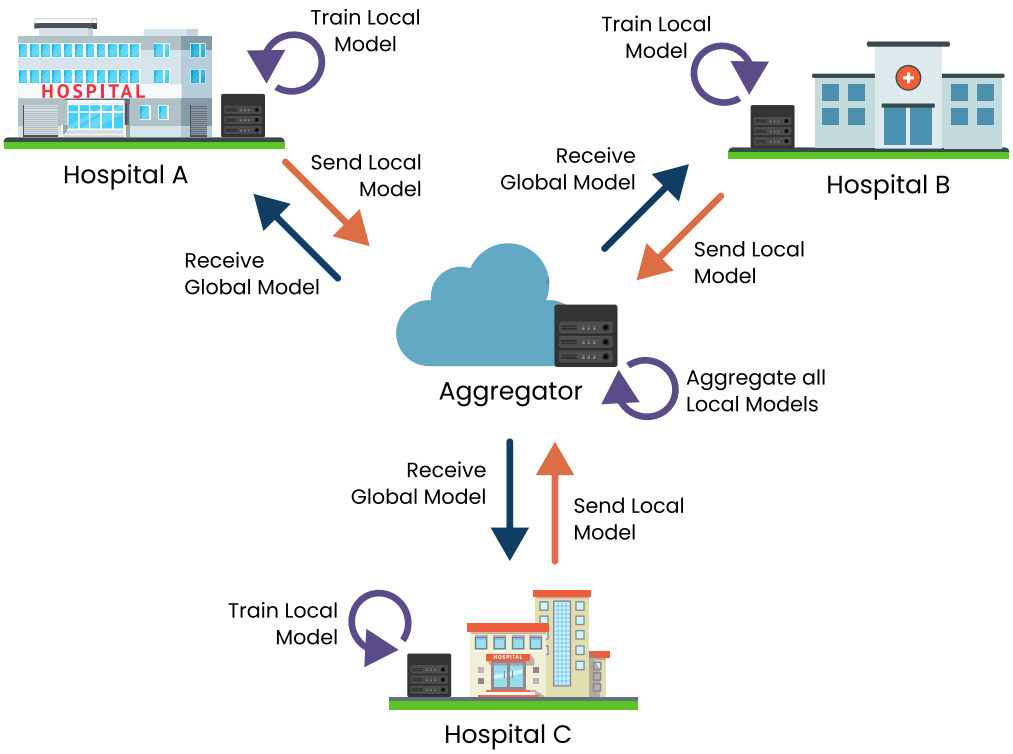
hainet

Privacy-Preserved Learning in Health AI Through Federated Learning (PriFed)



Health AI for All Network
www.hainet.org

Privacy-Preserved Learning in Health AI Through Federated Learning

Binod Bhattarai[1,2,3,5]

Bibek Niroula[2]

Aavash Chhetri[2]

Kiran Raj Pandey[1,2]

Yash Raj Shrestha[4]

Niyoj Oli[1,2]

[1] Health AI for All Network

[2] Nepal Applied Mathematics and Informatics Institute for Research

[3] University of Aberdeen

[4] University of Lausanne

[5] University College London

Acknowledgment

The whitepaper was officially released at the PriFed Symposium 2026[a] The authors gratefully acknowledge the support of the Swiss State Secretariat for Education, Research and Innovation (SERI) through the Leading House South Asia at ZHAW School of Engineering, Switzerland. This work was carried out under the Connect & Collaborate Grant (2025–2026), awarded for the project titled *"Swiss–South Asia Symposium on Secure and Equitable Health AI for Resource-Limited Environments."*

Version: Final

[a]Binod Bhattarai serves as Chair of the PriFed Symposium 2026, with Kiran Raj Pandey and Yash Raj Shrestha as Co-chairs. Bibek Niroula, Avash Chhetri, and Niyoj Oli serve as Committee Members.

Contents

Executive Summary

Healthcare systems in resource limited settings face persistent challenges, including shortages of skilled professionals, limited access to diagnostic services, and fragmented healthcare infrastructure. Artificial intelligence (AI) has emerged as a promising tool to address these gaps by augmenting clinical decision-making, improving diagnostic accuracy, and enabling scalable healthcare delivery. However, the development of effective AI systems relies heavily on access to large, diverse, and high-quality datasets.

Traditional approaches to training AI models in healthcare follow a centralized paradigm, where data from multiple institutions is aggregated into a single repository. While this approach has driven early successes, it is fundamentally misaligned with the realities of healthcare systems. Strict data privacy regulations, institutional data silos, security risks, and infrastructural constraints make large-scale data pooling difficult, particularly in resource-constrained environments.

Federated Learning offers a compelling alternative. By enabling models to be trained collaboratively across multiple institutions without requiring patient data to leave local environments, Federated Learning addresses key limitations of centralized approaches. It allows healthcare providers to contribute to shared AI systems while preserving data privacy and maintaining institutional control. Despite its promise, Federated Learning introduces its own set of technical, institutional, and operational challenges, including data heterogeneity, communication overhead, governance complexities, and coordination across stakeholders. These challenges are further compounded in resource limited settings by limited connectivity, infrastructure gaps, and shortages of technical expertise. Nevertheless, Federated Learning is uniquely well-suited to the needs of healthcare systems in resource limited settings, where data is inherently distributed and centralized infrastructure is often impractical. By enabling privacy-preserving collaboration, Federated Learning provides a pathway for institutions to collectively develop robust and generalizable AI models.

This whitepaper examines the limitations of centralized AI development, introduces **Pri**vacy-preserved learning in Health AI through **Fed**erated Learning (PriFed) , and analyzes its relevance for healthcare systems in resource limited settings. It further outlines key challenges to adoption and provides strategic recommendations for enabling PriFed ecosystems. Central to this effort is the role of collaborative networks, such as the Health AI for All Network (HAINet), in fostering partnerships, building capacity, and advancing equitable access to Health AI solutions.

Introduction

Healthcare Challenges in the Resource Limited Settings

Healthcare systems across the resource limited settings operate under a unique set of systemic pressures. The most acute barrier is the severe shortage of specialized medical professionals such as radiologists, pathologists, and cardiologists. This scarcity often disproportionately affects rural and marginalized populations, forcing patients to endure long travel times and delayed diagnoses for critical conditions. For example, Africa carries an estimated 25 percent of the world's disease burden but accounts for only 3 percent of the global healthcare workforce.[1] Public hospitals and clinics often suffer from overcrowding, outdated equipment, and medical supply shortages due to insufficient funding. Furthermore, medical infrastructure is often highly inconsistent. Many facilities operate with a fragmented mix of legacy and modern equipment, leading to healthcare data that are frequently incomplete or missing entire diagnostic modalities (e.g., a rural clinic may have basic X-ray capability but lack MRI or advanced pathology tools).

Figure 1: Healthcare disparity in resource limited settings

Role of AI in Addressing Challenges in Healthcare

Artificial Intelligence (AI)–assisted interventions offer an opportunity to address these systemic constraints and bridge the gap in healthcare quality between advanced and resource-limited settings. AI can enable effective task shifting by delegating certain diagnostic and decision-support functions to non-specialist healthcare workers such as nurses or community health workers, to address workforce shortages and improve access to care. These systems can support semi-trained healthcare workers

[1] "COVID-19 Management: Curriculum for Community Health Workers," Africa CDC, November 2020.

in making accurate initial triage decisions without requiring a specialist on site. AI-backed point-of-care (POC) diagnosis solutions have demonstrated real-world positive impact by enabling rapid, accurate assessments in resource-limited settings, often using smartphones or low-cost devices. In Malawi, AI-assisted fetal monitoring software on ultrasound devices reduced stillbirths by 82%.[2] AI tools like Qure.ai's qXR, deployed in India, analyze chest X-rays at rural clinics, referring fewer false positives to specialists and scaling to millions screened amid specialist shortages.[3] Complementary to AI-assisted diagnostics, telemedicine platforms enable clinical consultation and treatment from a distance. Telemedicine platforms like India's eSanjeevani and Rwanda's Babyl provide remote consultations, cutting travel costs and serving millions in rural areas with limited facilities. In Sub-Saharan Africa, where 70% of the world's maternal deaths happen, m-mama service connects pregnant women and newborns with volunteer drivers for transportation to the nearest healthcare facility. This AI-driven emergency referral system is believed to have led to a 38% and over 40%[4] reduction in maternal deaths and newborn deaths, respectively. Advanced machine learning frameworks can also be designed to infer clinically meaningful insights even when certain diagnostic modalities are unavailable. This capability is particularly important in resource-constrained environments, where healthcare facilities may only be able to collect a subset of the diagnostic signals typically used in well-equipped hospitals.

Taken together, these examples illustrate that AI-assisted healthcare, beyond theoretical promise, is an increasingly demonstrated reality across the resource limited settings. As the field matures, the central challenge will shift from proving feasibility to ensuring equitable deployment. Realizing this potential will demand sustained collaboration between AI developers, local health systems, governments, and global health institutions to build solutions that are not only technically capable but contextually grounded and locally owned.

---

[2]Caroline Kimeu, "How AI monitoring is cutting stillbirths and neonatal deaths in a clinic in Malawi," The Guardian, 6 December 2024.
[3]Qure.ai, "AI makes TB screening more accessible in South Africa," Impact Story, 2024, https://www.qure.ai/impact-stories/ai-makes-tb-screening-more-accessible-in-south-africa.
[4]Vodafone Foundation, "Lifesaving maternal and newborn transport system 'm-mama' expands to Malawi," 20 September 2023, https://foundation.vodafone.com/news/vodafone-foundation/lifesaving-maternal-newborn-transport-system-m-mama-expands-malawi.

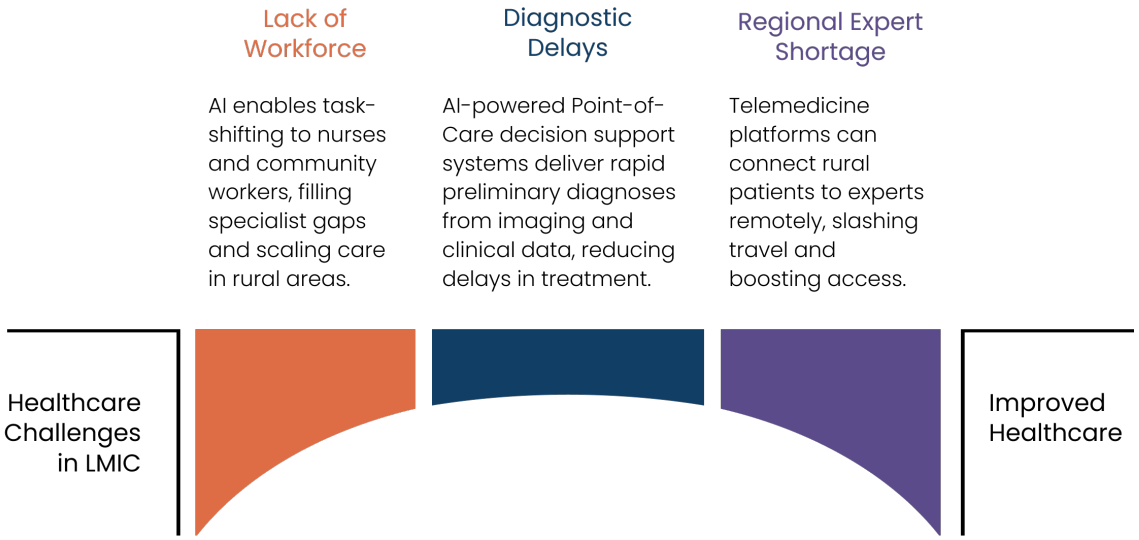


Figure 2: Bridging the gaps in health care in resource-limited settings with AI

Health Artificial Intelligence

Deep Learning, a subfield of machine learning (ML), has seen a dramatic resurgence in past years, largely driven by increases in computational power and the availability of massive new datasets. The integration of deep learning into healthcare has moved rapidly from proof-of-concept to clinical implementation, largely driven by the ability of deep neural networks to extract high-dimensional patterns from complex biomedical data.

Evidence from Transformative Deep Learning Applications

Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) excel in medical imaging by detecting subtle anomalies with expert-level accuracy. A seminal 2017 paper demonstrated that a CNN trained on 129,450 dermoscopic images achieved dermatologist-level performance (AUC 0.91 for melanoma), enabling scalable skin lesion triage.[5] In Radiology, CheXNet [6] used DenseNet-121 model trained on the ChestX-ray14 dataset surpassomg average radiologist performance in pneumonia detection (AUC 0.768 vs. 0.732), and it is now deployed in clinical pliots for triage. Deep learning models have also demonstrated the ability to infer

[5]Andre Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, 2017.
[6]Pranav Rajpurkar et al., "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning," 2017.

genetic mutations directly from histopathology images. A deep residual CNN (Inception v3) applied to lung adenocarcinoma slides predicted clinically relevant mutations such as EGFR and TP53 with AUC values between 0.85 and 0.95 using only H&E-stained images.[7]

Natural language processing (NLP) models extract insights from unstructured electronic health records (EHRs), enabling automated coding, risk prediction, and clinical decision support. Transformer-based language models have significantly improved performance on clinical text tasks. A domain-adapted BERT model, BioBert trained on large-scale clinical notes demonstrated improvements across tasks such as de-identification, mortality prediction, and hospital readmission forecasting.[8] Deep learning methods applied to electronic health records have also demonstrated strong performance in clinical coding and phenotyping tasks. Recurrent neural networks such as LSTM and GRU models applied to the MIMIC-III dataset have achieved accuracies approaching 95% for automated ICD coding and phenotyping tasks such as obesity detection (F1 = 0.92), substantially reducing the need for manual chart review.[9]

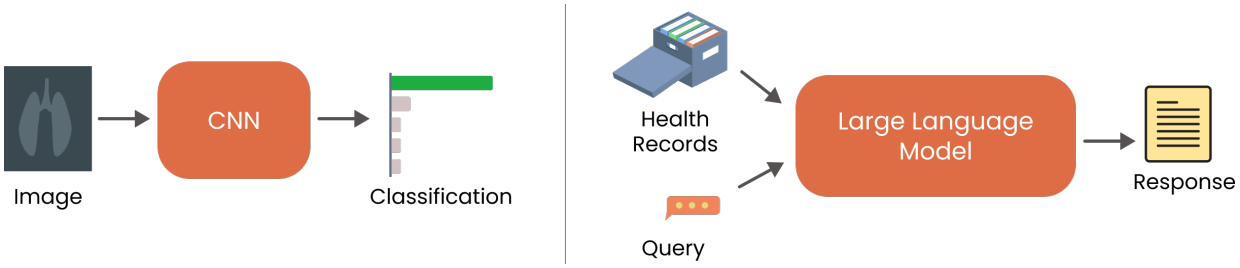


Figure 3: Use of AI models for medical image classification (left) and use of LLMs to query over large EHR (right)

Clinicians naturally combine diverse diagnostic inputs including imaging, laboratory data, omics, and pathology reports, to make holistic clinical decisions. This has inspired the development of multimodal AI models designed to mirror such integrative reasoning. A bimodal deep learning framework combining CNN-based image features from chest radiographs with LSTM representations of clinical text improved mortality prediction performance on the MIMIC-CXR dataset (377,000 image–report pairs), achieving an AUC of 0.85 compared with 0.78 for unimodal models.[10] PneumoFusion-Net

[7]Nicolas Coudray et al., "Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning," Nature Medicine, 2018.
[8]Emily Alsentzer et al., "Publicly Available Clinical BERT Embeddings," Proceedings of the Clinical Natural Language Processing Workshop, 2019.
[9]C. Xiao et al., "Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record," IEEE Journal of Biomedical and Health Informatics, 2018.
[10]Scott Huang et al., "Fusion of Medical Imaging and Electronic Health Records using Bi-modal

integrates CT images, clinical text, numerical lab test results, and radiology reports, achieving diagnostic accuracy approaching 96% across more than 10,000 clinical cases and outperforming single-modality approaches by a substantial margin of 15%.[11]

Training AI Models in Healthcare

Traditional Paradigm: Centralized Model Training

The remarkable success of the deep learning applications mentioned above is traditionally predicated on one fundamental requirement: massive amounts of centralized data. Deep learning models are highly data-hungry, often requiring thousands to millions of annotated examples. Large datasets improve statistical reliability, while diversity ensures generalization across patient demographics, clinical practices, and data acquisition protocols. The standard lifecycle for developing Health AI models typically follows a centralized pipeline, where data from multiple healthcare sources is collected and used to train a model in a single location.

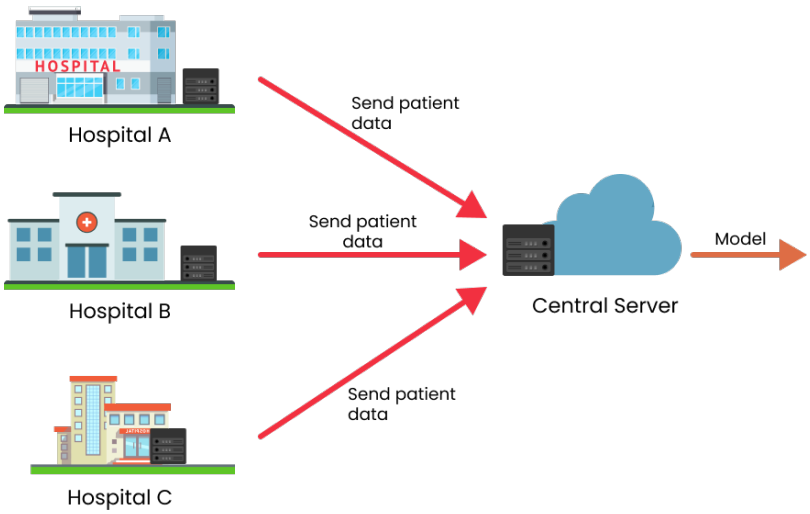


Figure 4: Centralized learning in Health AI. Patient data is aggregated into a central repository to train a single shared model.

1. *Data Aggregation and Pooling.* The process would begin with collecting patient

Deep Learning," 2020.
[11]Yujie Wang, et al., "PneumoFusion-Net: A Multi-modal Deep Learning Solution for Precise Pneumonia Diagnosis," 2025.

data from various healthcare sources, such as hospitals, diagnostic centres, or medical devices. This may include medical images, EHRs, laboratory test results, and other clinical information. This data is transferred over networks to a central repository managed by the organization developing the AI model.

2. *Standardization and Preprocessing.* This stage may involve aligning data formats, removing incomplete records, correcting inconsistencies, and organizing the data into a unified training dataset.

3. *Centralized Model Training.* The AI model is then trained using this centralized dataset on powerful computing infrastructure. During training, the model repeatedly analyzes the data, learns patterns and gradually improves its predictions. Because the model has access to all available data in one location, this process is relatively straightforward and efficient from a technical perspective.

4. *Deployment.* After the model has achieved satisfactory performance during testing and validation, it is finalized and deployed for real-world use.

Limitations of Centralized AI Development

While centralized training has driven early successes in Health AI, applying this paradigm to real-world healthcare environments reveals several structural limitations. The centralized approach assumes that clinical data can be freely aggregated and pooled into a single repository for model development. In practice, this assumption rarely holds in healthcare systems. Legal restrictions, institutional policies, and infrastructural limitations often prevent large-scale data consolidation. These challenges are particularly pronounced in resource-constrained or highly fragmented healthcare systems.

Privacy and Regulatory Constraints. The most immediate barrier to centralized AI is the stringent regulatory landscape governing protected health information (PHI). Regulatory frameworks such as HIPAA in the United States and the GDPR in Europe impose strict controls on how patient data can be stored, transferred, and used. Legally aggregating large volumes of sensitive records into a single, centralized repository requires navigating complex compliance requirements. For many institutions, the legal friction and immense liability associated with exporting patient data off-site significantly limits their ability to participate in collaborative AI development.

Institutional Data Silos in Healthcare. Healthcare data is inherently fragmented across institutional boundaries. Hospitals, research centers, and clinical networks

often view their proprietary datasets as valuable assets. As a result, organizations may be reluctant to share patient data with external developers or competing healthcare systems. This lack of interoperability and limited data sharing leads to models being trained on relatively narrow and localized datasets. Consequently, clinical insights remain confined within individual institutions, and models trained under such conditions may perform poorly when deployed across different healthcare settings or patient populations.

Security Risks. Centralized model development requires the aggregation of large volumes of sensitive patient information in a single location. This architecture can create a high-value target for cyberattacks. A breach of a centralized server could potentially expose an entire training dataset, including highly sensitive medical records. In addition, data may be vulnerable during the transfer and consolidation process. IBM's 2024 Cost of a Data Breach Report[12] confirms healthcare data breaches averaged $9.77 million per incident—the highest across industries for the 14th straight year.

Infrastructure and Computational Barriers. Training robust Health AI models demands for formidable computational resources. In centralized architectures, these requirements are concentrated within a single computing environment, typically requiring high-performance GPU clusters and large-scale storage infrastructure. Furthermore, transferring large healthcare datasets—such as high-resolution medical images or volumetric MRI scans—from distributed clinical sites to a central server can demand significant network bandwidth. For many healthcare institutions, particularly in the resource limited settings, the infrastructure required to support such large-scale data movement and secure storage is either limited or prohibitively expensive. These practical constraints further complicate the adoption of centralized AI development pipelines.

The Need for Cross-Institution Collaboration

Achieving robust and equitable Health AI systems requires moving beyond isolated, institution-specific development. Although individual healthcare institutions may possess substantial volumes of clinical data, these datasets are inherently limited in scope due to demographic, geographic, and procedural constraints. As a result,

[12]Doug Bonderud, "Ransomware on the rise: Healthcare industry attack trends 2024," IBM, 2024.

models trained in isolation often fail to generalize effectively across diverse clinical environments. Cross-institutional collaboration provides a mechanism to aggregate knowledge across heterogenous data sources, enabling the development of models that better reflect real-world variability.

Overcoming the Small Data Problem. As we have established before, Deep Learning models rely heavily on large and diverse datasets to achieve strong generalization performance. While large hospitals may process significant patient volumes, the frequency of rare conditions, such as specific pediatric cancers or complex fracture patterns, may remain low within any single institutions. Consequently, locally trained models are prone to learning institution-specific artifacts rather than generalizable, clinically relevant features. Cross-institution collaboration increases both the scale and diversity of training data. By incorporating data from multiple sources, models are exposed to a broader distribution of cases, improving their ability to generalize across different clinical settings and patient populations.

Addressing Regional Heterogeneity and Geographic Bias. Clinical data distributions vary significantly across regions due to differences in genetics, environmental conditions, socioeconomic factors, and healthcare practices. Models trained exclusively on a data from a specific geographic region, particularly in high-resource settings, may exhibit degraded performance when deployed in underrepresented regions. This issue is especially critical for conditions with region-specific manifestations (e.g. tropical infectious diseases). Cross-institution collaboration, preferably across borders, ensures that models are trained on geographically diverse datasets, reducing the risk of algorithmic bias. This would result in models with higher diagnostic reliability across varied deployment contexts.

Collective Intelligence for Outbreak Response. Timely access to distributed clinical data is essential for effective response to emerging public health threats. Traditional research and publication pipelines introduce delays that are incompatible with the rapid progression of infectious disease outbreaks. A collaborative framework enables near real-time integration of clinical observations into shared models. This allows institutions in unaffected or early-stage regions to deploy diagnostic systems that have already been trained on relevant patterns, thereby reducing response latency. For example, during the early phase of COVID-19 outbreak in Wuhan, the hospitals reported characteristic chest CT patterns (e.g., ground-glass opacities). If a collaborative framework would have been in place, these institutions could contribute model updates capturing these features. Hospitals in regions such as Nepal or Ethiopia, prior to experiencing widespread transmission, ccould then deploy models already trained to recognize early indicators of the disease. Such a collaborative

approach can significantly enhance early detection and containment efforts during emerging epidemics.

Distributed Computational Requirements. Training large-scale deep learning models demands substantial GPU resources, memory, and storage infrastructure. For many institutions, particularly smaller hospitals or those in resource-constrained settings, such investments are economically and logistically prohibitive. Cross-institution collaboration provides a pathway to distribute these computational burdens; by enabling decentralized training paradigms, institutions can contribute local compute resources toward a shared objective, reducing the need for centralized, high-cost infrastructure. This is especially consequential in resource limited settings, where computational infrastructure is often severely limited. Lowering the barrier to participation in this way promotes more inclusive development of Health AI systems and ensures that advancements are not confined to well-funded organizations.

Despite the clear benefits of collaboration, the sharing of medical data is constrained by strict privacy, legal, and ethical requirements. This necessitates the development of new collaborative paradigms that enable knowledge sharing without requiring direct data exchange.

Privacy-Preserved Learning in Health AI through Federated Learning

Federated Learning (FL) [13] [14] addresses challenges for traditional centralized training and enables collaboration by fundamentally re-imagining the training lifecycle. While traditional AI development requires all data to be moved to a single location, Federated Learning operates on a simple but powerful shift: instead of bringing the data to the model, we bring the model to the data. This approach allows multiple hospitals to collaborate on a "Global Model" without ever sharing their patients' raw information. Federated Learning, hence, ensures that collaboration does not come at a cost of privacy.

[13] Kaissis G.A., et al., "Secure, privacy-preserving and federated machine learning in medical imaging", Nature Machine Intelligence, 2020.
[14] Jie Xu, et al., "Federated Learning for healthcare informatics." Journal of Healthcare Informatics Research, 2021.

Training Procedure of Federated Learning

Federated Learning follows an iterative, coordinated training process involving a central server and multiple distributed clients (data owners). The objective is to train a global model by utilizing decentralized data while ensuring that raw data never leaves local institutions.

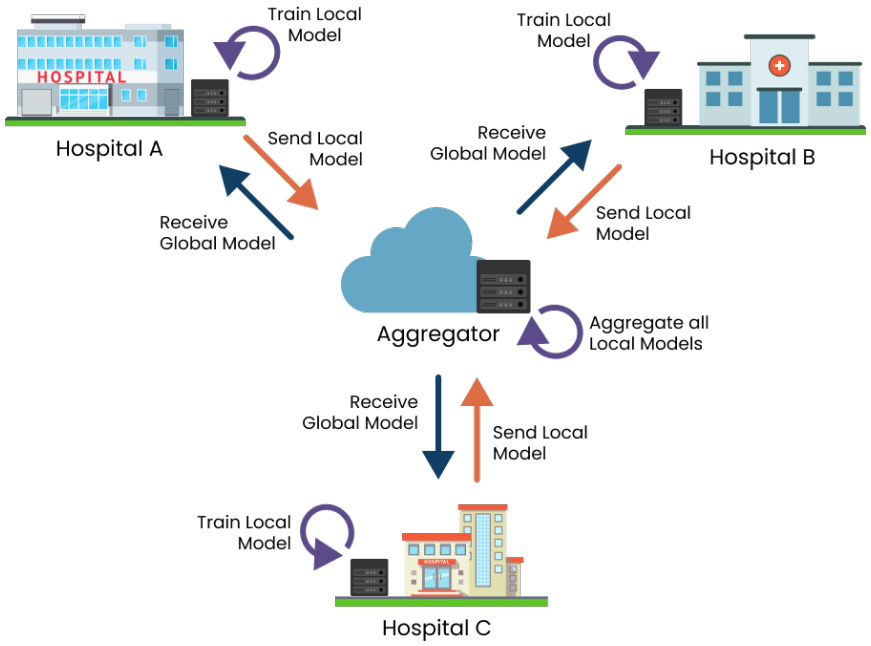


Figure 5: Privacy-preserved learning in Health AI through Federated Learning. Data remains local and secure while only model updates are shared collaboratively.

1. *Initial Model Distribution.* Central server begins the process by initializing a global machine learning model and distributes this global model to a selected subset of participating clients. The global model serves as the common starting point for all partiipants.

2. *Local Model Training.* Upon receiving the global model, each client trains the model on its local dataset. During this phase, model parameters are updated based solely on the patterns present in the local data. Critically, raw data remains on the client throughout this step and is never transmitted outside the local environment.

3. *Global Model Update.* After local training, clients send their updated model parameters back to the central servers. These updates represent what the model learned from the client's local data, but do not expose the data itself. The central server then performs global model update by aggregating the received client updates. A common approach is Federated Averaging (FedAvg),

[15] where the server computes an average of the client model parameters to produce an improved global model. This new global model reflects the collective learning across all participating clients.

4. *Iterative Refinement.* This process continues through iterative refinement, where the updated global model is redistributed to clients for further rounds of local training. With each iteration, the model progressively improves until it reacches a desired level of performance or convergence.

As figure 5 illustrates, in Federated Learning local data never needs to leave the local server where it is stored.

Privacy-Preserving Machine Learning (PPML) and secure aggregation.

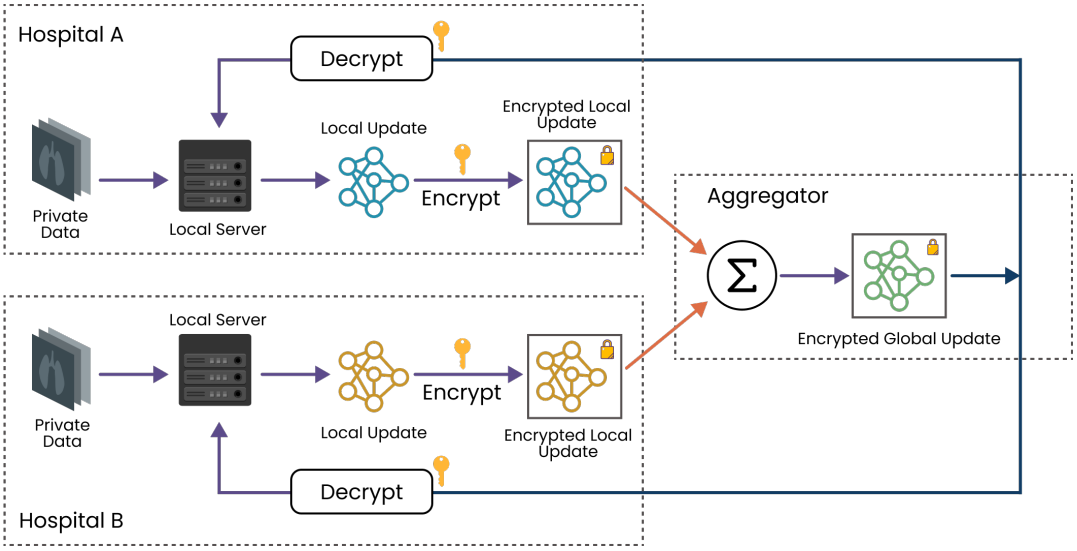


Figure 6: Homomorphic encryption: enabling the server to aggregate model updates without decrypting them, preserving client data privacy throughout.

Although Federated Learning reduces the need to centralize sensitive health data, it does not by itself eliminate privacy risks. Model updates transmitted during training may still reveal information about individual records, as shown by membership inference and reconstruction attacks.[16] This has given rise to the broader field of Privacy-Preserving Machine Learning (PPML), which integrates cryptographic and statistical techniques to strengthen confidentiality while enabling collaborative

[15] H. Brendan McMahan, et al., "Communication-Efficient Learning of Deep Networks from Decentralized Data," Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), 2017.
[16] Reza Shokri, et al., "Membership inference attacks against machine learning models", IEEE Symposium on Security and Privacy (SP), 2017.

model development. Among the principal approaches in PPML are differential privacy, which introduces controlled statistical noise to ensure that the contribution of any single record cannot be identified,[17] and homomorphic encryption, which allows computations to be performed on encrypted values.

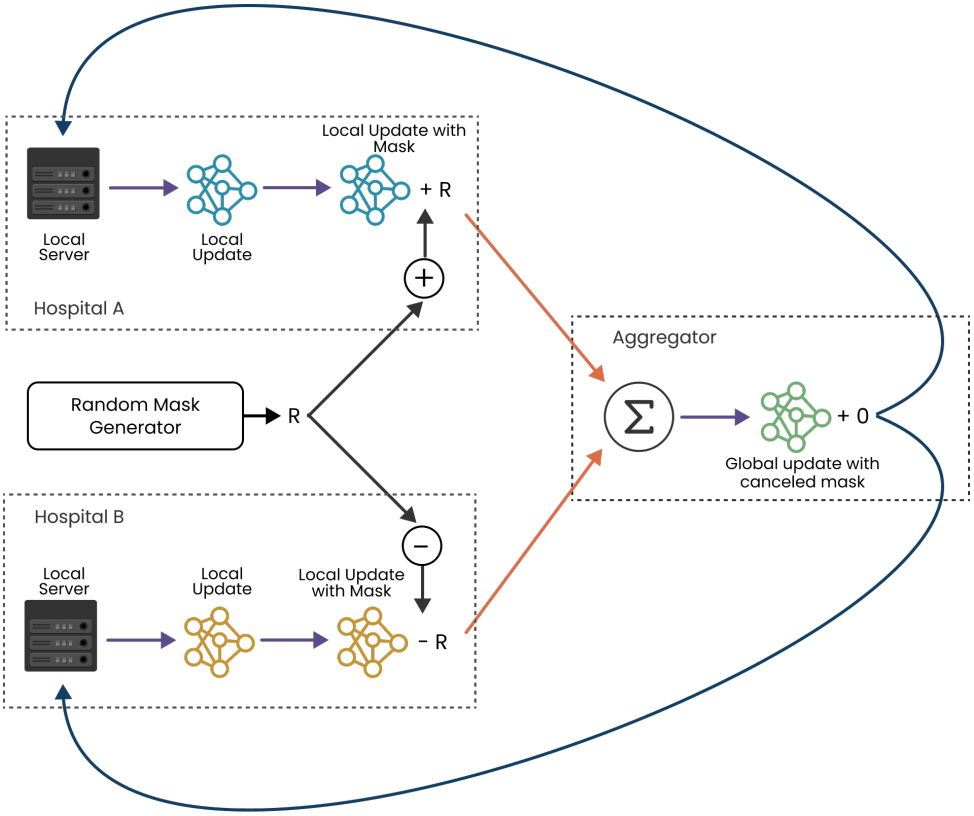


Figure 7: Secure aggregation: The random masks are designed such that they cancel out upon summation. So the server learns only A + B, without accessing them individually thereby protecting individual client updates from exposure.

A particularly important technique in the context of Federated Learning is secure aggregation.[18] Secure Aggregation is a privacy-preserving protocol that ensures the server can only see the aggregated result (e.g., sum or average of model updates), not the individual client updates. Rather than transmitting raw parameter updates, each participant encrypts or masks its local contributions in such a way that only the combined result is revealed to the coordinating server. This ensures that no individual institution's data distribution can be reconstructed while still enabling effective global model training. In healthcare, this allows hospitals and research centers to participate in federated networks without fear that their local datasets

[17]Martin Abadi, et al., "Deep learning with differential privacy", Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 2016.
[18]Keith Bonawitz, et al., "Practical secure aggregation for privacy-preserving machine learning", Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 2017.

may be exposed, thereby fostering trust among stakeholders.

Federated Learning in Healthcare

The adoption of Federated Learning in healthcare has expanded rapidly, with emerging research validating its utility across multiple clinical tasks and data modalities. Almufareh et al.[19] showed that a privacy-preserving, collaborative FL framework can significantly improve early detection performance for breast cancer across distributed clinical networks. Similarly, a cross-institutional feasibility study by Lee et al.[20] demonstrated that FL achieves tumor segmentation performance comparable to centralized learning. Moreover, on an independent meningioma dataset, the federated approach outperformed centralized models, highlighting its generalization capability. Large-scale collaborations further validate the practical utility of FL. The EXAM study, led by Mass General Brigham and NVIDIA, involved 20 hospitals across five continents.[21] The study used FL to predict supplemental oxygen requirements in COVID-19 patients based on Electronic Medical Records, achieving a 38% improvement in generalizability compared to models trained at individual institutions. Beyond unimodal settings, multimodal Federated Learning is gaining traction, enabling institutions to jointly leverage heterogeneous data sources—including medical imaging, electronic health records, and omics data—while maintaining strict data privacy constraints.[22] In the context of cardiovascular diagnosis, Xu et al.[23] evaluated FL for Acute Myocardial Infarction diagnosis using Electronic Medical Records from three hospitals in China, comprising 3,614 patient records. The federated model consistently outperformed models trained locally at individual institutions, further demonstrating the benefits of collaborative learning.

[19] Almufareh MF, et al., "A Federated Learning Approach to Breast Cancer Prediction in a Collaborative Learning Framework", Healthcare, 2023.
[20] Wei-Kai Lee, et al., "Federated Learning: A Cross-Institutional Feasibility Study of Deep Learning Based Intracranial Tumor Delineation Framework for Stereotactic Radiosurgery", Journal of Magnetic Resonance Imaging, 2023.
[21] Ittai Dayan, et al., "Federated Learning for predicting clinical outcomes in patients with COVID-19", Nat Med 27, 2021.
[22] Jacob Thrasher, et al., "Multimodal Federated Learning in Healthcare: a Review", Journal of Healthcare Informatics Research, 2025.
[23] Xu Jie, et al., "High performance of privacy-preserving acute myocardial infarction auxiliary diagnosis based on Federated Learning: a multicenter retrospective study", Ann Transl Med, 2022.

Federated Learning Adoption

While Federated Learning offers a promising paradigm for collaborative and privacy-preserving AI development, its adoption in real-world healthcare systems remains complex. Beyond the conceptual advantages, deploying Federated Learning across institutions requires addressing a range of technical, institutional, and infrastructural challenges. These challenges are particularly pronounced in resource limited settings, where healthcare systems operate under additional resource and coordination constraints.

Challenges in Federated Learning Adoption

The widespread adoption of Federated Learning is hindered by a complex set of interrelated obstacles that span both technical and institutional domains. Overcoming these barriers demands coordinated effort across disciplines — bringing together data scientists, clinicians, ethicists, and policymakers in sustained collaboration.

Technical Challenges

1. *Data Heterogeneity (Non-IID Data).* Data heterogeneity remains one of the most pressing concerns at the technical level. In federated environments, the participating entitites, whether hospitals, or research institutions, generate data under varying conditions. The data variability may be caused by the use of different instruments, acquisition protocols and/or patient demographics. This statistical non-uniformity across distributed datasets can negatively impact model convergence and compromise the generalizability of the global model.

2. *Communication Overhead.* FL requires frequent exchange of model updates between participating nodes and a central coordinating server. Transmitting these updates—particularly for high-dimensional deep learning models—can introduce significant communication bottlenecks, especially in bandwidth-constrained environments.

3. *System Heterogeneity.* Participating institutions may intermittently disconnect, possess limited computational resources, or follow asynchronous training schedules. These factors lead to “straggler” problems that complicate coordination

and can introduce instability into the global training process.[24] Ensuring consistent model performance and reliable communication across such a non-uniform ecosystem of hardware, software, and network infrastructure across healthcare institutions requires careful systems engineering and adaptive coordination protocols.

4. *Simulation environments.* A critical concern is the persistent gap between simulation and real-world deployments. Many existing studies remain confined to controlled experimental environments that, while methodologically useful, fail to capture the full complexity of operational healthcare settings, where data are inherently diverse, sensitive and specific to local contexts. Real-world deployments require secure communication protocols, flexible frameworks, and the capacity to adapt to evolving privacy needs.

5. *Research Barrier.* Compounding these challenges is a fundamental limitation within the research community: the lack of standardized evaluation frameworks. In the absence of common benchmarks, Federated Learning studies are difficult to compare, reproduce, and systematically build upon. This lack of standardization not only slows scientific progress but also hinders the reliable translation of research into real-world healthcare systems. Addressing this gap requires the development of robust, community-driven benchmarking efforts that enable consistent and transparent evaluation. Emerging initiatives such as *FLamby*[25] and *Med-MMFL*[26] represent important early steps in this direction, providing curated datasets and evaluation protocols tailored to federated and multimodal healthcare settings. Continued adoption and extension of such benchmarks will be critical for establishing reproducibility and accelerating progress in the field.

6. *Reliability of Models.* The AI models themselves pose challenge in healthcare settings. Current approaches often favor highly complex neural networks, which, although powerful, are resource-intensive and difficult to reproduce in constrained healthcare settings. Simpler, interpretable, and more personalized approaches could be more effective but remain underexplored.

[24] Bonawitz, et al., "Towards Federated Learning at Scale: System Design," Proceedings of Machine Learning and Systems, 2019.
[25] Jean Ogier du Terrail, et al., "FLamby: datasets and benchmarks for cross-silo Federated Learning in realistic healthcare settings", In Proceedings of the 36th International Conference on Neural Information Processing Systems (NeurIPS '22), 2022.
[26] Aavash Chhetri, et al., "Med-MMFL: A Multimodal Federated Learning Benchmark in Healthcare", CoRR 2026.

Institutional and Policy Challenges

Beyond the technical dimension, Federated Learning faces substantial institutional and policy headwinds. Effective inter-institutional collaboration requires not only shared technical standards but also organizational trust, aligned incentives, and clearly negotiated data governance agreements — none of which can be assumed in practice. Ethical considerations add another layer of complexity. Questions of algorithmic fairness, accountability, and transparency must be addressed proactively, particularly in high-stakes domains such as healthcare, where model decisions carry direct consequences for patients and communities. Regulatory compliance compounds these challenges further. Federated Learning initiatives must navigate a fragmented landscape of data protection frameworks — including GDPR, HIPAA, and various national regulations — whose requirements do not always map cleanly onto the distributed, cross-jurisdictional nature of federated systems. Achieving compliance while maintaining the collaborative utility of the framework requires thoughtful legal and policy design. Accountability and liability represent equally serious concerns. When a federated diagnostic model contributes to an incorrect clinical decision, the distributed nature of the framework makes assigning responsibility deeply ambiguous — it is not readily clear whether fault lies with the institution contributing the local model, the developer of the aggregation server, or the provider of the Federated Learning framework itself. This diffusion of responsibility has direct implications for patient safety and institutional trust. Without clear mechanisms for audit trails, model provenance, and responsibility attribution, clinical adoption will remain constrained by legitimate institutional risk aversion.

Implementation Challenges in Resource Limited Settings

Implementing Federated Learning in resource-constrained settings involves a specific set of infrastructural, operational, and institutional barriers that differ substantially from those encountered in high-income contexts. At the infrastructural level, most standard Federated Learning architectures assume the availability of high-end local GPUs, yet in many clinics, especially in resource limited settings, even basic server infrastructure may be absent, necessitating the use of lightweight frameworks or commodity hardware. This is compounded by unreliable internet connectivity and high data transmission costs, which render the frequent exchange of large model

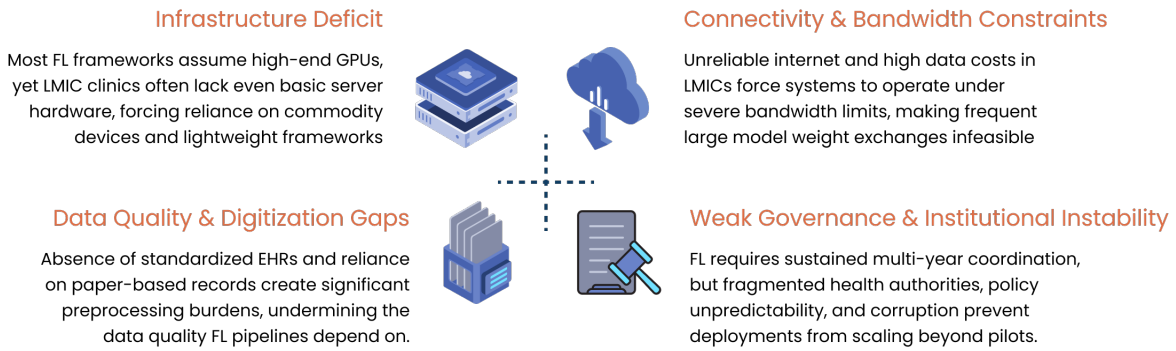


Figure 8: Implementation challenges of PriFed in resource limited settings

weights infeasible without aggressive gradient compression or sparse update strategies.[27] The lack of standardized Electronic Health Records (EHR) and the prevalence of unstructured or paper-based data further mean that the pre-processing effort required to make local data machine-ready for a federated pipeline is considerably higher than in well-digitized settings. Beyond infrastructure, the challenge is compounded by deeply rooted cultural attitudes toward health data. In healthcare systems at resource limited settigns, data are afforded low institutional priority and are rarely integrated into clinical decision-making. Health workers and administrators may not recognize data collection and management as core professional responsibilities, resulting in incomplete records, inconsistent coding practices, and limited engagement with digital health tools. This cultural undervaluation of health information poses a significant barrier to assembling the quality and volume of local data that Federated Learning pipelines depend upon. Political and governance challenges present an equally formidable obstacle. Federated Learning initiatives require sustained, multi-year coordination across institutions and administrative levels — a condition that is difficult to maintain in environments characterized by weak leadership structures, fragmented communication between community and national health authorities, and unpredictable shifts in policy and regulation. Corruption and institutional instability further undermine the long-term commitments necessary for federated infrastructure to take root and scale. Without stable governance frameworks that prioritize health data as a strategic national asset, even technically sound deployments risk stalling at the pilot stage.

Recent initiatives such as the African tuberculosis imaging study, spanning Ethiopia, Ghana, Mozambique, Nigeria, and other sub-Saharan countries,[28] have begun to

[27] Lim, et al., "Federated Learning in Mobile Edge Networks: A Comprehensive Survey," IEEE Communications Surveys & Tutorials, 2020.
[28] Jorge Fabila, et al., "Federated Learning in low-resource settings: A chest imaging study in Africa – challenges and lessons learned," 2025.

demonstrate Federated Learning's potential in such contexts. Yet significant hurdles remain. Addressing these gaps requires sustained investment in local capacity building, the development of interoperability frameworks tailored to fragmented health IT environments, and governance reforms that establish clear accountability for health data stewardship at both institutional and national levels.

Imperative for Federated Learning Adoption in Resource Limited Settings

Despite these challenges, Federated Learning presents a uniquely suitable paradigm for healthcare systems in resource limited settings. These regions face a distinctive set of structural, infrastructural, and resource-related constraints that make centralized approaches impractical, and it is precisely these constraints that Federated Learning is equipped to address. Healthcare delivery in resource limited settings spans a wide network of rural clinics, district hospitals, community health posts, and tertiary referral centers that operate with varying degrees of connectivity and institutional capacity. Federated Learning accommodates this fragmentation by design, enabling model training across distributed sites without requiring data to leave the point of care. Many resource limited settings also lack the centralized data infrastructure that conventional AI development assumes. Within a federated framework, institutions that would otherwise be excluded from Health AI research can participate meaningfully without first investing in costly centralized systems. By enabling collaborative development while keeping data local, FL provides a pathway to bridge the "AI Divide," ensuring that medical models are trained on indigenous datasets and are thus biologically and environmentally relevant to the populations they serve.[29] Models adapted from high-income contexts, by contrast, are unlikely to be representative of local populations and typically exhibit poor diagnostic and prognostic performance in these settings. Federated Learning also creates the conditions for regional and cross-border collaboration, allowing neighboring countries to collectively train more robust models without centralizing sensitive national health data. This is particularly valuable given that resource limited settings collectively represent vast, diverse patient populations with disease profiles, genetic variation, and clinical presentations that remain severely underrepresented in globally dominant Health AI datasets. Beyond model quality, such collaboration fosters the development of shared governance norms and interoperability standards across participating institutions. Operational

[29] Vayena, et al., "Machine learning in low-and middle-income countries: ethical and practical challenges," Lancet Digital Health, 2018.

resilience is a further consideration. Healthcare facilities in resource limited settings frequently contend with intermittent internet connectivity and power instability. Federated Learning frameworks can be designed to operate asynchronously, tolerating periods of disconnection and resuming participation when connectivity is restored. This makes federated approaches more operationally viable than cloud-dependent centralized alternatives in settings where reliable infrastructure cannot be guaranteed.

Finally, participation in Federated Learning networks offers resource limited settings tangible long-term value beyond improved model performance. Local institutions that contribute to federated training pipelines develop technical expertise, institutional protocols, and research infrastructure that compound over time. This positions Federated Learning as a vehicle for meaningful engagement in the global AI research ecosystem and for sustained capacity development in digital health across the resource limited settings.

Case Studies of Real-World Initiatives

The translation of Federated Learning from theoretical framework to clinical deployment requires both robust technical infrastructure and domain-specific application. The CAFEIN platform and the TRUSTroke project together illustrate how this transition can be achieved in practice, offering a concrete example of Federated Learning operating within the constraints of real-world healthcare systems.

CAFEIN: A Federated Learning Infrastructure for Medical Research

CAFEIN is a secure and modular Platform-as-a-Service (PaaS) for Federated Learning in medicine, designed by CERN in collaboration with Politecnico di Milano and the Consiglio Nazionale delle Ricerche.[30] The platform was conceived to address the dual challenges of data sovereignty and regulatory compliance in Europe, where the General Data Protection Regulation (GDPR) strictly governs the storage and

[30]CERN Knowledge Transfer, "CAFEIN – Federated network platform for the development and deployment of AI-based analysis and prediction models,"

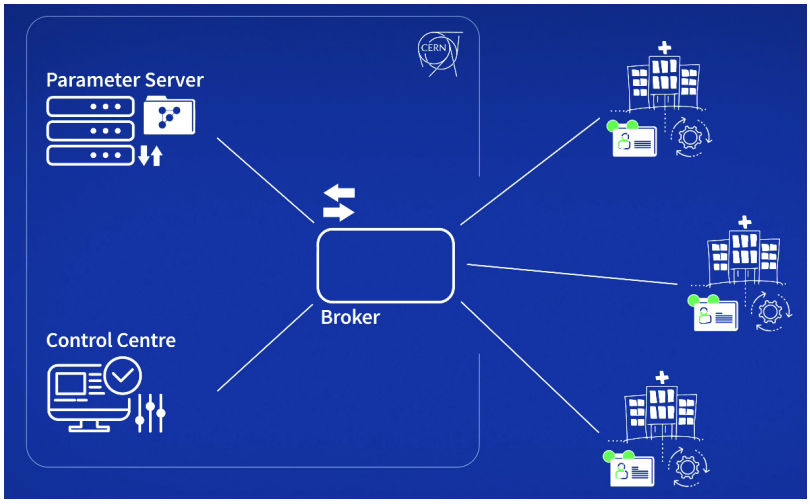


Figure 9: Federated Learning Architecture of CERN CAFEIN

processing of personal health data.[31] Rather than requiring hospitals and research institutions to exchange raw patient data, CAFEIN enables collaborative model training across distributed sites, preserving privacy while allowing collective knowledge to be shared across institutions. The platform has demonstrated its versatility across a range of medical applications. During the COVID-19 pandemic, CAFEIN was deployed to distinguish COVID-19 pneumonia from other forms of viral and bacterial pneumonia. Subsequent applications have extended this capability to Brain MRI anomaly detection and multi-pathology classification, establishing CAFEIN as a generalizable infrastructure for federated Health AI.

TRUSTroke: Federated Learning for Stroke Management

Building on the CAFEIN infrastructure, the TRUSTroke project represents a targeted clinical deployment aimed at improving stroke management through trustworthy AI-based predictive models.[32] The project focuses on key clinical endpoints including treatment response, discharge probability, readmission risk, and stroke recurrence, each of which carries direct implications for patient outcomes and resource allocation. TRUSTroke leverages CAFEIN's distributed training capabilities while ensuring compliance with a demanding regulatory landscape. Meeting these requirements within a federated framework demonstrates that performance and compliance are not competing objectives, and can be pursued in concert through careful infrastructure

[31] Paul Voigt and Axel Von dem Bussche. The EU General Data Protection Regulation (GDPR). Springer, 2017.
[32] Trustroke project. https://trustroke.eu/.

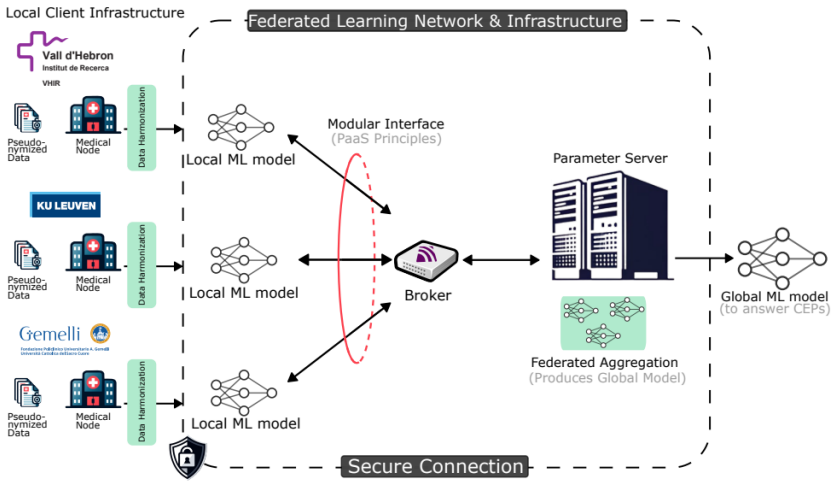


Figure 10: TRUSTroke project architecture built as PaaS on top of CAFEIN

design.

Lessons from the CAFEIN–TRUSTroke Coupling

The relationship between CAFEIN and TRUSTroke reflects a broader principle in the deployment of Federated Learning: the distinction between infrastructure and application is as important as the technology itself. CAFEIN provides the technological foundation — a secure, configurable, and scalable framework for distributed training. TRUSTroke demonstrates how that foundation can be adapted to a specific, high-stakes clinical context, producing models that are both performant and compliant with stringent ethical and legal standards. Taken together, the two initiatives illustrate a viable pathway from experimental AI systems to real-world medical practice. They also highlight the importance of designing Federated Learning platforms with regulatory compliance, clinical relevance, and institutional interoperability as first-order concerns rather than afterthoughts.

Roadmap for Federated Learning Adoption in Healthcare

The preceding sections have outlined both the transformative potential of Federated Learning in healthcare and the considerable obstacles that impede its adoption, particularly in resource-constrained settings. Realizing this potential requires deliberate,

coordinated actions across technical, institutional and policy domains. The following recommendations are intended as a practical roadmap for stakeholders seeking to advance Federated Learning from isolated pilots towards sustainable, large-scale deployment.

Technical Recommendations

Develop and Adopt Standardized Federated Learning Frameworks. The absence of common technical standards remains one of the most immediate barriers to progress. Investment in open, modular Federated Learning frameworks — designed with healthcare-specific requirements such as multimodal data support, differential privacy, and asynchronous communication — would lower the barrier to entry for institutions across resource settings and enable more meaningful comparison of research findings. Initiatives such as those building on platforms like CAFEIN offer early blueprints worth expanding and standardizing.

Build Interoperable Health Data Systems. Federated Learning cannot function effectively where underlying data are fragmented, unstructured, or encoded inconsistently. Parallel investment in health data interoperability — through the adoption of standards such as HL7 FHIR and SNOMED CT[33] — is a prerequisite for meaningful federated collaboration. This is especially critical in resource limited settings, where the transition from paper-based to structured digital records remains incomplete.

Establish Rigorous Benchmarking and Evaluation Infrastructure. The field requires standardized benchmarks that enable reproducible, transparent, and clinically meaningful evaluation of Federated Learning systems. Benchmarks must reflect real-world heterogeneity — across modalities, disease domains, and institutional settings — rather than idealized experimental conditions. Works such as Med-MMFL[34] represent essential steps in this direction and should be supported, expanded, and adopted as community standards.

Invest in Lightweight and Resilient Deployment Architectures. Technical frameworks must be designed to accommodate the operational realities of resource limited settings, including intermittent connectivity, limited compute resources,

[33] Ayan Chatterjee, et al., "HL7 FHIR with SNOMED-CT to Achieve Semantic and Structural Interoperability in Personal Health Data: A Proof-of-Concept Study", Sensors(Basel), 2022.
[34] Aavash Chhetri, et al., "Med-MMFL: A Multimodal Federated Learning Benchmark in Healthcare", 2026.

and minimal IT support. Research and development efforts should prioritize gradient compression techniques, asynchronous federated protocols, and edge-deployable model architectures that function reliably under constrained conditions.

Prioritize Explainability and Interpretability in Federated Model Design. The clinical adoption of Federated Learning hinges on the degree to which clinicians can understand, interrogate, and trust model outputs. In federated settings, interpretability is particularly challenging, as models are trained across heterogeneous data distributions, making it difficult to attribute predictions to specific features or trace the contribution of individual participating sites. Health AI systems must therefore incorporate explainability mechanisms as core design requirements rather than post-hoc additions.

Article 13 of the *EU AI Act*[35] explicitly requires transparency for high-risk systems. High-risk AI must include instructions detailing accuracy, robustness, risks, and mechanisms for interpreting outputs, ensuring deployers (e.g., clinicians) can apply them correctly. This applies directly to AI-assisted medical decisions, classifying them as high-risk with obligations for technical documentation and human oversight. Hence, Federated Learning research and deployment efforts should treat explainability as a first-class objective, investing in methods that remain coherent and informative even when models are trained across distributed, non-uniform data sources.

Institutional Recommendations

Establish Regional Health AI Networks. Federated Learning's value scales with the diversity and number of participating institutions. Regional consortia, modeled on existing examples in Europe and emerging efforts in Sub-Saharan Africa and South Asia, should be formalized with clear governance structures, shared technical standards, and equitable participation agreements. Such networks create the institutional substrate for sustained collaboration and shared ownership of resulting models.

In this context, the *Health AI for All Network (HAINet)* is well-positioned to play a catalytic role as a global convener of stakeholders across research, healthcare, and policy domains. By fostering cross-institutional partnerships, supporting knowledge exchange, and promoting equitable participation frameworks, HAINet can help

[35]European Union, "Artificial Intelligence Act," Article 13: Transparency and Provision of Information to Deployers, 2024 (entry into force 2 August 2026).

enable the development of regional Federated Learning ecosystems that are aligned with the needs of resource-limited settings.

Invest in Local Capacity Building. Technical infrastructure alone is insufficient without the human capital to deploy and maintain it. Healthcare institutions, particularly in resource limited settings, require targeted investment in training data scientists, clinical informaticists, and IT professionals who understand both the technical requirements of federated systems and the clinical contexts in which they operate. Academic partnerships, South-South collaboration, and international research fellowships all have a role to play in building this capacity sustainably.

Integrate Clinicians and Ethicists into System Design. Federated Learning systems that are designed without clinical input risk producing models that are technically sound but operationally irrelevant or ethically problematic. Institutional frameworks should mandate the meaningful involvement of clinicians, patient advocates, and ethicists from the earliest stages of system design, ensuring that federated models address genuine clinical needs and reflect the values of the communities they serve.

Establish Clear Accountability and Liability Frameworks. Institutions participating in Federated Learning consortia require clarity on questions of responsibility, audit, and liability before committing to deployment. Health AI networks should develop explicit multi-party agreements that define model provenance, update protocols, and accountability structures in the event of adverse clinical outcomes. These agreements should be developed in consultation with legal experts and updated as regulatory landscapes evolve.

Policy Recommendations

Enact Privacy-Preserving Data Governance Frameworks. Governments and regulatory bodies should develop data governance frameworks that explicitly accommodate federated and distributed learning paradigms. Existing regulations such as GDPR were designed with centralized data processing in mind and require adaptation to address the nuances of federated model training, gradient sharing, and cross-border collaboration. Clear, forward-looking guidance will reduce institutional uncertainty and accelerate responsible adoption.

Develop International Standards for Federated Health AI. The cross-jurisdictional nature of Federated Learning makes international coordination indispensable. Bodies

such as the World Health Organization, the International Telecommunication Union, and regional health authorities should work toward harmonized standards for federated AI governance, covering data minimization, model auditing, consent frameworks, and cross-border data flows. Harmonization reduces compliance fragmentation and enables the kind of global collaboration that Federated Learning makes technically possible.

Incentivize Participation from Underrepresented Health Systems. Policy mechanisms — including tax incentives, research credits, and preferential funding access — should be designed to lower the cost of participation for under-resourced institutions. Without deliberate structural incentives, Federated Learning networks risk replicating existing inequities, concentrating model development among well-resourced institutions while peripheral contributors bear participation costs with limited benefit.

Federated Learning offers a rare convergence of technical capability and ethical alignment — a framework that advances the state of Health AI while respecting the privacy, sovereignty, and autonomy of the communities whose data make that progress possible. Achieving this promise, however, is contingent on the willingness of technologists, clinicians, institutions, and governments to act with both ambition and coordination. The recommendations outlined above are not exhaustive, but they reflect the core investments and commitments without which Federated Learning will remain a compelling idea confined largely to research environments. The path from promise to practice runs through sustained, collective action — and the time to begin is now.

Conclusion

Artificial intelligence has the potential to significantly improve healthcare delivery, but its success depends on training approaches that align with real-world constraints. The traditional centralized model, while effective in controlled settings, is often impractical in healthcare environments characterized by data fragmentation, privacy regulations, and limited infrastructure.

Federated Learning offers a viable alternative by enabling collaborative model development without requiring sensitive patient data to be centralized. This makes it particularly well-suited for resource limited settings, where data is inherently distributed and centralized infrastructure is limited. However, realizing this potential requires addressing

both technical and institutional challenges, including data heterogeneity, system reliability, governance, and capacity constraints. Overcoming these barriers will depend on coordinated investments in infrastructure, skills, and collaborative frameworks.

Collaborative Networks such as the Health AI for All Network (HAINet) can play a key role in enabling these efforts by fostering partnerships and supporting the development of inclusive, privacy-preserving AI ecosystems. Through such collaboration, Federated Learning can help unlock the full potential of AI for more equitable and scalable healthcare systems, particularly in resource-constrained settings in the resource limited settings.

Health AI for All Network (HAINet)