Internet of Things and Cloud Computing
2025, Vol. 13, No. 3, pp. 62-76
https://doi.org/10.11648/j.iotcc.20251303.12

Research Article

A Privacy-Preserving Data Governance in Cross-Border Telemedicine Using Federated Learning and Differential Privacy in Kenya

Michael Meyo[1, *], Cynthia Ikamari[2], Anthony Mile[1]

[1]Department of Computer Science and Information Technology, Co-operative University of Kenya, Nairobi, Kenya
[2]Department of Mathematical Sciences, Co-operative University of Kenya, Nairobi, Kenya

Abstract

This study presents a privacy-preserving learning model designed for cross-border telemedicine in East Africa that keeps raw patient records in country while hospitals collaborate on model quality. The core of this approach is to keep sensitive patient records localized within each country, with hospitals training models locally and only sharing model updates. Using synthetic EHRs split across seven hospitals in Kenya, Tanzania, and Uganda, we compare centralized training, standard federated learning, and federated learning with differential privacy. Federated learning improves utility while maintaining data localization, with accuracy rising by about 0.0665, recall for the positive class improving by about 0.1193, and F1 increasing by about 0.0657 relative to centralized training. Adding differential privacy made the system more resilient to attacks. The success rate of model-inversion attacks dropped from 0.696 in the centralized training scenario to 0.686 with standard FL and further to 0.638 with FL + DP. This represents an absolute reduction of 0.058, or about 8.4 percent, in attack success. Membership-inference leakage has an AUC of around 0.50. The trade-off is tunable utility at a chosen privacy budget, for example accuracy near 0.530 at $\varepsilon = 0.30$. The originality is practical, we pair federated learning with an attack simulator and an ε register that turns privacy into an auditable setting hospitals can manage during cross-border care.

Keywords

Telemedicine, Federated Learning, Differential Privacy, Data Governance, Cybersecurity, Cross-Border Data Transfer, Privacy-Preserving Machine Learning, Model Inversion Attack

1. Introduction

Telemedicine has rapidly evolved from pilot deployments to mainstream clinical workflows, enabling remote use, diagnosis, and longitudinal care particularly for underserved and rural populations. As adoption accelerates in East Africa, cross-border teleconsultations and cloud-hosted services are increasingly common, raising hard questions about data sovereignty, lawful processing, and cybersecurity safeguards in the movement and computation of patient data across jurisdictions. In Kenya, these concerns are amplified by heterogeneous regulatory regimes among partner countries and the operational realities of distributed e-health platforms [9]. Consequently, privacy-preserving computation is a founda-

*Corresponding author: michaemeyo0@gmail.com (Michael Meyo)

Received: 19 August 2025; **Accepted:** 1 September 2025; **Published:** 19 September 2025

tional requirement for safe, lawful telemedicine at scale [4]. Centralized electronic health record (EHR) architectures where raw data are aggregated to a single analytical store remain common, but they have a single-point compromise, create cross-border transfer exposure, and weaken patient privacy [4]. Kenyan and regional legal analyses have repeatedly flagged centralized stores as prime targets for breach and unlawful processing, especially where data leave national borders or are mirrored in third-country clouds without adequate safeguards. These legal and security deficits motivate an architectural pivot towards decentralization, where computation is brought to the data and only privacy-safe data move across sites. Recent comparative analyses underline that Kenya's telemedicine growth coexists with regulatory gaps and infrastructure instances, especially for cross-border services. Policy harmonization and secure data-sharing are repeatedly cited as prerequisites for sustainable scale, which directly motivates a federated, DP approach that keeps data in jurisdiction while enabling collaboration [23].

Kenya's policy landscape recognizes eHealth and telemedicine as legitimate modes of health service but historically emphasized ICT system controls over the processing of health data, leaving substantive governance questions, require safeguarding of patient-level records and access controls, yet they do not fully operationalize privacy-by-design for analytics, and a comprehensive eHealth regulatory framework envisioned in 2019 has been slow to materialize [4]. For cross-border use, compliance must also align with instruments like the GDPR's transfer regime (lawfulness and appropriate safeguards) and domestic rules (e.g., Kenya's Data Protection Act, 2019) that elevate health data to a special category requiring heightened protection.

Privacy-enhancing technologies directly address this problem. Federated Learning allows institutions to collaboratively train models while keeping raw EHR data local, only model parameters/updates are shared for aggregation. When combined with Differential Privacy which perturbs updates to bound information leakage about any single individual the pipeline acquires formal privacy guarantees that are auditable against regulatory tests. [5, 6]. In African healthcare and Internet-of-Medical-Things (IoMT) contexts, FL's decentralized training has been highlighted for preserving jurisdictional control and reducing exposure during collaboration, while DP and secure aggregation harden the update channel against inference attacks and cross-site snooping [7].

From a cybersecurity standpoint, cross-border telemedicine systems face adversarial risks beyond classical network intrusion. Model inversion and membership inference attacks exploit trained parameters to reconstruct sensitive attributes or determine if a patient contributed to training [8]. Empirical guidance and regulatory analyses now treat robust anonymization of outputs as necessary, not optional, in high-risk machine learning for health. Differential privacy is state-of-the-art for this purpose, providing quantifiable privacy budgets (ε) and enabling structured trade-offs between utility and confidentiality. Our implementation foregrounds these threats and defenses the federated design limits blast radius by keeping data local, and the DP layer purposefully degrades an attacker's ability to extract patient-specific signals from the global model.

This paper operationalizes a privacy-preserving federated learning (PPFL) framework for Kenyan cross-border telemedicine through a multi-node simulation that (i) trains local models at hospital nodes without exporting raw data, (ii) aggregates updates via federated averaging, (iii) injects calibrated DP noise to bound leakage, and (iv) stress-tests privacy with a model inversion attack module. The design objective is to maximize compliance and confidentiality while maintaining clinically meaningful sensitivity for target tasks, thereby aligning technical security controls with the Kenya Data Protection Act and GDPR transfer principles. The paper opens by setting the context, reviewing telemedicine privacy, cross-border data governance, federated learning, and differential privacy. It then introduces the proposed framework and threat model, explains the dataset and preprocessing, and outlines the system architecture and evaluation approach. Next comes the experimental setup and a comparison of centralized training, federated learning, and federated learning with differential privacy, assessed with utility metrics and a model inversion attack. The discussion interprets the findings against Kenyan and international data protection requirements, highlights practical trade-offs and limitations, and draws out deployment implications. The paper closes with key takeaways and directions for future work, including a hospital pilot.

2. Problem Definition

Cross-border telemedicine in East Africa needs a way to train models without pooling raw electronic health records across jurisdictions that apply different privacy regimes. Centralized aggregation increases exposure to privacy risk and complicates lawful transfer assessments. Kenya's Data Protection Act recognizes health data as sensitive, yet implementing technical safeguards and auditable controls in routine analytics remains unclear in practice, which leaves hospitals with process-level guidance but little operational assurance. At the same time, healthcare-specific deployments of federated learning rarely pair client-side differential privacy with measurable attacker-centric tests, so risk owners cannot quantify privacy gain or audit ε selections over time. The problem is to design and validate a federated learning workflow that keeps raw EHRs in-country, applies differential privacy at the client, logs ε and training metadata for audit, and quantifies adversarial risk with model-inversion and membership-inference tests while maintaining acceptable clinical utility.

3. Literature Review

Telemedicine adoption across Sub-Saharan Africa remains

uneven, with persistent gaps between urban centers and rural counties due to connectivity, cost, and fragmented policy frameworks. Recent comparative evidence focused on Kenya, South Africa, and Nigeria shows Kenya's strength in mHealth and app-based services, yet highlights regulatory ambiguity around data privacy, reimbursement, and cross-border consultations issues that slow scale-up without clear safeguards for patient data. A systematic review argues that harmonized policy, broadband investment, and capacity-building are prerequisites for sustainable telemedicine expansion in the region, placing data governance and privacy at the center of operational readiness. These regional realities motivate privacy-preserving analytics that keep patient records local while still enabling multi-site insight and quality improvement [23, 1]. In healthcare, federated learning (FL) is widely documented as a way to train models across sites without centralizing raw electronic health records, improving data minimization and reducing single-store breach risks. Surveys of FL for smart healthcare emphasize both the promise and the practical challenges: non-IID data, straggler effects, and communication overhead that complicate clinical deployments. Recent healthcare-focused surveys consolidate applications, aggregation choices, and data partitioning issues in clinical FL [22]. Privacy-preserving FL (PPFL) taxonomies also catalog adversarial risks membership inference and model inversion in particular and recommend differential privacy and secure aggregation to limit leakage through model updates. In telemedicine data-sharing contexts, recent designs combine heterogeneous FL with security controls including model-alignment and threshold identity authentication to harden endpoints in resource-constrained clinics to coordinate learning across disparate endpoints, which is critical when hospital and clinic systems vary in bandwidth and compute [21]. African low-resource deployments further report constraints like intermittent connectivity and limited GPUs, reinforcing the need for resilient aggregation, on-device pre-processing, and privacy layers that tolerate not so ideal conditions with real hospitals facing connectivity gaps, heterogeneous data quality, and policy fragmentation [5, 25]. Differential privacy is the state-of-the-art defense against record-level leakage in trained models. Medical imaging studies that integrate DP into FL report measurable privacy gains at modest utility costs when ε is kept in a reasonable range and aggregation is tuned to clinical label balance [20]. Emerging work on inclusive, compliance-aware DP-FL goes further by aligning privacy budgets and participation weighting with regulatory constraints and equity goals, suggesting practical ways to meet data-protection tests while maintaining diagnostic performance. These results support the use of DP-bounded parameter sharing in our experiments and inform the ε choices we evaluate later and compliance-aware noise adaptation that preserves accuracy for under-resourced sites [19, 20, 24]. Across Africa, more than half of countries now have data-protection statutes, and most classify health and genetic data as "sensitive," warranting heightened protection. Several laws include research or scientific-use provisions, but they typically require meeting core processing principles and, in some cases, fresh consent for secondary use. Regional instruments like the African Union's Malabo Convention and model laws in ECOWAS and SADC set reference points for cross-border collaboration, yet implementation remains uneven. For Kenya, practice increasingly echoes GDPR concepts such as data-protection impact assessments (DPIAs), privacy-by-design default, and strengthened consent and transparency duties for processing health data. These instruments together set a compliance bar that FL with DP can help meet, because raw records remain in country and only privacy-bounded updates cross borders under documented safeguards [3, 24]. The literature converges on one point, telemedicine is expanding fast and privacy must be designed in from the start, yet most sources stop short of showing how to do it in practice [12, 13]. Ayo-Farai and colleagues review telemedicine's rise in Africa and argue it can close care gaps across the continent, but their narrative remains high level and doesn't operationalize concrete privacy-by-design controls for analytics [1]. Corrales Compagnucci and Fenwick map the legal terrain for cross-border health data, stressing transfer-risk assessments, due diligence, and structured safeguards when sensitive data moves between jurisdictions, though their guidance is primarily legal and not tied to specific machine-learning mechanisms [2]. In Kenya, Ogonjo et al. explain that health data is a constitutional and statutory privacy priority under the Data Protection Act, and that sector guidance stresses confidentiality and secure handling, yet the country's ICT standards have historically focused more on protecting systems than regulating how health data are actually processed, with the envisioned eHealth regulatory framework still unresolved, which leaves an implementation gap for analytics workflows [3]. Mensah's analysis of AI telemedicine in Ghana reaches similar conclusions, highlighting the need for privacy-enhancing technologies and stronger compliance practice, but it offers policy recommendations rather than a tested technical pathway for privacy-preserving model training [10]. Technical surveys from the IoT and smart-health community reinforce why centralizing data is brittle, noting heterogeneous devices, policies, and non-IID distributions that challenge traditional centralized learning, while calling out open problems and the need for decentralized approaches, which they don't instantiate or evaluate under health-data law constraints. Taken together, prior work establishes the risks, the regulatory expectations, and the architectural direction, but stops short of a Kenyan, cross-border-ready implementation that pairs federated learning with formal privacy guarantees and demonstrates measurable resilience to inference attacks. This paper fills that gap with an auditable, privacy-preserving federated learning pipeline that keeps raw EHR data local, applies differential privacy to client updates before aggregation, and quantifies the privacy–utility trade-off using a model-inversion attack, aligning the technical design with the DPA and cross-border

transfer principles. This work treats participating hospitals as data controllers and the coordinating service as a processor. We ran a DPIA focused on cross-border aggregation risk, set privacy budgets accordingly, and documented the transfer mechanism and rights-handling workflow so raw EHR never leaves originating jurisdictions.

4. Methodology

This study used publicly available synthetic electronic health records (EHRs) generated by the Synthea™ platform, which is widely applied in healthcare machine learning research because it produces structured, standards-compliant records without exposing real patient identities. Synthea simulates patient lifecycles from birth to death, drawing on epidemiological data, care guidelines, and population statistics. The generated EHRs reflect plausible disease progressions, medications, and encounters, while adhering to HL7 FHIR formats that are interoperable across health information systems. To model the East African cross-border telemedicine setting, we partitioned the dataset into seven distinct hospital nodes, each representing an institution in Kenya, Tanzania, and Uganda. Partitioning was stratified to preserve diversity of age, gender, and disease prevalence across nodes, simulating realistic inter-hospital variations. Each node therefore trained locally on its share of records before contributing model updates to the federated aggregation step. Although Synthea produces clinically plausible records, it does not replicate the exact disease distributions or referral pathways of East African populations. This is both a strength and a limitation. On one hand, synthetic cohorts protect privacy and allow safe experimentation on federated learning and differential privacy techniques without regulatory hurdles. On the other hand, outcomes may not fully capture local, under-resourced facility patterns, or socio-economic drivers of health that influence real-world telemedicine data. There has been no formal clinical validation of the generated cohorts against Kenyan or regional epidemiological baselines in this study.

We designed an auditable, privacy-preserving learning pipeline for cross-border telemedicine that compares three training modes on the same synthetic EHR cohort: a centralized baseline, federated learning, and FL with differential privacy. The study uses 3,459 FHIR-derived patient records with natural class imbalance, focusing on a binary target $y \in \{0,1\}$ that flags respiratory infection. Records are partitioned across seven hospitals to mimic a cross-border network, three in Kenya, two in Tanzania, and two in Uganda, reflecting the legal and operational realities emphasized in the regional privacy and governance literature. Consistent with policy advice that raw health data should not leave originating facilities, training occurs locally and only model artifacts flow. We flatten FHIR JSON to tabular form, standardize types, derive age from date of birth when necessary, normalize whitespace and categorical strings, and keep the natural imbalance by design. Let the cleaned dataset at hospital k be

$$D_k = \left(x_i^{(k)}, y_i^{(k)}\right)\}_{i=1}^{n_k} \; with \; n = \sum n_k$$

where each $x_i \in R^d$ is a feature vector and $y_i \in 0,1$ is the label for respiratory infection. We stack features such as age, gender, and site identifiers into a matrix $X_k \in R^{n_k \times d}$ and the labels into a vector $y_k \in 0,1^{n_k}$. This makes clear that learning happens on site-specific data, not a central pool.

Features comprise age, gender, hospital and country identifiers. For each experiment we build a stratified train–test split once and reuse it across modes to avoid split-induced variance.

Table 1. *PPFL Framework and Procedure.*

Phase	What happens	Inputs	Technique/Settings	Outputs	Metrics & Logs
0. Nodes & Data	Partition synthetic EHR by hospital and jurisdiction.	Synthea EHR batches per hospital; site policies	FHIR/CSV parsing; de-identification checks	Local training-ready datasets at each node	Row counts; class balance; schema checks
1. Preprocessing	Clean, impute, encode, and scale features per node.	Local EHR tables	Train/val/test split; SMOTE or class weights if needed	X_train, y_train; X_val, y_val; X_test, y_test	Preprocess report; feature list; leakage checks
2. Local Training	Train the local model for E local epochs.	Preprocessed splits; current global model	Optimizer; lr; batch; epochs E	Local weight updates or gradients	Local loss; accuracy/recall/F1
3. Differential Privacy	Clip per-sample gradients and add noise.	Per-batch gradients	DP-SGD with clip C; noise σ; ε, δ accountant	DP-noised updates	ε consumed per round; δ; clipping stats
4. Secure Upload	Transmit updates over authenticated channel.	DP-noised updates	TLS: client authentication; optional secure aggregation. This row addresses the ex-	Encrypted update at server	Handshake logs; integrity checks

Phase	What happens	Inputs	Technique/Settings	Outputs	Metrics & Logs
			ternal eavesdropper threat on the update channel, while client-side DP limits what even an honest-but-curious coordinator can infer from received updates.		
5. Aggregation	Aggregate updates to form new global model.	Client updates	FedAvg or weighted median; drop outlier updates	Global model round t	Round t hash; participating clients; timing
6. Evaluation	Compute utility and adversarial risk.	Global model; held-out data; attack datasets	Accuracy, recall, F1; model inversion; membership inference	Utility and privacy scores	AUCs; attack success; confusion matrices
7. Audit & Compliance	Register privacy and security posture per round.	Round metrics; ε, δ; model hash; roles and safeguards	ε register; DPIA checklist; cross-border transfer log	Auditable log; compliance summary	ε history; DP accountant trace; incident flags

In Centralized baseline, We fit a Random Forest (RF) on pooled data by minimizing empirical risk

$$\min_{\theta} \frac{1}{n} \sum_{k} \sum_{i=1}^{n_k} \ell\left(f_{\theta}\left(x_i^{(k)}\right), y_i^{(k)}\right)$$

Minimizing empirical risk selects a model that reduces the mean loss over a pooled dataset. The pooled dataset is defined, and the empirical risk is expressed.

$$D = \cup_k D_k$$

For a Random Forest (RF), the objective is approximated by training trees on bootstrap samples and choosing splits that reduce impurity in aggregate this seeks to minimize the empirical risk on the combined data. This pooled baseline provides an upper bound on data movement and a utility reference against privacy-preserving configurations.

$$\hat{R}(f) = \frac{1}{|D|} \sum_{(x,y) \in D} \ell(f(x), y)$$

We chose Random Forests for several reasons that match our deployment constraints. First, RFs are robust on tabular EHR features with mixed types and missingness, which improves recall under natural class imbalance. Second, trees train reliably on hospital hardware, so each site can participate without GPUs or deep-learning toolchains. Third, RF outputs are simple to aggregate across sites using probability-level, which avoids parameter averaging and still outputs a single global posterior. Neural networks could increase utility if feature interactions are complex, but they raise implementation cost and require parameter-level FedAvg or secure aggregation. In a regulated setting, that extra complexity can slow adoption and complicate auditing. We therefore treat RF as the operational baseline and use DP to tune privacy while keeping the process simple and auditable.

In Federated learning. Each hospital trains a local RF f_{θ_k} by

$$\min_{\theta_k} \frac{1}{n_k} \sum_{i=1}^{n_k} \ell\left(f_{\theta_k}\left(x_i^{(k)}\right), y_i^{(k)}\right).$$

The local empirical risk is the average loss over the samples held by hospital k. Training minimizes this risk to produce a local model. Because only model signals are shared, raw EHR data never leaves each hospital—supporting privacy and compliance.

Because classic FedAvg parameter averaging is ill-posed for tree ensembles, we aggregate probabilities with data-size weights. Probability-level fusion is a practical fit for RF because sites can emit calibrated class probabilities without exposing internal structure. For neural networks we would need parameter-level aggregation which introduces optimizer alignment and secure-aggregation dependencies that increase engineering effort and audit scope.

$$\hat{p}(y = 1 \mid x) = \sum_{k=1}^{K} \alpha_k \, \hat{p}_k(y = 1 \mid x), \; \alpha_k = \frac{n_k}{\sum_j n_j}.$$

Because decision-tree ensembles don't have compatible parameters to average across sites, we combine their predictions instead. For each case, every hospital outputs a probability for the positive class. The server then forms one global probability by taking a weighted average of those hospital probabilities. A hospital's weight equals its share of the total training examples across all sites, so larger datasets have proportionally more influence while every site still contributes. This yields a single deployable global prediction that respects local training differences and keeps raw patient records at the source.

With RFs, adding DP to client updates primarily unsettles the local gradient or statistics used in training, which reduces

overconfident posteriors at aggregation and lowers inversion success with a predictable cost to accuracy.

To provide differential privacy, local statistics g_k are clipped and Gaussian noise is added

$$\bar{g}_k = \frac{g_k}{\max(1, \|g_k\|_2 / C)}, \quad \tilde{g}_k = \bar{g}_k + \mathcal{N}(0, \sigma^2 C^2 I).$$

Before a site sends its update, we cap each per-sample gradient at a fixed L2 limit C. If a gradient is larger than C, we scale it down so its length equals C. This bounds how much any single record can influence the update and fixes the sensitivity. We then add zero-mean Gaussian noise whose standard deviation is proportional to C and controlled by a tunable sigma. The noise masks individual contributions while the true learning signal emerges when many updates are averaged. A privacy accountant turns the chosen sigma, the clipping C, and the number of training steps into an epsilon–delta guarantee, which we record for audit.

The standard deviation σ of the added noise is chosen to satisfy an (ε, δ) differential privacy guarantee. Differential privacy is treated as an operational control. We fix ε as the target privacy budget and set δ as a very small failure probability. Per-sample clipping C bounds each record's influence so sensitivity is finite. Given the sampling rate and training steps, a privacy accountant maps the noise scale σ, together with C, to an (ε, δ) guarantee, we raise σ until the composed loss stays within the target ε, and we log ε, δ, σ, C, sampling rate, steps, and the accountant for auditability. This is implemented with DP-SGD and recorded per round, including clipping statistics and the ε register. Lower ε strengthens confidentiality but can reduce utility, so ε is chosen with clinical risk appetite, data sensitivity, and cross-border obligations in mind under Kenya's DPA and GDPR transfer rules. In practice, FL already shrinks exposure by keeping raw EHRs in-country, while DP directly suppresses inference risk, which should depress model-inversion and membership-inference success rates, this aligns with guidance that anonymization of ML outputs must be demonstrable in healthcare.

$$\sigma \geq \frac{\Delta_2}{\varepsilon} \sqrt{2 \ln \frac{1.25}{\delta}}$$

Choose ε as our privacy budget and δ as a very small failure probability. Clipping at C fixes the maximum influence any single record can have. Given our sampling rate and the number of training steps, the privacy accountant reports how much privacy is spent for a chosen noise level σ. Increase σ until the composed privacy loss stays at or below the target ε. Larger σ adds more noise, which strengthens privacy but usually costs some utility. Record ε, δ, σ, C, the sampling rate, the number of steps, and the accountant used so the guarantee is auditable.

We adopt standard PPFL threat models that include curious or compromised aggregators and honest peers. Attacks considered are (i) membership inference, where an adversary estimates if a record participated in training, and (ii) model inversion, where an adversary reconstructs sensitive attributes from gradients or model outputs. These threats are realistic in hospital networks with mixed trust boundaries and multi-site learning [5, 20]. We evaluate two realistic adversaries (external black box adversary and internal aggregator) for cross-border telemedicine. The external attacker queries the deployed global model, sees only class probabilities or hard labels, and has no access to training data, weights, or gradients. The internal coordinator is assumed curious but protocol-following. It receives only client-side differentially private updates and aggregated posteriors, not raw records or clean gradients. Our simulations mirror this by limiting the attacker's view to the same post-aggregation outputs used for evaluation, which matches what a hospital IT admin or cloud coordinator can inspect in practice. We exclude collusion between clients in this study, and we do not attempt white-box gradient inversion, which requires access to parameters or per-sample gradients. With Random Forests and probability-level fusion, that information is not available by design. The external black-box model maps to a hospital or vendor exposing a triage API or a clinician dashboard. The honest coordinator maps to a national or cloud orchestration service. Channel eavesdroppers are mitigated by TLS and client authentication in our pipeline, while the client-side DP layer reduces the value of any outputs an attacker can lawfully or unlawfully obtain. This is why our leakage metrics focus on how much post-deployment model outputs reveal under these two access levels.

The attack success rate (ASR) for model inversion is defined as the fraction of test points on which the attacker correctly infers the true label.

$$\text{ASR} = \frac{1}{m} \sum_{j=1}^{m} \mathbb{1}\left[h(\hat{p}(\cdot \mid x_j)) = y_j\right].$$

The attack success rate counts how often a model-inversion attack correctly recovers the target attribute from the model's outputs. We run the attack on a held-out set and, for each case, compare the attacker's guess to the true value, then divide the number of correct guesses by the total number tested. Lower ASR means the model is leaking less. A drop in ASR when moving from centralized training to FL, and then to FL with differential privacy, signals stronger protection. For a balanced binary target, values near 50% indicate performance close to random guessing.

Differential privacy should depress both ASR and membership AUC, while FL without DP mainly reduces the attack surface by keeping raw EHR in-country.

We evaluate the utility of the models using standard metrics based on the confusion matrix [14].

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN},$$

$$\text{Precision} = \frac{TP}{TP+FP}, \quad \text{Recall} = \frac{TP}{TP+FN},$$

$$F_1 = 2 \cdot \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

These metrics summarize performance using the counts in the confusion matrix. Accuracy is the share of all predictions that are correct across both classes. Precision for the positive class answers, of the records we flagged as positive, how many were truly positive, so it reflects false alarms. Recall for the positive class answers, of all truly positive records, how many we correctly detected, so it reflects missed cases. F1 combines precision and recall into a single score by taking their harmonic mean, which rewards models that keep both high. In imbalanced health data, accuracy can look high even when positives are missed, so we prioritize positive-class recall and F1 when judging clinical usefulness.

Given the natural imbalance, we emphasize the positive-class F1 and recall, and we avoid synthetic re-sampling to preserve the deployment distribution. This quantitative evaluation is complemented by a mixed method and qualitative approach [15] and draws on sampling best practices for limited datasets [16].

We keep hyperparameters fixed across modes (trees, depth, seed) and evaluate on identical splits to isolate architectural effects. The Streamlit dashboard reproduces the workflow in real time upload, centralized training, FL, FL+DP with an ε slider (0.1–0.5) attack preview with per-model simulate inversion attack, and compare models where two live plots make trade-offs visible: a privacy–utility scatter (accuracy vs. membership-AUC) and an ε-effect plot with twin axes (accuracy vs. ε and membership-AUC vs. ε). The Compliance view overlays attack metrics against heuristic HIPAA/GDPR/Kenya-DPA risk bands to support governance discussions. This methodology ties the system to concrete legal and security expectations raw data remain local, only DP statistics cross borders and privacy risk is quantified with interpretable ε budgets and adversarial readouts an approach consistent with current telemedicine guidance.

This study trained and evaluated on Synthea™ synthetic cohorts only no identifiable patient data were used. Because synthetic records can still encode clinical regularities, we treated ethics as an operational concern. We scoped roles so each hospital remains controller of its local EHR, limited processing to on-premise training, and exposed only probability-level signals for aggregation. We registered privacy budgets (ε), model hashes, and attack scores per round to support audit and cross-border transfer reviews [4].

5. Results

We modeled a cross-border telemedicine network with 7 hospital nodes across Kenya, Tanzania, and Uganda. Data never left the hospital boundary during FL/FL+DP training, which reduces exposure and narrows the attack surface compared to central pooling. Federated training keeps raw records inside each hospital boundary, so the confidentiality risk is localized and governed by the hospital's own controls. The global model still learns across jurisdictions by ex-changing non-identifying signals, which is exactly what you want for cross-border processing. Class balance is intentionally natural, so the model's decisions reflect what security teams will defend in production rather than a lab-balanced scenario. Because these cohorts are Synthea-generated, we did not clinically validate epidemiologic representativeness findings should be read as technical proof-of-concept. The RF family behaved as expected in this setting, probability fusion lifted recall across sites, and adding DP reduced inversion success by damping overconfident posteriors at the cost of a moderate accuracy drop.

Table 2. *Node sizes and class balance.*

Hospital	Records
Aga Khan Ke	456
KNH Ke	478
MTRH Ke	450
Bugando Tz	293
Muhimbili Tz	329
Mulago Ug	695
Nsambya Ug	758

Table 3. *Class Balance.*

Target (Respiratory infection)	Count
1 (Present)	2205
0 (Absent)	1254

The results, summarized in the privacy-utility scatter plot below, show that federated learning (FL) improves on the centralized baseline in both performance and security posture. Furthermore, adding differential privacy (DP) provides measurable hardening against inference attacks at a predictable utility cost.

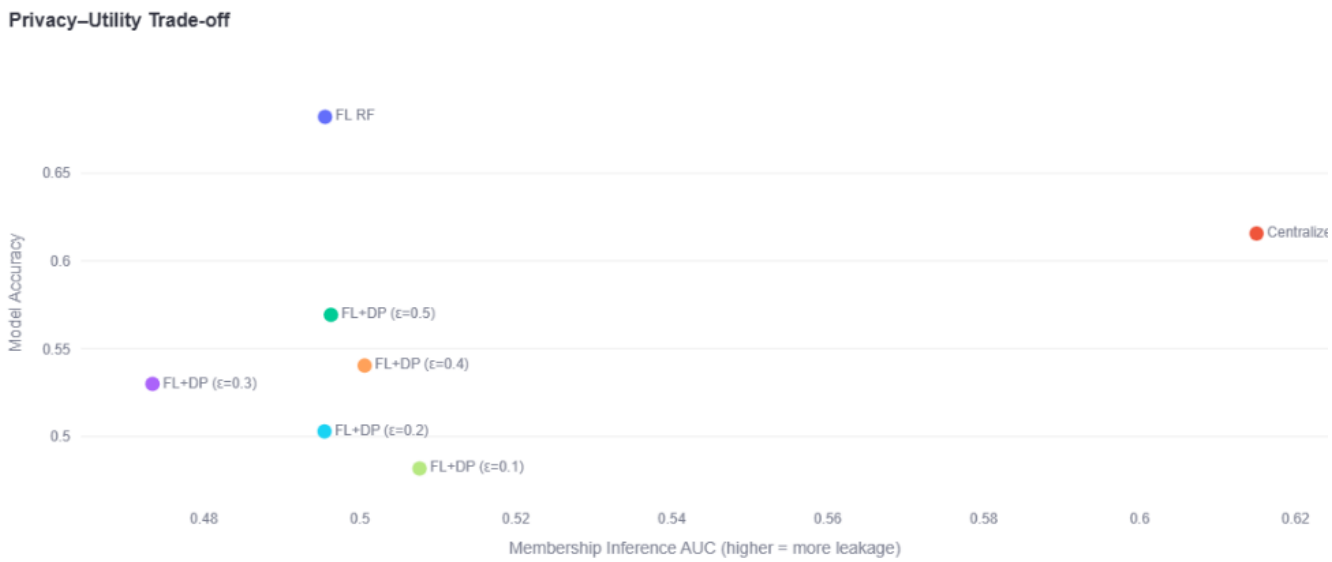

Figure 1. *Privacy-Utility Trade-off.*

FL outperforms the centralized model on accuracy while lowering leakage; FL+DP shifts leakage further left as ε decreases. Relative to centralized, FL improves accuracy by +0.0665 and F1 by +0.0657; FL+DP reduces inversion success by 8.4% absolute (0.696 → 0.638).

The centralized baseline gives a good reference point for utility, accuracy 0.616 and F1 0.706 on the positive class, but it achieves this by pooling raw patient data across borders. That design increases the blast radius of any breach, a single compromise exposes the full cohort. It also creates a high-value target, a central datastore with rich EHR features that simplifies model inversion attempts. Practically, a centralized build requires strong transfer agreements, encryption key management, and continuous monitoring of the central pipeline, which raises operating risk and cost.

Table 4. *Centralized RF metrics (test set).*

Metric	Value
Accuracy	0.6156
Precision (class 1)	0.6887
Recall (class 1)	0.7251
F1 (class 1)	0.7064
ROC-AUC	0.6233

Federated learning improves utility and shrinks exposure. Accuracy rises to 0.682 and F1 to 0.772, with a large jump in recall to 0.844 [17]. That gain is useful in clinical triage, missing fewer respiratory cases, and it lands without moving raw data. Technically, aggregating probabilities lets each hospital's model capture its own case, then the server averages signals rather than parameters. That keeps the aggregator stateless about patient identities. The remaining risks concentrate at the federation endpoints, not in a central data lake.

Table 5. *Federated Learning.*

Metric	Value
Accuracy	0.6821
Precision (class 1)	0.7112
Recall (class 1)	0.8444
F1 (class 1)	0.7721
ROC-AUC	0.7143

Table 6. *Federated Learning + Differential Privacy (ε = 0.30).*

Metric	Value
Accuracy	0.5299
Precision (class 1)	0.6617
Recall (class 1)	0.5378
F1 (class 1)	0.5933
ROC-AUC	0.5358

Adding differential privacy hardens the federation against inference, with the expected drop in utility. At ε = 0.30 accuracy is 0.530 and F1 is 0.593. The loss comes from deliberate noise injected at aggregation that reduces overconfident posteriors, which is exactly what frustrates inversion and mem-

bership tests. Treat ε as a configurable control, not a constant. Lower ε buys more privacy at a measurable utility cost, higher ε does the reverse. In deployment, you'd pick ε during a Data Protection Impact Assessment, record it with the model version and training round, then adjust if risk appetite changes.

The specific performance at this privacy budget is visualized below.

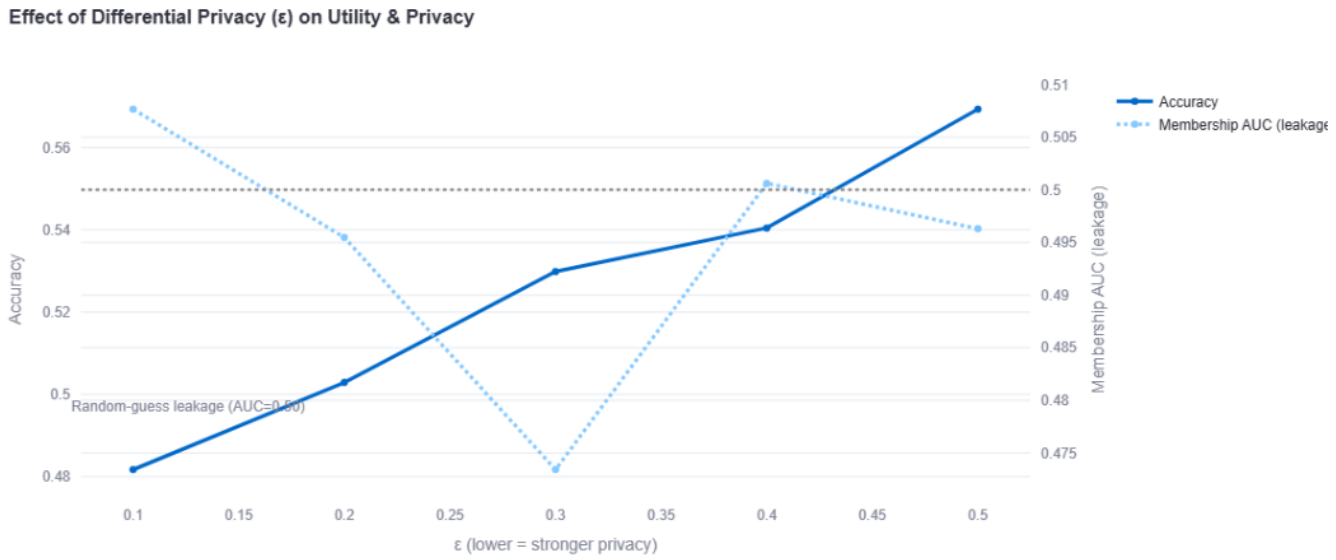

Figure 2. *Effect of DP (ε) on Utility.*

Accuracy rises as ε increases, while membership-AUC stays ~0.50 (near random-guess leakage) ε = 0.30 is our balanced operating point. As discussed in the methodology, ϵ acts as a policy dial, a lower value provides stronger privacy at a greater utility cost, while a higher value prioritizes utility. This allows organizations to make auditable, risk-based decisions that align with their Data Protection Impact Assessment (DPIA). Chosen ε, model hash, round, and attack scores are logged for auditability.

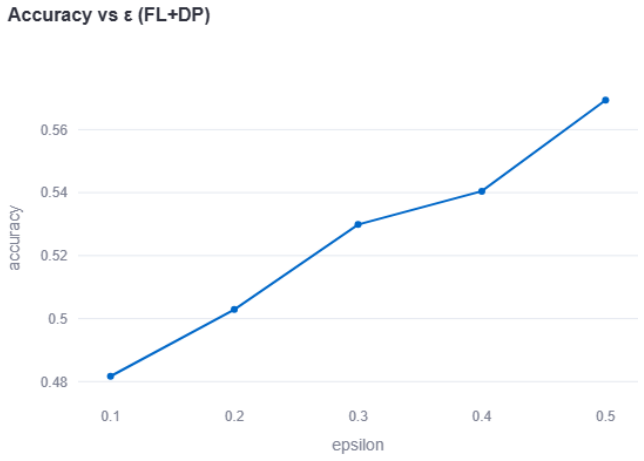

Figure 3. *Accuracy vs ε 1.*

Accuracy improves as privacy relaxes, matching the trend in Figure 2.

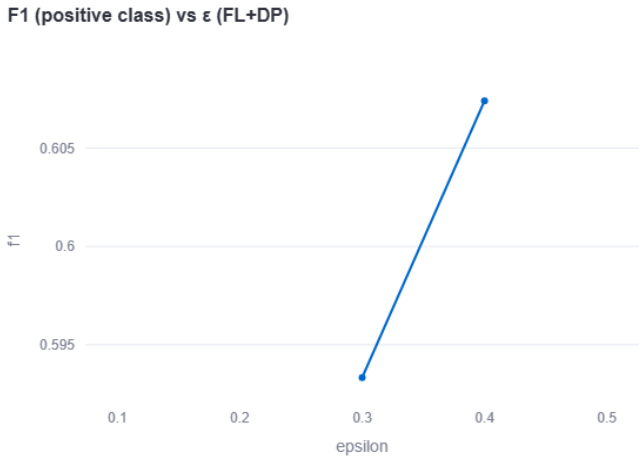

Figure 4. *F1 (positive class) vs ε 1.*

F1 for the positive class increases with ε, which matters more than accuracy.

The inversion results show the security benefit clearly. Centralized model gives an attack success of 0.696. Moving to Federated Learning drops that to 0.686, which reflects less informative outputs when the model has learned from decentralised, non-pooled distributions. Enabling Differential privacy pushes success down to 0.638 at ε = 0.30. The membership-inference AUC sits near 0.50 for all modes, which is

good, the attacker can't reliably tell if a point was in training.

The adversarial scores reported below reflect a black-box external attacker querying model outputs and an honest coordinator observing only post-aggregation signals, which are the two access levels hospitals actually face in deployment.

Table 7. *Adversarial risk signals.*

Model	Inversion Success
Centralized	0.696
Federated	0.6861
Federated+Differential (ε=0.30)	0.6376

Taken together, the results show that you don't have to trade privacy for performance as a binary choice. Federated learning alone improves both utility and security posture by eliminating central pooling. Federated learning and differential privacy gives you a tunable privacy budget that can be dialed to stricter settings when data leaves national borders or when the use case is high stakes. The attacker still extracts some signal from clinical regularities, which is why the inversion accuracy remains above 0.60 for many tabular tasks, but differential privacy demonstrably pushes that success down.

6. Discussion

Compared with existing work, our contribution puts in operation privacy-by-design for cross-border telemedicine and shows measurable privacy and utility outcomes. African telemedicine overviews map opportunities, without auditable safeguards readers can run, which is where we add value [1]. Legal scholarship lays out transfer-risk duties but doesn't tie them to concrete ML mechanisms, we have laid down a way to solve that, by keeping raw EHR in-country and logging ε, δ, C, rounds and model hashes for audit [2]. Kenyan sector analyses stress that health data is sensitive and that guidance has focused more on ICT system controls than on privacy-preserving analytics, leaving an implementation gap that our federated, client-side-DP model fills [3, 4]. Technical surveys call for decentralized learning with formal guarantees yet stop at threat notes we introduce those guarantees and report attacker-centric metrics [5, 6]. IoMT-oriented work emphasizes integrity and secure collaboration, while our evaluation centers formal DP budgets and measured leakage [7]. Unlike anonymization-only approaches, we provide a quantifiable budget and demonstrate resistance to model-inversion and membership inference in practice [8]. Empirically, FL improves utility over centralized training in our setting, accuracy 0.682 vs 0.616, with a large recall gain for the positive class, and FL+DP reduces inversion success from 0.696 to 0.638 at ε = 0.30 while keeping membership AUC near 0.50. These measured changes make the privacy–utility trade-off explicit and auditable for regulators and security teams. This study set out to answer if can we deliver useful clinical predictions across borders without centralizing raw health records, and can we make that posture measurable to a regulator and a red-team? The results say yes. Federated learning alone raised utility over the centralized baseline, and differential privacy reduced attackability further, while keeping membership-inference leakage near random guessing. In a cross-border setting, that's the mix you want: keep data at the hospital, export only minimal signals, and prove that those signals don't spill patient-level information.

Moving from centralized training to Federated learning increased accuracy, with a large jump in recall for respiratory infections. That jump matters because triage systems are punished more for misses than for extra reviews. Crucially, we got that gain while leaving raw EHR inside the hospitals. This matches the direction of telemedicine guidance that warns against wholesale aggregation and instead pushes computation to the edge, then coordinates results centrally. From a cybersecurity angle, the attack surface changes shape. There's no data lake to breach. Adding differential privacy delivered the expected trade-off, lower ε reduced inversion success. The ε register in the dashboard, paired with attack scores and model hashes, becomes an auditable privacy record. Our design keeps raw data in-country by default and lets a hospital prove that even shared model signals have bounded contribution at the individual level. Model-inversion success dropped from the centralized baseline to Federated learning and fell further under federated learning and differential privacy. The lesson is that the outputs are measurably less useful to an adversary once you stop central pooling and add Differential privacy. Federated learning lines up with what cross-border data governance scholarship keeps recommending, minimize transfers, reduce identifiability of what you do transfer, and document the risk treatment across the chain. Our hospital-wise splits were a policy simulation. They show that the largest Ugandan nodes remain local controllers, that Kenyan and Tanzanian nodes participate without sending raw records, and that the aggregator only ever sees noisy statistics under Differential privacy. Mensah's telemedicine compliance review makes a similar point in the West African context, lawful processing isn't just a legal basis, it's a technical posture that reduces what you expose in the first place. A final concern in public systems is whether the approach runs on ordinary infrastructure. The Random Forest and probability aggregation path is deliberately lightweight. It doesn't need parameter-level FedAvg or heavy secure aggregation to start delivering value. Where devices are even tighter, edge-efficient models and compression strategies from the telemedicine optimization literature [18] can slot in without changing the governance picture. The evidence supports a security-first architecture for cross-border telemedicine, Federated learning beats the centralized baseline on both utility and risk exposure, and Federated learning and differ-

ential privacy gives a tunable, regulator-friendly privacy budget that demonstrably depresses attack success. This aligns with regional privacy scholarship on health data, which pushes minimization, localization, and verifiable safeguards over grand centralization.

6.1. Practical Implications (Workflow) for Hospitals

The diagram outlines a repeatable loop hospitals can run each round. It begins with preparation and a brief DPIA that sets roles and a starting privacy budget ε and δ. Each site then trains locally on its own EHR and, when required, applies DP-SGD by clipping per-sample gradients at C and adding Gaussian noise with σ. The site sends only the noised updates over an authenticated, encrypted channel to a coordinator, which performs weighted aggregation and returns a new global model. Teams evaluate both sides of the trade-off, utility metrics and attacker audits for inversion and membership, then record ε, δ, C, σ, model hash, round, and scores in an ε register. Using those results, they tune ε, redeploy, and repeat the cycle.

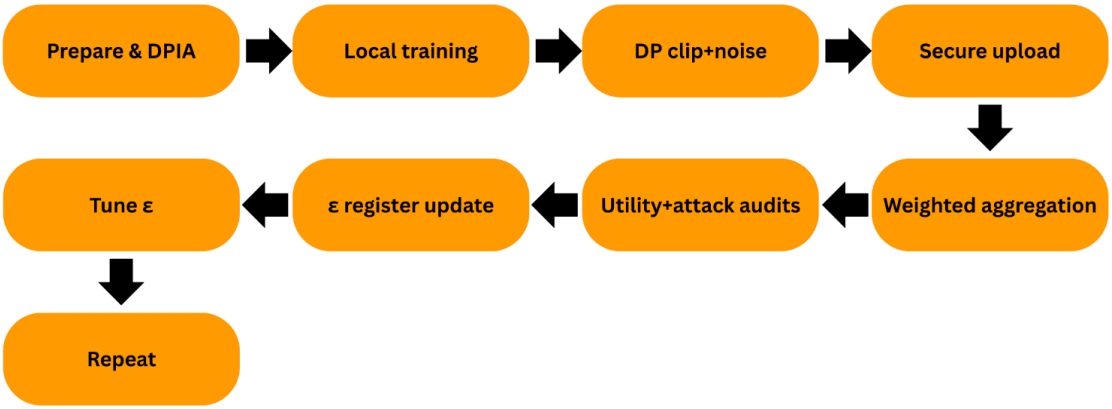

Figure 5. Replication Workflow.

6.2. Legal and Policy Analysis: Kenya in Context

Several analysis shows Africa's legal landscape is maturing, over half of countries have enacted data-protection laws, and most elevate health and genetic data to "sensitive" status that triggers enhanced safeguards. Many systems provide research solutions but still require core principles, accountability, and in some settings new consent for secondary use. For cross-border analytics, regional instruments (Malabo/ECOWAS/SADC) complement national statutes but leave interoperability gaps. Aligning with GDPR-style obligations, Kenyan guidance stresses DPIAs, privacy-by-design default, and explicit transparency when processing health data criteria that our federated, DP-bounded design was built to satisfy [3, 24]. HIPAA illustrates a different approach it centers on organizational safeguards and distinguishes treatment disclosures from other uses, which reinforces the value of technical privacy controls when collaborating across systems [3]. Kenya treats health data as sensitive and requires necessity, proportionality, and appropriate safeguards, including specific rules for cross-border transfers under Sections 48–49 of the Data Protection Act and sector policy that emphasizes confidentiality, consent, and patient rights. Our design aligns with that posture by keeping raw EHR local, moving only bounded model signals, and recording ε and other audit metadata so safeguards are demonstrable [4]. Comparatively, other African regimes converge on similar principles while differing in mechanics. Deploying to Nigeria means aligning with national data protection obligations (e.g., lawful basis, data subject rights, transfer safeguards) and documenting transfer mechanisms when any processing touches external infrastructure. South Africa's environment similarly elevates health data and requires appropriate measures for cross-border sharing, so a localization-first, PETs-enabled workflow remains the safest fall back option. For EU-facing collaborations, the GDPR's transfer regime requires a lawful transfer tool plus "supplementary measures" where needed. The ε-register, local-only raw data, and minimized outputs operate as technical and organizational safeguards in that assessment, complementing contractual terms. In short, the same PPFL controls that satisfy Kenya's necessity and safeguards logic also help meet EU transfer expectations because they reduce the identifiability and value of anything that leaves a hospital boundary [2]. Operationally, broader deployments need three adaptations. First, role mapping and processor contracts: name each hospital as a controller of its local EHR and the coordinator as a processor, then bind the processor to privacy budgets, logging, and breach duties. Second, transfer assessments and controls: run a DPIA that records the chosen ε, the threat model, and the residual risk after PETs, and attach supplementary measures where a partner sits outside the tar-

get jurisdiction. Third, rights management at runtime: ensure access, correction, and objection rights can be honored without touching raw cross-site data by scoping rights responses to local controllers and documenting how model outputs are handled as non-raw artifacts. These steps reflect current cross-border health-data scholarship and Kenyan sector analyses that call for executable safeguards rather than policy statements alone [2]. Finally, regional work on e-health governance notes implementation gaps around securing data in transit and at rest. PETs help close that gap by design no central pool, encrypted channels, and DP-bounded outputs. This is consistent with Kenyan analysis that recommend a standard guidance on consent, rights, and data sharing in the health sector, and it addresses the persistent critique that formal frameworks have not fully translated into operational security [4, 11].

7. Conclusion

East African telemedicine can deliver clinically useful models without centralizing raw patient records, and that privacy can be measured, not guessed. We framed the problem as a security and governance challenge first, then an optimization task second. The core contribution is an auditable learning process that trains per-hospital, aggregates only privacy-bounded signals, and exposes live risk processes through attack simulations and ε budgets. In our seven-node simulation across Kenya, Tanzania, and Uganda, federated learning improved utility over a centralized baseline while leaving data in place. Federated learning with differential privacy then reduced attack success further, with the expected accuracy cost that a risk owner can tune by moving ε. This closes the loop between engineering choices and compliance duties that regional analyses have called for in health data transfers and localization debates. From a cybersecurity standpoint, the architecture changes the attack surface in our favor. Central pooling created a single breach point and expanded cross-border exposure. Moving computation to the hospitals removed that honey-pot. The coordinator handled only model signals and audit metadata, which are easier to protect with channel security, client authentication, and logging. Our adversarial tests reflect that shift. Model inversion success fell when we moved from centralized training to federated learning, and it dropped again when we injected calibrated Gaussian noise at aggregation. This is exactly what privacy-enhancing technologies are supposed to do in high-risk health analytics, lower the extractable signal from outputs while keeping decisions useful. The policy story is equally important. While this study is grounded in East African law, many research partners and cloud vendors fall under U.S. jurisdiction. Our design helps meet key HIPAA obligations without depending on blanket de-identification. First, federated learning satisfies the minimum necessary principle, by keeping raw PHI on-prem and transmitting only model signals. Differential privacy at aggregation reduces residual re-identification risk in outputs and aligns with the Security Rule's requirement to protect against reasonably anticipated"= threats, while our attack surface. In short, Federated learning as the default and FL+DP for higher-risk exchanges provide a HIPAA-aware posture that limits PHI exposure, makes safeguards auditable, and reduces the need for cross-border transfer of identified records. Kenya's Data Protection Act elevates health data to a special category and expects necessity, proportionality, and appropriate safeguards for any cross-border movement. GDPR's transfer regime takes a similar line. Our system operationalizes those expectations. Data minimization is satisfied because raw EHR never leaves the hospital by default. Purpose and proportionality are enforced because the only thing that moves is the minimum signal needed to form a consensus model, and that signal is clipped and noised under a documented privacy budget. The dashboard's ε register, model hash, and attack scores create an audit trail that a Data Protection Impact Assessment can reference. This responds to regional calls for moving beyond policy statements to executable safeguards that can be inspected and verified in practice. The governance package that comes out of this work is practical. Run federated learning as the default, not the exception. Treat ε as a policy control, not a research hyper-parameter. Record ε, model version, training date, and the two attack scores for every training round. This is how a hospital shows a regulator that privacy-by-design is a control loop with thresholds, and actions. The strategic implication for Kenya, Tanzania, and Uganda is clear. Regional telemedicine can scale faster when we stop arguing about a single neutral data lake and start federating models across hospitals. The sites keep sovereignty over their records. When a cross-border transfer is unavoidable, what travels is already privacy-bounded and documented. That reduces the risk down to what a security team can actually monitor. It also creates a path for continuous improvement. You can tighten ε for sensitive populations or use cases, relax it when evidence shows low risk. This research shows that privacy-preserving federated learning is a workable foundation for cross-border telemedicine in East Africa. It improves clinical utility over a centralized baseline, it narrows the attack surface, and it produces evidence that privacy controls are functioning. Adding differential privacy gives you a dial to meet different risk appetites and transfer contexts, and it measurably lowers an adversary's success. The approach fits the spirit and letter of Kenya's DPA and the broader transfer rules that govern sensitive health data. Most importantly, it gives hospitals a way to collaborate without surrendering their patients' privacy. That is a practical win for cybersecurity, for compliance, and for care.

8. Recommendations

Run federated learning as the default for cross-border analytics and reserve FL+DP for higher-risk transfers. Treat ε as a policy control, not a tuning trick. Record ε, model hashes,

rounds, and attack scores per training cycle to create an auditable privacy log. Adopt a DPIA-first workflow for any cross-border aggregation and document roles, safeguards, and rights handling before deployment. Align operational hardening with Kenya's DPA and HIPAA/GDPR by keeping raw EHR in-country, authenticating clients, securing the channel, and monitoring post-deployment inversion/membership drift.

Abbreviations

FL	Federated Learning
DP	Differential Privacy
PPFL	Privacy Preserving Federated Learning
PETs	Privacy Enhancing Technologies
EHR	Electronic Health Records
FHIR	Fast Healthcare Interoperability Resources
IOMT	Internet Of Medical Things
PHI	Protected Health Information
DPAK	Data Protection Act (Kenya)
GDPR	General Data Protection Regulation
HIPAA	Health Insurance Portability and Accountability Act
DPIA	Data Protection Impact Assessment
RF	Random Forest
ASR	Attack Success Rate
ROC-AUC	Receiver Operating Characteristic Area Under The Curve
F1	F1 Score (Harmonic Mean Of Precision and Recall)
FedAvg	Federated Averaging
ε (Epsilon)	Privacy Budget In Differential Privacy
σ (Sigma)	Standard Deviation Of Gaussian Noise

Acknowledgments

The authors wish to extend their heartfelt thanks to the Co-operative University of Kenya for the invaluable academic and technical support received during this research. Heartfelt appreciation is extended to Dr. Cynthia Ikamari and Dr. Anthony Mile for their unwavering guidance, insightful feedback, and support throughout the model development and evaluation stages. Gratitude is also extended to colleagues, peers, and all those who contributed directly or indirectly to the successful completion of this study. Their support and insights remain deeply appreciated.

Data Availability Statement

This study made use of publicly available synthetic healthcare records generated using the Synthea™ open-source patient population simulator. The dataset employed for experimentation was accessed from Kaggle under the repository "Synthea Dataset JSONs EHR"
https://www.kaggle.com/datasets/krsna540/synthea-dataset-jsons-ehr/data

The dataset is openly accessible to the public for research and educational purposes under the terms specified by the original authors. No additional permissions were required to obtain or use the data in this study. These data are synthetic and contain no PHI we did not conduct a formal clinical validation against Kenyan, Tanzanian, or Ugandan reference populations, so representativeness is not claimed.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] O. Ayo-Farai, O. Ogundairo, C. P. Maduka, C. C. Okongwu, A. O. Babarinde, and O. T. Sodamade. Telemedicine in healthcare: A review of progress and challenges in Africa. Matrix Science Pharma, 7(4): 124–132, 2023. https://journals.lww.com/mtsp/fulltext/2023/07040/telemedicine_in_health_care__a_review_of_progress.4.aspx

[2] M. Corrales Compagnucci and M. Fenwick. A multidisciplinary perspective on cross-border health data transfers. 2024. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4934468

[3] F. A. Ogonjo, R. Achieng, and M. Zalo. An overview of data protection in the Kenyan health sector. Strathmore University, 2022. https://cipit.org/an-overview-of-data-protection-in-kenyan-health-sector/

[4] S. O. Matagi and S. Kaneko. Challenges and opportunities on data protection and privacy in healthcare. International Journal of Scientific Research Updates, 5(1): 23–41, 2023. https://orionjournals.com/ijsru/content/challenges-and-opportunities-data-protection-and-privacy-healthcare

[5] D. C. Nguyen, Q. V. Pham, P. N. Pathirana, M. Ding, A. Seneviratne, Z. Lin, and W. J. Hwang. Federated learning for smart healthcare: A survey. ACM Computing Surveys, 55(3): 1–37, 2022. https://doi.org/10.1145/3501296

[6] X. Yin, Y. Zhu, and J. Hu. A comprehensive survey of privacy-preserving federated learning: A taxonomy, review, and future directions. ACM Computing Surveys, 54: 72, 2021. https://doi.org/10.1145/3460427

[7] A. K. Alkhalifa, M. H. Alanazi, K. Mahmood, W. S. Almukadi, M. A. Qurashi, A. H. Alshehri, and A. A. Mohamed. Harnessing privacy-preserving federated learning with blockchain for secure IoMT applications in smart healthcare systems. Fractals, 2024, article 2540020. https://doi.org/10.1142/S0218348X25400201

[8] I. E. Olatunji, J. Rauch, M. Katzensteiner, and M. Khosla. A review of anonymization for healthcare data. Big Data, 12(6): 538–555, 2024. https://doi.org/10.1089/big.2021.0169

[9] J. Kitili and N. Karanja. The new wave of eHealth: AI and privacy concerns? A case study of Kenya. 2023. https://idl-bnc-idrc.dspacedirect.org/bitstreams/000b82d8-ceb2-4c3d-94b1-357839684a5c/download

[10] G. B. Mensah. Privacy, security, and compliance in AI telemedicine. 2023. https://www.researchgate.net/profile/George-Benneh-Mensah/publication/381584030_Privacy_Security_and_Compliance_in_AI_Telemedicine/links/66754457d21e220d89c5767c/Privacy-Security-and-Compliance-in-AI-Telemedicine.pdf

[11] G. N. O. Munyolo. Cyber-security in E-health: A critical analysis of the regulatory framework in Kenya. Doctoral dissertation, University of Nairobi, 2021. https://www.asbatdigitallibrary.org/resource/6412290e8609de63e8c0d78e/view

[12] S. Dalkin, N. Forster, P. Hodgson, M. Lhussier, and S. M. Carr. Using computer assisted qualitative data analysis software (CAQDAS; NVivo) to assist in the complex process of realist theory generation, refinement and testing. International Journal of Social Research Methodology, 24(1): 123–134, 2021. https://doi.org/10.1080/13645579.2020.1803528

[13] J. Pyo, W. Lee, E. Y. Choi, S. G. Jang, and M. Ock. Qualitative research in healthcare: Necessity and characteristics. Journal of Preventive Medicine & Public Health, 56: 162–170, 2023. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9925284/

[14] A. Rahman and M. G. Muktadir. SPSS: An imperative quantitative data analysis tool for social science research. International Journal of Research and Innovation in Social Science, 2021. https://www.academia.edu/download/79116055/300-302.pdf

[15] E. Smajic, D. Avdic, A. Pasic, A. Prcic, and M. Stancic. Mixed methodology of scientific research in healthcare. Acta Informatica Medica, 30(1): 57–61, 2022. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9226784/

[16] M. Parsaeian, M. Mahdavi, M. Saadati, P. Mehdipour, A. Sheidaei, S. Khatibzadeh, and S. Shahraz. Introducing an efficient sampling method for national surveys with limited sample sizes: Application to a national study to determine quality-of-care. 2021. https://doi.org/10.1186/s12889-021-11441-0

[17] R. Scofano, A. Monteiro, and L. Motta. Evaluation of the experience with the use of telemedicine in a home dialysis program—A qualitative and quantitative study. BMC Nephrology, 23: 34, 2022. https://doi.org/10.1186/s12882-022-02824-5

[18] D. Stilinski. Optimizing convolutional neural network models for resource-constrained devices in telemedicine: A lightweight approach. 2024. https://easychair.org/publications/preprint/5JXp/open

[19] Parampottupadam, S., Coşğun, M., Pati, S., Zenk, M., Roy, S., Bounias, D.,... & Maier-Hein, K. (2025). Inclusive, Differentially Private Federated Learning for Clinical Data. arXiv preprint arXiv: 2505.22108. https://arxiv.org/abs/2505.22108

[20] Adnan, M., Kalra, S., Cresswell, J. C., Taylor, G. W., & Tizhoosh, H. R. (2022). Federated learning and differential privacy for medical image analysis. Scientific reports, 12(1), 1953. https://www.nature.com/articles/s41598-022-05539-7

[21] Wang, N., Zhang, J., Huang, J., Ou, W., Han, W., & Zhang, Q. (2024). Telemedicine data secure sharing scheme based on heterogeneous federated learning. Cybersecurity, 7(1), 56. https://link.springer.com/article/10.1186/s42400-024-00250-8

[22] Madathil, N. T., Dankar, F. K., Gergely, M., Belkacem, A. N., & Alrabaee, S. (2025). Revolutionizing healthcare data analytics with federated learning: A comprehensive survey of applications, systems, and future directions. Computational and Structural Biotechnology Journal. https://www.sciencedirect.com/science/article/pii/S2001037025002223

[23] Agbeyangi, A. O., & Lukose, J. M. (2025, March). Telemedicine adoption and prospects in sub-Sahara Africa: a systematic review with a focus on South Africa, Kenya, and Nigeria. In Healthcare (Vol. 13, No. 7, p. 762). MDPI. https://www.mdpi.com/2227-9032/13/7/762

[24] Munung, N. S., Staunton, C., Mazibuko, O., Wall, P. J., & Wonkam, A. (2024). Data protection legislation in Africa and pathways for enhancing compliance in big data health research. Health Research Policy and Systems, 22(1), 145. https://link.springer.com/article/10.1186/s12961-024-01230-7

[25] Fabila, J., Garrucho, L., Campello, V. M., Mart ń-Isla, C., & Lekadir, K. (2025). Federated learning in low-resource settings: A chest imaging study in Africa--Challenges and lessons learned. arXiv preprint arXiv: 2505.14217. https://arxiv.org/abs/2505.14217

Biography

Michael Meyo is an IT Manager at Fadhili Development Programme. He is a continuing student at the Co-operative University of Kenya, pursuing his master's degree. He holds a Bachelor's degree in Information Technology from the Jomo Kenyatta University of Agriculture and Technology and a Diploma in Business Information Technology from the Co-operative University of Kenya. He has also undertaken CISCO networking training through the Kenyatta University Networking Academy. His professional and academic journey reflects a strong commitment to advancing technology-driven solutions, digital innovation, and secure information management.

Cynthia Ikamari is a Lecturer at the Cooperative University of Kenya and a researcher in Mathematical Finance and Quantitative Economics at Multimedia University of Kenya. Her expertise lies in stochastic modeling, quantitative risk analysis, and financial mathematics. She also applies advanced mathematical and statistical methods to interdisciplinary fields, including data governance and privacy-preserving computation, supporting research on secure and resilient models for cross-border telemedicine.

Anthony Mile is a Lecturer at the Co-operative University of Kenya with expertise in Information Technology, computer and network security, and cloud computing. His research spans network architecture, secure communication, and applied computing, with contributions that strengthen the foundations of cybersecurity and data governance in emerging digital ecosystems.

Research Field

Michael Meyo: Cybersecurity, Privacy-Preserving Machine Learning, Federated Learning, Differential Privacy, Data Governance in Telemedicine, Artificial Intelligence for Healthcare

Cynthia Ikamari: Financial Mathematics, Quantitative Risk Analysis, Stochastic Modeling, Applied Statistics in Data Security, Economic Modeling & Data Governance, Simulation of Complex Systems

Anthony Mile: Computer and Network Security, Cloud Computing, Network Architecture, Secure Communication Protocols, Wireless Networks & IoT Security, Data Governance and Digital Systems Security