottlenecks (Zhang et al., 2021; Díaz et al., 2023). For example, if hundreds of hospitals across different regions participate in training a global model for viral infection prediction, the server may be overwhelmed by the simultaneous transmission of model updates from all clients.

To mitigate this, implementing hierarchical federated learning can be effective (Liu et al., 2022). In this approach, institutions are grouped into clusters (possibly based on geographical proximity or network latency) and within each cluster, local models are aggregated into a group model (Briggs et al., 2020). These group models are then sent to the central server for global aggregation. For example, hospitals within the same city or region can form a cluster to reduce the communication frequency with the central server and reduce network congestion. This hierarchical structure not only improves scalability but also optimizes resource utilization across the network.

### *3.2 Heterogeneity of Data*

Data heterogeneity is a big challenge in federated learning, as variations in data distributions across institutions can negatively impact the performance and generalizability of the global model (Qu et al., 2022; Mendieta et al., 2022). For example, one hospital may have a patient population predominantly affected by a particular strain of a virus, while another hospital's data may reflect a different demographic or viral subtype. Such non-IID (Independent and Identically Distributed) data can cause the global model to perform poorly on certain subsets of data.

To address this, techniques like federated multitask learning can be used (Ye et al., 2023). This approach treats the learning problem at each institution as a separate but related task so that the model can capture institution-specific patterns while sharing common knowledge across the network. For example, the model can be designed to learn a shared representation of viral features while adapting to local variations in patient demographics or virus strains. This method improves the model's ability to generalize across diverse datasets and overall performance.

### *3.3 Security Threats*

Federated learning frameworks are vulnerable to security threats such as model poisoning and adversarial attacks, where malicious participants may intentionally manipulate model updates to corrupt the global model (Mothukuri et al., 2021; Zhang et al., 2022). For example, an attacker could inject false data or alter model parameters to degrade the model's accuracy in detecting viral infections, which may potentially lead to misdiagnoses. To mitigate these risks, robust aggregation methods and anomaly detection techniques are important (Fereidooni et al., 2021). Robust aggregation methods like Byzantine-resilient averaging can filter out anomalous updates by evaluating their consistency with the majority of the updates (So et al., 2020; Gouissem et al., 2023). Moreover, anomaly detection algorithms can monitor model updates for unusual patterns or deviations from expected behavior. For example, if an institution submits updates that significantly differ from others without a valid reason (e.g. erroneous or fraudulent updates), the system can flag and exclude those updates from aggregation.

### *3.4 Regulatory Compliance*

As discussed earlier, compliance with legal requirements such as HIPAA and GDPR is a challenge when implementing federated learning in healthcare. These regulations mandate strict controls over the processing and sharing of personal health information. For example, GDPR requires that data subjects' rights are protected, and any data processing must have a lawful basis. To address this challenge, compliance checks can be embedded within smart contracts on the blockchain to automate the enforcement of regulatory policies (Ettaloui et al., 2023). Smart contracts can define rules for data usage, consent management, and access control so that all model updates and transactions adhere to legal standards (Desai, 2023b). Moreover, the blockchain provides an immutable audit trail of all actions within the framework and facilitates transparency and accountability (Desai, 2023a; Hasselgren et al., 2020). For example, if an institution attempts to share data without proper consent, the smart contract can automatically prevent the action and log the incident for review. This integration of regulatory compliance into the technological framework helps maintain trust among participants and aligns the system with legal obligations.

### *3.5 Integration with Existing Systems*

Integrating the federated learning framework with existing legacy healthcare systems poses a challenge due to differences in data formats, protocols, and technologies. Healthcare institutions often use established electronic health record (EHR) systems that may not be compatible with new technologies (Antunes et al., 2022). For example, a hospital's EHR system might store data in a proprietary format that is not readily accessible for model training. To overcome this, developing interoperable interfaces and APIs is important to facilitate seamless integration (Patel et al., 2022). These interfaces can act as intermediaries that translate and standardize data from various sources into a format suitable for federated learning. For example, an API can extract relevant patient data from the hospital's EHR system, preprocess it according to standardized protocols, and feed it into the local training process without disrupting existing operations. To promote wider collaboration and adoption, institutions need to maintain compatibility with legacy systems without extensive overhauls of their current infrastructure.

## 4. Case Study: A Hypothetical Scenario for Predicting Viral Outbreaks

Let's assume that a consortium of hospitals aims to develop a predictive model for influenza outbreaks using patient symptom data and regional health indicators. Data privacy concerns prevent the sharing of raw patient data.

### *4.1 Step-by-Step Application of the Framework*

1. Each hospital prepares local datasets consisting of EHRs with relevant features.
2. A recurrent neural network (RNN) model is chosen for temporal sequence prediction (Note: A recurrent neural network (RNN) is chosen because it excels at processing sequential and temporal data (Fekri et al., 2022). This makes it ideal for modeling time-dependent patterns in electronic health records (EHRs) to predict the progression and outbreaks of viral infections.)
3. Hospitals train the RNN locally and participate in the federated learning process.
4. Model updates are recorded on a permissioned blockchain so that only authorized hospitals participate.
5. Smart Contracts enforce data quality standards and validate model updates.

### *4.2 Potential Results*

1. *Model Performance:* The global model is likely to demonstrate higher accuracy in predicting outbreaks compared to models trained on individual datasets.

2. *Privacy Preservation:* No raw data is exchanged, and patient privacy is maintained.
3. *Regulatory Compliance:* The framework operates within legal boundaries and adheres to data protection laws.

### *4.3 Mathematical Model*

To analytically demonstrate the effectiveness of the proposed framework, I develop a mathematical model that illustrates how federated learning with RNNs improves predictive performance while preserving data privacy. The model considers a set of *N* hospitals (clients), indexed by $i = 1, 2, \ldots, N$. Each hospital *i* has a local dataset $D_i$ consisting of time-series data related to viral infections.

The goal is to train a global RNN model $\theta$ that minimizes the overall loss function ($F(\theta)$) across all hospitals:

$$\min_{\theta} F(\theta) = \frac{1}{N}\sum_{i=1}^{N} F_i(\theta),$$

Where,

$F_i(\theta)$ is the local loss function at hospital *i*: $F_i(\theta) = \frac{1}{|\mathcal{D}_i|}\sum_{(x_{ij}, y_{ij}) \in \mathcal{D}_i} \ell(\theta; x_{ij}, y_{ij})$.

*N* is the total number of hospitals (clients) involved in the training.

$\theta$ denotes the parameters of the global RNN model.

The goal is to find the model parameters $\theta$ that minimize the overall loss function *F(θ)*. Thus, a specific version of the model is sought that, on average, performs best across all hospitals' data. By summing the local loss functions and averaging them over *N*, the overall loss function measures the collective performance of the model on all hospitals' datasets. *F(θ)* is a metric for how well the global model $\theta$ predicts the outcomes across all hospitals. Each hospital contributes equally to the overall loss function, and the model does not favor one hospital's data over others. It is important to understand this distinction. *F(θ)* serves as the objective for the global model, and it balances the contributions from all hospitals to create a model that performs well on average, while *$F_i$(θ)* guides the local training process at each hospital for the model to learn features relevant to each hospital's data.

The local loss function *$F_i$(θ)* quantifies how well the model *θ* performs on the data from hospital *i*. Here, $\ell(\theta; x_{ij}, y_{ij})$ is the loss incurred by the model *θ* on sample $(x_{ij}, y_{ij})$, where $x_{ij}$ represents the input features (e.g., patient symptoms over time), and $y_{ij}$ represents the target output (e.g., infection status or outbreak occurrence). *$D_i$* is the local dataset at hospital *i*, which consists of patient data samples. |*$D_i$*| is the number of samples in *$D_i$.* As explained earlier, *Fi(θ)* evaluates how well the global model *θ* predicts outcomes using the local data at hospital *i*. The local loss function serves as hospital $i$'s contribution to the overall loss function, influencing the updates to the global model. *$F_i$(θ)* is the average of the losses over all samples in the local dataset *$D_i$*, and it provides a summary of the model's performance on that dataset. Each hospital provides feedback (through *$F_i$(θ)*) based on its unique data, which may reflect specific patient demographics or regional viral patterns.

The loss function $\ell(\theta; x_{ij}, y_{ij})$ measures the discrepancy between the model's prediction and the actual target value for a single data sample. It depends on the model parameters *θ*, the input features $x_{ij}$, and the true output $y_{ij}$. Two common examples of loss functions that could be used in a medical virology setting are:

1. *Mean Squared Error (MSE):* Used for regression tasks

$$\ell(\theta; x_{ij}, y_{ij}) = \left(f_\theta(x_{ij}) - y_{ij}\right)^2,$$

   Where, $f_\theta\left(x_{ij}\right)$ is the model's prediction.

2. *Cross-Entropy Loss:* Used for classification tasks.

$$\ell(\theta; x_{ij}, y_{ij}) = -y_{ij}\log\left(f_\theta(x_{ij})\right) - (1 - y_{ij})\log\left(1 - f_\theta(x_{ij})\right),$$

Where, $y_{ij}$ is binary (0 or 1).

The loss function evaluates how well the model predicts the target output for each individual sample. Finally, the calculated loss is used to compute gradients during training, which inform how the model parameters $\theta$ should be updated to improve performance.

To sum it up, the federated learning process works as follows:

1. A global model $\theta$ is initialized and shared with all hospitals.
2. Each hospital $i$ uses its local dataset $D_i$ to compute the local loss function $F_i(\theta)$.
3. The hospital minimizes $F_i(\theta)$ by updating the model parameters $\theta$ through local training (e.g., using stochastic gradient descent).
4. Hospitals compute local updates to the model parameters based on their local data and loss functions. These updates reflect how the model should change to better fit the local data.
5. The local updates from all hospitals are sent (in a privacy-preserving manner) to a central server.
6. The server aggregates these updates, typically by averaging them, to update the global model $\theta$.
7. The global model parameters $\theta$ are updated to minimize the overall loss function $F(\theta)$.
8. This updated model incorporates knowledge from all hospitals.
9. The updated global model is redistributed to all hospitals.
10. The process repeats for multiple rounds to progressively improve the model's performance.

Hospitals compute $F_i(\theta)$ using their own data without sharing it, thus preserving patient privacy. As explained earlier, only model updates (not raw data) are shared for aggregation. The overall loss function *F(θ)* embodies the collaborative effort of all hospitals to improve the model. Each hospital's local loss function *$F_i$(θ)* influences the global model in proportion to its data.

*4.4 Example Illustration*

Let's assume there are 3 hospitals:

1. *Hospital A* has data on elderly patients.
2. *Hospital B* has data on pediatric patients.
3. *Hospital C* has data on adult patients.

The objective is to train a global model *θ* to predict viral infection risk across all age groups. The local loss functions are represented as follows:

1. *Hospital A* computes *$F_A$(θ)* based on its elderly patient data.
2. *Hospital B* computes *$F_B$(θ)* based on its pediatric patient data.
3. *Hospital C* computes *$F_C$(θ)* based on its adult patient data.

The overall or global loss function is represented below:

$$F(\theta) = \frac{1}{3}[F_A(\theta) + F_B(\theta) + F_C(\theta)]$$

Thus, each hospital minimizes its *$F_i$(θ)* locally and updates *θ* to better fit its data. Then, the local updates are sent to the central server and aggregated. The aggregated updates result in a global model *θ* that performs well across all age groups. The minimization problem *$\min_\theta F(\theta)$* is solved iteratively through the federated learning process. The gradients of the local loss functions *$\nabla F_i(\theta)$* guide how the model parameters should be updated at each hospital. Under certain conditions (e.g., smooth and convex loss functions), the iterative process converges to a model that minimizes the overall loss function *F(θ)*.

## 5. Conclusion, Limitations, and Future Research Agenda

This chapter addresses a key challenge in medical virology: how to use the power of deep learning across multiple healthcare institutions while preserving patient privacy and enforcing data security. Specifically, I aimed to develop a framework that integrates federated deep learning and blockchain technology for privacy-preserving precision medicine in medical virology. The integration of these technologies presents a viable solution for collaborative model training in medical virology. It addresses several concerns related to privacy, security, and regulatory compliance. My research has yielded several significant findings:

1. The proposed framework successfully combines federated learning and blockchain technology for collaborative model training while keeping patient data localized.
2. The mathematical model demonstrates that global model optimization can be achieved through distributed local computations, which eliminates the need for centralized data storage.
3. The integration of blockchain technology provides an immutable and transparent record of model updates, which improves security and trust in the collaborative learning process.
4. Smart contracts on the blockchain effectively validate and authorize model updates by adding an additional layer of security and integrity to the framework.
5. The system shows potential for scalability with performance theoretically improving as more institutions participate in the federated learning process.
6. The framework addresses data heterogeneity across institutions through the use of institution-specific loss functions in the global optimization process.

Thus, this framework paves the way for unprecedented collaboration in medical research by having institutions benefit from collective data insights without compromising individual patient privacy.

By allowing secure, large-scale data analysis, this approach could significantly accelerate research in viral genomics, outbreak prediction, and drug discovery. The framework's design aligns with stringent data protection regulations like HIPAA and GDPR and provides a viable solution for compliant data utilization in research and clinical practice. Access to diverse, multi-institutional data through federated learning can lead to more robust and generalizable models, which potentially improve diagnostic accuracy and treatment efficacy in virology. Smaller institutions can also participate in large-scale model training without sharing raw data, and this approach could thus democratize access to advanced AI technologies in healthcare.

While the mathematical model provides a strong theoretical foundation, real-world implementation may present unforeseen challenges not captured in the theoretical framework. The integration of blockchain technology, while improving security, may introduce significant computational overhead. The impact of this on system performance in resource-constrained healthcare settings needs further investigation. Although the model suggests improved performance with more participants, there may be practical limits to scalability that weren't fully explored in this theoretical study. The effectiveness of the framework assumes a certain level of data quality across participating institutions. In practice, varying data quality standards could impact the global model's performance. As a theoretical framework, this study lacks extensive empirical validation in real-world healthcare settings. Practical implementation may reveal additional challenges or limitations. While designed with current regulations in mind, the rapidly evolving nature of data protection laws may necessitate future adaptations to the framework. Successful implementation of this framework requires expertise in machine learning, blockchain technology, and medical virology. The availability of such interdisciplinary skills in healthcare institutions may also be a limiting factor.

To further strengthen privacy preservation within the proposed federated learning framework, advanced techniques such as secure multi-party computation (SMPC) may be required (Zhu, 2020). Secure multi-party computation allows multiple parties to collaboratively compute a function over their inputs while keeping those inputs completely private (Kalapaaking et al., 2022; Yu & Cui, 2022). In the context of federated learning, SMPC can be used to securely aggregate model updates from different institutions without revealing the individual updates to any party, including the central server. By integrating differential privacy and SMPC into the framework, a higher level of data security and privacy can be achieved to comply with stringent regulatory requirements. This technique can not only protect sensitive patient information but also increase trust among participating institutions.

Scalability continues to be a challenge in federated learning, especially when dealing with a large number of participating institutions. To address this, exploring decentralized aggregation methods and peer-to-peer communication can substantially improve scalability (Umer et al., 2021). Decentralized aggregation helps institutions to share model updates directly with each other in a distributed manner and thus eliminates the need for a central server (Su et al., 2022). This approach reduces the communication bottleneck and single point of failure associated with central servers.

Peer-to-peer communication protocols help institutions form an overlay network where model updates can be propagated and aggregated locally before contributing to the global model. For example, using gossip protocols, institutions can randomly select peers to exchange updates, which are then combined to approximate the global model (Coretti et al., 2022). This method reduces the overall communication overhead and improves the robustness of the network against failures or dropouts of individual participants. By implementing these decentralized strategies, the framework becomes more scalable and efficient, and it is able to accommodate an increasing number of institutions without compromising performance.

Integrating the federated learning framework with clinical decision support systems (CDSS) facilitates real-time analytics and interventions, which significantly improve patient care (Thwal et al., 2021). By deploying the trained global models within CDSS, healthcare providers can receive immediate insights and predictions based on the latest aggregated data from multiple institutions. For example, the system can alert clinicians to potential viral outbreak risks in their region or predict patient deterioration due to viral infections so that proactive measures can be implemented.

Real-time integration helps with models being continuously updated and reflecting the most current trends and patterns in medical virology. This is particularly important for rapidly evolving viruses, where timely information can make a significant difference in patient outcomes and public health responses. Also, incorporating the framework into CDSS facilitates seamless workflow integration for clinicians and provides decision support within their existing EHR systems. This synergy between federated learning and clinical practice accelerates the translation of "data-driven" insights into tangible clinical actions.

While my proposed framework is designed for medical virology, its principles and architecture are highly adaptable to other medical domains such as oncology or cardiology. In oncology, for example, federated learning can help collaborative training of models for cancer detection or treatment response prediction using imaging data from multiple hospitals, all while preserving patient privacy Similarly, in cardiology, institutions can jointly develop models to predict cardiac events or analyze electrocardiogram patterns without sharing sensitive patient data.

Applying the framework across different domains increases its impact on precision medicine by facilitating the development of robust, generalized models that benefit from diverse datasets. It addresses common challenges in healthcare data analysis, such as data scarcity and heterogeneity, by pooling knowledge without compromising confidentiality. Moreover, cross-domain applications promote interdisciplinary collaboration and innovation and have the potential to bring about breakthroughs in disease understanding, diagnosis, and treatment across various fields of medicine.

## Note: Use of Generative AI tools

During the preparation of this work, the author used Anthropic's Claude AI in order to check for spelling and grammar errors in the write-up and make corrections. After using this tool/service, the author reviewed and edited the content as needed and take full responsibility for the content of the publication.

## References

Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016, October). Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security* (pp. 308-318). https://doi.org/10.1145/2976749.2978318

AbdulRahman, S., Tout, H., Ould-Slimane, H., Mourad, A., Talhi, C., & Guizani, M. (2020). A survey on federated learning: The journey from centralized to distributed on-site learning and beyond. *IEEE Internet of Things Journal*, *8*(7), 5476-5497. https://doi.org/10.1109/JIOT.2020.3030072

Ahmed, F., Kang, I. S., Kim, K. H., Asif, A., Rahim, C. S. A., Samantasinghar, A., ... & Choi, K. H. (2023). Drug repurposing for viral cancers: A paradigm of machine learning, deep learning, and virtual screening-based approaches. *Journal of Medical Virology*, *95*(4), e28693. https://doi.org/10.1002/jmv.28693

Androulaki, E., Barger, A., Bortnikov, V., Cachin, C., Christidis, K., De Caro, A., ... & Yellick, J. (2018, April). Hyperledger fabric: a distributed operating system for permissioned blockchains. In *Proceedings of the thirteenth EuroSys conference* (pp. 1-15). https://doi.org/10.1145/3190508.3190538

Angraal, S., Krumholz, H. M., & Schulz, W. L. (2017). Blockchain technology: applications in health care. *Circulation: Cardiovascular quality and outcomes*, *10*(9), e003800. https://doi.org/10.1161/CIRCOUTCOMES.117.003800

Antunes, R. S., André da Costa, C., Küderle, A., Yari, I. A., & Eskofier, B. (2022). Federated learning for healthcare: Systematic review and architecture proposal. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *13*(4), 1-23. https://doi.org/10.1145/3501813

Arabahmadi, M., Farahbakhsh, R., & Rezazadeh, J. (2022). Deep learning for smart Healthcare—A survey on brain tumor detection from medical imaging. *Sensors*, *22*(5), 1960. https://doi.org/10.3390/s22051960

Attaran, M. (2022). Blockchain technology in healthcare: Challenges and opportunities. *International Journal of Healthcare Management*, *15*(1), 70-83. https://doi.org/10.1080/20479700.2020.1843887

Banabilah, S., Aloqaily, M., Alsayed, E., Malik, N., & Jararweh, Y. (2022). Federated learning review: Fundamentals, enabling technologies, and future applications. *Information processing & management*, *59*(6), 103061. https://doi.org/10.1016/j.ipm.2022.103061

Briggs, C., Fan, Z., & Andras, P. (2020, July). Federated learning with hierarchical clustering of local updates to improve training on non-IID data. In *2020 international joint conference on neural networks (IJCNN)* (pp. 1-9). IEEE. https://doi.org/10.1109/IJCNN48605.2020.9207469

Collins, L., Hassani, H., Mokhtari, A., & Shakkottai, S. (2022). Fedavg with fine tuning: Local updates lead to representation learning. *Advances in Neural Information Processing Systems, 35,* 10572-10586. Retrieved from Link.

Coretti, S., Kiayias, A., Moore, C., & Russell, A. (2022, November). The Generals' Scuttlebutt: Byzantine-Resilient Gossip Protocols. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security* (pp. 595-608). https://doi.org/10.1145/3548606.3560638

Dabbagh, M., Choo, K. K. R., Beheshti, A., Tahir, M., & Safa, N. S. (2021). A survey of empirical performance evaluation of permissioned blockchain platforms: Challenges and opportunities. *computers & security*, *100*, 102078. https://doi.org/10.1016/j.cose.2020.102078

Desai, H. (2023a). Infusing Blockchain in accounting curricula and practice: expectations, challenges, and strategies. *The International Journal of Digital Accounting Research, 23,* pp. 97-135. https://doi.org/10.4192/1577-8517-v23_5

Desai, H. (2023b). An Auditor's Perspective on Smart Contracts and DAOs. *The CPA Journal*, *93*(7/8), 60-63.

Desai, H. (2023c). Crypto Record-keeping Technologies for Tax Professionals: Developments, Challenges, and Ethical Considerations. *Bloomberg BNA Tax Management Memorandum, 63(12)*. http://dx.doi.org/10.13140/RG.2.2.15151.61600/1

Díaz, J. S. P., & García, Á. L. (2023). Study of the performance and scalability of federated learning for medical imaging with intermittent clients. *Neurocomputing*, *518*, 142-154. https://doi.org/10.1016/j.neucom.2022.11.011

Ettaloui, N., Arezki, S., & Gadi, T. (2023, November). An Overview of blockchain-based electronic health record and compliance with GDPR and HIPAA. In The International Conference on Artificial Intelligence and Smart Environment (pp. 405-412). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-48573-2_58

Fang, Y., Li, K., Zheng, R., Liao, S., & Wang, Y. (2021). A communication-channel-based method for detecting deeply camouflaged malicious traffic. *Computer Networks*, *197*, 108297. https://doi.org/10.1016/j.comnet.2021.108297

Fekri, M. N., Grolinger, K., & Mir, S. (2022). Distributed load forecasting using smart meter data: Federated learning with Recurrent Neural Networks. *International Journal of Electrical Power & Energy Systems*, *137*, 107669. https://doi.org/10.1016/j.ijepes.2021.107669

Fereidooni, H., Marchal, S., Miettinen, M., Mirhoseini, A., Möllering, H., Nguyen, T. D., ... & Zeitouni, S. (2021, May). SAFELearn: Secure aggregation for private federated learning. In *2021 IEEE Security and Privacy Workshops (SPW)* (pp. 56-62). IEEE. https://doi.org/10.1109/SPW53761.2021.00017

Gouissem, A., Abualsaud, K., Yaacoub, E., Khattab, T., & Guizani, M. (2023). Collaborative byzantine resilient federated learning. *IEEE Internet of Things Journal*, *10*(18), 15887-15899. https://doi.org/10.1109/JIOT.2023.3266347

Hanzely, F., & Richtárik, P. (2020). Federated learning of a mixture of global and local models. *arXiv preprint arXiv:2002.05516*. https://doi.org/10.48550/arXiv.2002.05516

Hasselgren, A., Wan, P. K., Horn, M., Kralevska, K., Gligoroski, D., & Faxvaag, A. (2020). GDPR compliance for blockchain applications in healthcare. *arXiv preprint arXiv:2009.12913*. https://doi.org/10.48550/arXiv.2009.12913

Kalapaaking, A. P., Stephanie, V., Khalil, I., Atiquzzaman, M., Yi, X., & Almashor, M. (2022). Smpc-based federated learning for 6g-enabled internet of medical things. *IEEE Network*, *36*(4), 182-189. https://doi.org/10.1109/MNET.007.2100717

LeCun Y, Bengio Y, Hinton G. *Deep learning.* Nature. 2015 May 28;521(7553):436-44. https://doi.org/10.1038/nature14539

Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. *IEEE signal processing magazine*, *37*(3), 50-60. https://doi.org/10.1109/MSP.2020.2975749

Liu-Wei, W., Kafkas, Ş., Chen, J., Dimonaco, N. J., Tegnér, J., & Hoehndorf, R. (2021). DeepViral: prediction of novel virus–host interactions from protein sequences and infectious disease phenotypes. *Bioinformatics*, *37*(17), 2722-2729. https://doi.org/10.1093/bioinformatics/btab147

Liu, L., Zhang, J., Song, S., & Letaief, K. B. (2022). Hierarchical federated learning with quantization: Convergence analysis and system design. *IEEE Transactions on Wireless Communications*, *22*(1), 2-18. https://doi.org/10.1109/TWC.2022.3190512

Liu, X., Xie, L., Wang, Y., Zou, J., Xiong, J., Ying, Z., & Vasilakos, A. V. (2020a). Privacy and security issues in deep learning: A survey. *IEEE Access*, *9*, 4566-4593. https://doi.org/10.1109/ACCESS.2020.3045078

Liu, Y., Garg, S., Nie, J., Zhang, Y., Xiong, Z., Kang, J., & Hossain, M. S. (2020b). Deep anomaly detection for time-series data in industrial IoT: A communication-efficient on-device federated learning approach. *IEEE Internet of Things Journal*, *8*(8), 6348-6358. https://doi.org/10.1109/JIOT.2020.3011726

Mao, Y., Zhao, Z., Yan, G., Liu, Y., Lan, T., Song, L., & Ding, W. (2022). Communication-efficient federated learning with adaptive quantization. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *13*(4), 1-26. https://doi.org/10.1145/3510587

McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017, April). Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics* (pp. 1273-1282). PMLR. Retrieved from https://proceedings.mlr.press/v54/mcmahan17a/mcmahan17a.pdf

Mendieta, M., Yang, T., Wang, P., Lee, M., Ding, Z., & Chen, C. (2022). Local learning matters: Rethinking data heterogeneity in federated learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 8397-8406). https://doi.org/10.48550/arXiv.2111.14213

Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: review, opportunities and challenges. *Briefings in bioinformatics*, *19*(6), 1236-1246. https://doi.org/10.1093/bib/bbx044

Mothukuri, V., Parizi, R. M., Pouriyeh, S., Huang, Y., Dehghantanha, A., & Srivastava, G. (2021). A survey on security and privacy of federated learning. *Future Generation Computer Systems*, *115*, 619-640. https://doi.org/10.1016/j.future.2020.10.007

Ng, D., Lan, X., Yao, M. M. S., Chan, W. P., & Feng, M. (2021). Federated learning: a collaborative effort to achieve better medical imaging models for individual sites that have small labelled datasets. *Quantitative Imaging in Medicine and Surgery*, *11*(2), 852. https://doi.org/10.21037%2Fqims-20-595

Ozfatura, E., Ozfatura, K., & Gündüz, D. (2021, July). Time-correlated sparsification for communication-efficient federated learning. In *2021 IEEE International Symposium on Information Theory (ISIT)* (pp. 461-466). IEEE. https://doi.org/10.1109/ISIT45174.2021.9518221

Patel, V. A., Bhattacharya, P., Tanwar, S., Gupta, R., Sharma, G., Bokoro, P. N., & Sharma, R. (2022). Adoption of federated learning for healthcare informatics: Emerging applications and future directions. *IEEE Access*, *10*, 90792-90826. https://doi.org/10.1109/ACCESS.2022.3201876

Pfitzner, B., Steckhan, N., & Arnrich, B. (2021). Federated learning in a medical context: a systematic literature review. *ACM Transactions on Internet Technology (TOIT)*, *21*(2), 1-31. https://doi.org/10.1145/3412357

Qu, L., Zhou, Y., Liang, P. P., Xia, Y., Wang, F., Adeli, E., ... & Rubin, D. (2022). Rethinking architecture design for tackling data heterogeneity in federated learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10061-10071). https://doi.org/10.48550/arXiv.2106.06047

Reisizadeh, A., Mokhtari, A., Hassani, H., Jadbabaie, A., & Pedarsani, R. (2020, June). Fedpaq: A communication-efficient federated learning method with periodic averaging and quantization. In *International conference on artificial intelligence and statistics* (pp. 2021-2031). PMLR. Retrieved from https://proceedings.mlr.press/v108/reisizadeh20a/reisizadeh20a.pdf

Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., ... & Cardoso, M. J. (2020). The future of digital health with federated learning. *NPJ digital medicine*, *3*(1), 1-7. https://doi.org/10.1038/s41746-020-00323-1

Saeed, H., Malik, H., Bashir, U., Ahmad, A., Riaz, S., Ilyas, M., ... & Khan, M. I. A. (2022). Blockchain technology in healthcare: A systematic review. *Plos one*, *17*(4), e0266462. https://doi.org/10.1371/journal.pone.0266462

Saidani, A., Jianwen, X., & Mansouri, D. (2020). A lossless compression approach based on delta encoding and T-RLE in WSNs. *Wireless Communications and Mobile Computing*, *2020*(1), 8824954. https://doi.org/10.1155/2020/8824954

Schreiber, R., Koppel, R., & Kaplan, B. (2024). What Do We Mean by Sharing of Patient Data? DaSH-A Data Sharing Hierarchy of Privacy and Ethical Challenges. *Applied Clinical Informatics*. https://doi.org/10.1055/a-2373-3291

Shamshirband, S., Fathi, M., Dehzangi, A., Chronopoulos, A. T., & Alinejad-Rokny, H. (2021). A review on deep learning approaches in healthcare systems: Taxonomies, challenges, and open issues. *Journal of Biomedical Informatics, 113*, 103627. https://doi.org/10.1016/j.jbi.2020.103627

So, J., Güler, B., & Avestimehr, A. S. (2020). Byzantine-resilient secure federated learning. *IEEE Journal on Selected Areas in Communications*, *39*(7), 2168-2181. https://doi.org/10.1109/JSAC.2020.3041404

Sookhak, M., Jabbarpour, M. R., Safa, N. S., & Yu, F. R. (2021). Blockchain and smart contract for access control in healthcare: A survey, issues and challenges, and open issues. *Journal of Network and Computer Applications, 178,* 102950. https://doi.org/10.1016/j.jnca.2020.102950

Su, Y., Li, J., Li, Y., & Su, Z. (2022). Edge-enabled: a scalable and decentralized data aggregation scheme for IoT. *IEEE Transactions on Industrial Informatics*, *19*(2), 1854-1862. https://doi.org/10.1109/TII.2022.3170156

Thwal, C. M., Thar, K., Tun, Y. L., & Hong, C. S. (2021, January). Attention on personalized clinical decision support system: Federated learning approach. In *2021 IEEE International conference on big data and smart computing (BigComp)* (pp. 141-147). IEEE. https://doi.org/10.1109/BigComp51126.2021.00035

Umer, K., Huang, Q., Khorasany, M., Afzal, M., & Amin, W. (2021). A novel communication efficient peer-to-peer energy trading scheme for enhanced privacy in microgrids. *Applied Energy*, *296*, 117075. https://doi.org/10.1016/j.apenergy.2021.117075

Wang, H., Guo, S., Qu, Z., Li, R., & Liu, Z. (2021). Error-compensated sparsification for communication-efficient decentralized training in edge environment. *IEEE Transactions on Parallel and Distributed Systems*, *33*(1), 14-25. https://doi.org/10.1109/TPDS.2021.3084104

Wang, Z., Wen, M., Xu, Y., Zhou, Y., Wang, J. H., & Zhang, L. (2023). Communication compression techniques in distributed deep learning: A survey. *Journal of Systems Architecture*, *142*, 102927. https://doi.org/10.1016/j.sysarc.2023.102927

Xia, G., Chen, J., Yu, C., & Ma, J. (2023). Poisoning attacks in federated learning: A survey. *IEEE Access*, *11*, 10708-10722. https://doi.org/10.1109/ACCESS.2023.3238823

Ye, M., Fang, X., Du, B., Yuen, P. C., & Tao, D. (2023). Heterogeneous federated learning: State-of-the-art and research challenges. *ACM Computing Surveys*, *56*(3), 1-44. https://doi.org/10.1145/3625558

Yu, S., & Cui, L. (2022). Secure multi-party computation in federated learning. In *Security and Privacy in Federated Learning* (pp. 89-98). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-19-8692-5_6

Zhang, J., Zhu, H., Wang, F., Zhao, J., Xu, Q., & Li, H. (2022). Security and privacy threats to federated learning: Issues, methods, and challenges. *Security and Communication Networks*, *2022*(1), 2886795. https://doi.org/10.1155/2022/2886795

Zhang, M., Wei, E., & Berry, R. (2021). Faithful edge federated learning: Scalability and privacy. IEEE Journal on Selected Areas in Communications, 39(12), 3790-3804. https://doi.org/10.1109/JSAC.2021.3118423

Zheng, R., Liu, J., Li, K., Liao, S., & Liu, L. (2020, August). Detecting malicious tls network traffic based on communication channel features. In *2020 IEEE 8th International Conference on Information, Communication and Networks (ICICN)* (pp. 14-19). IEEE. https://doi.org/10.1109/ICICN51133.2020.9205087

Zhong, Z., Zhou, Y., Wu, D., Chen, X., Chen, M., Li, C., & Sheng, Q. Z. (2021, May). P-FedAvg: Parallelizing federated learning with theoretical guarantees. In *IEEE INFOCOM 2021-IEEE Conference on Computer Communications* (pp. 1-10). IEEE. https://doi.org/10.1109/INFOCOM42981.2021.9488877

Zhu, H. (2020). On the relationship between (secure) multi-party computation and (secure) federated learning. *arXiv preprint arXiv:2008.02609*. https://doi.org/10.48550/arXiv.2008.02609

Ziller, A., Usynin, D., Braren, R., Makowski, M., Rueckert, D., & Kaissis, G. (2021). Medical imaging deep learning with differential privacy. *Scientific Reports*, *11*(1), 13524. https://doi.org/10.1038/s41598-021-93030-0