PRIVACY PRESERVING DATA GOVERNANCE IN CROSS BORDER TELEMEDICINE USING FEDERATED LEARNING AND DIFFERENTIAL PRIVACY IN KENYA

MICHAEL MEYO OTIENO

A THESIS SUBMITTED TO THE DEPARTMENT OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY IN THE SCHOOL OF MATHEMATICS AND COMPUTING IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF MASTER OF SCIENCE IN CYBER SECURITY OF THE CO-OPERATIVE UNIVERSITY OF KENYA.

2025

DECLARATION

Declaration by the candidate

This thesis is my original work and has not been presented for a degree in any other University or for any other award.

Name of candidate and Reg. Number……Michael Meyo……. C005/600022/2023…

Signature:  Date: 21 NOV 2025

Declaration by the supervisors

We confirm that the work reported in this thesis was carried out by the candidate under our supervision and has been submitted with our approval as university supervisors.

Name of Supervisor, Department, Faculty:

Dr. Cynthia Ikimari, Department of Computer Science and Information Technology, School of Computing and Mathematics, The Cooperative University of Kenya

Signature:  Date: 21 NOV 2025

Name of Supervisor, Department, Faculty:

Dr Anthony Mile, Department of Computer Science and Information Technology, School of Computing and Mathematics, The Cooperative University of Kenya

Signature:  Date 21 NOV 2025

ACKNOWLEDGMENT

I would like to express my deepest gratitude to The Almighty for granting me the strength, wisdom, and perseverance to undertake and complete this research. I am grateful for all of my supervisors' guidance, encouragement, and feedback, as he has been an invaluable asset to me as a researcher. Because of their extensive background in cybersecurity, telemedicine, and machine learning, they were an invaluable resource as I worked to refine my concepts and construct my thesis. I am incredibly grateful to the Cooperative University of Kenya for providing me with the resources, support, and environment that I needed to do my research. To my classmates and coworkers, I am eternally grateful for the insightful discussions, unwavering support, and inspiring words they have provided me with during this academic journey.

TABLES OF CONTENTS

LIST OF ABBREVIATIONS AND ACRONYMS

AI – Artificial Intelligence

API – Application Programming Interface

ASR – Attack Success Rate

AUC – Area Under the (ROC) Curve

CISO – Chief Information Security Office

DPA – Data Protection Act

DP – Differential Privacy

DPIA – Data Protection Impact Assessment

ECOWAS – Economic Community of West African States

EHR – Electronic Health Records

EU – European Union

FL – Federated Learning

GDPR – General Data Protection Regulation

HIPAA – Health Insurance Portability and Accountability Act

IoMT – Internet of Medical Things

IT – Information Technology

IID – Independent and Identically Distributed

ML – Machine Learning

MIA – Membership Inference Attack

MoH – Ministry of Health (Kenya)

MPC – Multi-Party Computation

ODPC – Office of the Data Protection Commissioner

PET – Privacy-Enhancing Technologies

PIPEDA – Personal Information Protection and Electronic Documents Act

PPFL – Privacy-Preserving Federated Learning

RF – Random Forest

SADC – Southern African Development Community

SGD – Stochastic Gradient Descent

SMOTE – Synthetic Minority Oversampling Technique

SPSS – Statistical Package for the Social Sciences

SVM – Support Vector Machine

TFF – TensorFlow Federated

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This thesis presents an auditable, privacy-preserving learning workflow for Kenyan cross-border telemedicine. Hospitals train models locally and share only model signals, so raw EHRs remain in country. Using synthetic Synthea EHRs, 3,459 records are partitioned across seven hospitals in Kenya, Tanzania, and Uganda to compare a centralized baseline, federated learning (FL), and FL with client-side differential privacy (DP). Random Forests are trained per site; probability-level fusion forms a global prediction without parameter averaging. The threat model covers a black-box external adversary and an honest-but-curious coordinator. We quantify privacy risk with membership-inference AUC and a model-inversion attack, and we log ε, δ, clipping C, noise σ, model hashes, rounds, and attack scores in an ε-register for audit. FL improves utility while maintaining localization: accuracy rises from 0.616 to 0.682 and F1 from 0.706 to 0.772, with positive-class recall reaching 0.844. Adding DP at $\varepsilon = 0.30$ reduces model-inversion success from 0.696 (centralized) to 0.638 (FL+DP), an absolute drop of about 8.4%, with membership-inference AUC near 0.50 (≈ random). Utility remains tunable at the chosen privacy budget, for example accuracy near 0.530 and F1 near 0.593 at $\varepsilon = 0.30$. The originality is practical: DP-bounded FL is paired with an attacker simulator and an ε-register that turns privacy into an operational, auditable control aligned with Kenya's Data Protection Act and GDPR transfer principles. The dataset is synthetic and not clinically validated for East African representativeness, so results indicate technical feasibility; a regulated hospital pilot is the next step.

CHAPTER ONE

1 INTRODUCTION

The background of the study, the statement of the problem, the study's objectives, and the research questions is explained in this chapter. It also outlines the significance, scope, and limitations of the study, as well as the study's organization.

1.1 Background of Study

1.1.1 Concept of Telemedicine

Telemedicine refers to the delivery of health services such as consultation, diagnosis, treatment, monitoring, and health education using information and communication technologies when the patient and provider are geographically separated. It is recognized globally as a key strategy for expanding access to care, especially in rural, underserved, or fragile settings, by overcoming distance and specialist shortages. International evidence from low- and middle-income countries shows that telemedicine can improve continuity of care, reduce travel costs, and enhance timeliness of diagnosis when appropriately governed and integrated into existing health systems.

1.1.2 Telemedicine Infrastructure and Adoption in Eastern Africa and Kenya

In sub-Saharan Africa, telemedicine adoption has accelerated, particularly after the COVID-19 pandemic, but remains uneven due to infrastructure and policy gaps. In Kenya, the National eHealth Policy 2016–2030 and the more recent Digital Health Act 2023 provide a formal mandate for digital health and telemedicine, emphasizing secure data exchange and interoperable

systems (Ministry of Health, 2016; Ministry of Health, 2023). Despite this progress, telemedicine services are still concentrated in a limited number of facilities. A rapid review of telemedicine use during the first year of the COVID-19 pandemic in sub-Saharan Africa identified only 11 empirical studies of telemedicine deployments across the entire region, which the authors characterize as only "moderate" uptake (Chitungo et al., 2021 Similar pilots and early-stage deployments exist in neighboring East African countries, but cross-border services remain constrained by fragmented infrastructure, uneven connectivity, and varied organizational readiness (Ayo-Farai et al., 2024). Within this study's tri-country focus, Kenya has moved furthest in formalizing digital health and telemedicine through the Kenya National eHealth Policy 2016–2030 and the Data Protection Act (2019), which together frame teleconsultations as high-risk digital health services that must comply with strict data protection safeguards (Munyolo, 2021). Uganda and Tanzania have introduced broader data protection and health-information reforms, but socio-legal reviews note that their telemedicine deployments remain more pilot-driven, with weaker harmonization of licensing, consent and cross-border data-transfer rules compared to Kenya, leaving providers to navigate a patchwork of obligations when collaborating across borders.

1.1.3 Machine Learning in Telemedicine

As telemedicine platforms mature, there is increasing interest in embedding machine learning (ML) models to support triage, risk prediction, and clinical decision support using electronic health records (EHRs), sensor data, and imaging. Supervised models such as random forests, support vector machines, artificial neural networks, and decision trees have been widely used for medical diagnosis and prognosis because of their ability to learn from structured EHRs and large clinical datasets (Habehh & Gohel, 2021). In telemedicine contexts, these models can

automate parts of screening, prioritize high-risk patients, and generate predictive insights that are difficult to obtain with manual review alone. However, traditional ML workflows usually assume that data from multiple facilities is centralized in a single repository, which raises concerns about data sovereignty, regulatory compliance, and systemic cybersecurity risk when such repositories span jurisdictions.

1.1.4 Concepts of Federated Learning and Their Relevance in Telemedicine

Federated Learning (FL) has emerged as a promising paradigm for training ML models collaboratively without pooling raw patient records in a single database. In cross-institutional healthcare settings, each hospital trains a local model on its own data and only shares model updates or predictions, while a coordinating service aggregates these signals into a global model (Mensah, 2024; Alkhalifa et al., 2024). This approach is attractive for East African telemedicine because it can respect data localization requirements by keeping raw records inside each country or facility while still allowing shared learning. Case studies have shown that FL can improve diagnostic performance in tasks such as diabetic retinopathy screening and respiratory disease detection without centralizing EHRs or images (Matagi & Kaneko, 2023; Mensah, 2024). Nevertheless, FL by itself does not fully eliminate privacy risk, as model updates and outputs can still leak sensitive information through advanced inference attacks, and its legal status under existing health and data protection frameworks remains under-explored.

1.1.5 The Law, Privacy and Telemedicine Across Borders

Cross-border telemedicine introduces additional governance complexity because patient data and model outputs may move between jurisdictions with different regulatory standards.

Kenya's Data Protection Act (2019) classifies health information as sensitive personal data and imposes strict safeguards for cross-border transfers, including requirements for adequacy, appropriate safeguards, and demonstrable privacy-by-design. At the same time, regional integration efforts mean Kenyan telemedicine providers increasingly interact with partners in countries whose laws may not fully align with the DPA or the EU's General Data Protection Regulation (GDPR) (Corrales Compagnucci & Fenwick, 2024). Global frameworks such as GDPR and HIPAA in the United States emphasize principles like data minimization, purpose limitation, and strong anonymization or de-identification before data is reused or shared. Emerging analyses of African digital health environments highlight tensions between the commercial growth of digital health, the use of foreign cloud infrastructure, and capacity gaps in enforcement of privacy rules (Kitili, 2023). In this context, federated learning combined with formal privacy technologies such as differential privacy offers a potential pathway to operationalize "privacy by design" and cross-border data governance, but there is limited evidence on how such approaches can be tailored and audited for East African telemedicine.

1.2 Statement of The Problem

Telemedicine in Kenya and the wider East African region is expanding under supportive policy frameworks, national digital health strategies and growing demand for remote care. However, most existing deployments either rely on centralized EHR systems, which concentrate sensitive data on a single platform, or on vendor-controlled cloud infrastructures hosted outside national borders, complicating data sovereignty and legal accountability (Munyolo, 2021). At the same time, there is increasing interest in including machine learning models into telemedicine workflows, yet conventional centralized training requires aggregating EHRs from multiple

hospitals into one environment, which is difficult to reconcile with Kenya's Data Protection Act, privacy-by-design expectations, and cross-border transfer safeguards. Federated learning offers a good alternative by keeping data local while sharing model signals, but by itself it does not provide a complete privacy or governance solution. Research has shown that even when raw data remains local, model updates and outputs can still leak information through membership inference and model inversion attacks, raising questions about how telemedicine providers can demonstrate that data reuse and cross-border analytics remain within acceptable privacy risk bounds. In the Kenyan context, there is limited guidance and almost no empirical evidence on how a federated learning–based telemedicine workflow could be made auditable, measurable, and aligned with Kenya's DPA, GDPR transfer principles, and sector-specific health regulations. The specific problem this study addresses is therefore the absence of an auditable, privacy-preserving data governance model for Kenyan cross-border telemedicine that combines federated learning with formal privacy guarantees and explicit alignment to applicable legal frameworks. Existing systems and policies do not yet provide a tested technical workflow that hospitals can adopt to collaborate on predictive models across borders while keeping raw EHRs in-country, quantifying privacy risk, and demonstrating compliance.

1.2.1 Knowledge Gap

Telemedicine and digital health in Kenya and Eastern Africa have been examined from infrastructural, organizational, and policy perspectives, and what has been noted are implementation barriers, fragmentation of regulations, and the need for harmonized digital health laws (Ayo-Farai et al., 2024). Second, machine learning in healthcare has been widely explored using centralized models such as random forests, support vector machines and deep neural

networks for tasks like diagnosis, risk prediction, and imaging analysis, but often assumes that data from multiple sites can be pooled in one repository (Habehh & Gohel, 2021). Third, emerging work on federated learning in African or Kenyan settings demonstrates that FL can support privacy-preserving diagnostics and maintain data localization, yet these studies largely focus on model utility or feasibility rather than on formal privacy accounting or governance (Matagi & Kaneko, 2023). Finally, there is a growing body of research on differential privacy and privacy attacks against ML models, but little of it is operationalized within East African telemedicine or mapped explicitly to Kenyan legal obligations. As a result, there is a specific gap at the intersection of telemedicine data governance, federated learning, and formal privacy guarantees, no existing study has developed and empirically evaluated an end-to-end workflow in which Kenyan and regional telemedicine nodes train models locally on synthetic EHRs, share only privacy-bounded signals, simulate realistic attacks, and log privacy budgets and governance metrics in a way that auditors and regulators can interpret. This thesis therefore seeks to fill that gap by proposing and testing a privacy-preserving federated learning (PPFL) model tailored to Kenyan cross-border telemedicine data governance.

1.3 Main Objectives of The Study

To develop a privacy-preserving federated learning–based data governance model for Kenyan cross-border telemedicine that aligns with national and international data protection requirements.

1.3.1 Specific Objectives

- To analyze the existing legal, governance and infrastructural context for cross-border telemedicine data exchange between Kenya and selected East African partners, with a focus on implications for machine learning and federated learning.
- To design a privacy-preserving federated learning architecture for cross-border telemedicine that incorporates differential privacy and is conceptually aligned with Kenya's Data Protection Act (2019), GDPR, and HIPAA safeguards.
- To implement a PPFL architecture using synthetic Synthea™ EHR datasets partitioned across multiple hospital nodes representing Kenya, Tanzania and Uganda.
- To evaluate the proposed model in terms of predictive utility, resistance to membership inference and model inversion attacks, and its ability to provide auditable evidence of compliance with applicable data protection and health data governance requirements.

1.4 Research Questions

- How do current legal, governance and infrastructural arrangements for cross-border telemedicine in Kenya and selected East African countries affect the design of machine learning and federated learning–based data governance models?
- In what way can a federated learning architecture incorporating differential privacy be designed to support privacy-preserving cross-border telemedicine while remaining aligned with Kenya's Data Protection Act, GDPR, and HIPAA safeguards?
- How can the proposed privacy-preserving federated learning model be implemented and simulated using synthetic EHR data distributed across hospital nodes that represent Kenyan and regional telemedicine partners?

- How does the proposed model perform in terms of predictive utility, measured privacy leakage under attack simulations, and the provision of auditable governance artefacts such as privacy budgets and compliance logs?

1.5 Expected Outcomes

The expected outcomes of this research are:

- A thorough report on Kenya's telemedicine current cross border governance and regulatory frameworks.
- A federated learning-based framework designed for safe and legally sound cross-border healthcare data governance.
- A federated learning prototype framework implementation examined for accuracy, robust security and compliance.
- Policy best practices to enhance telemedicine data governance in Kenya.

1.6 Significance of The Study

Implementation of federated learning will improve privacy of data, privacy of patient's compliance to regulations. A huge array of healthcare professionals, regulatory firms, and other professionals will benefit from the findings of this research which will boost enhancing frameworks that are standardized and comply with both national and international laws and regulations. Confidentiality of patient data, secure data management will also be guaranteed through utilization of federated learning, issues of unauthorized access will also be reduced as this study will help shape governance policies that address data breaches. This will promote trust in the telemedicine platforms.

1.6 Scope of the study

This study focuses on data governance and privacy-preserving machine learning for cross-border telemedicine involving Kenyan providers and selected East African partners. It is limited to a simulated environment using synthetic Synthea™ electronic health records that approximate telemedicine-relevant variables but do not represent real patients or clinical workflows. The federated learning setup models a small network of hospitals in Kenya, Tanzania and Uganda to illustrate cross-border collaboration under data localization and privacy constraints. The technical scope is confined to structured EHR data, Random Forest classifiers, probability-level aggregation, and Gaussian mechanism differential privacy. The study does not deploy the model in live clinical systems, does not evaluate other advanced cryptographic approaches such as homomorphic encryption, and does not attempt to redesign national policy. Instead, it provides a technically grounded proof-of-concept and governance template that can inform future pilots under real regulatory and organizational conditions.

1.7 Limitations of the study

A few constraints arise, first, the study relies entirely on synthetic Synthea™ EHR data, which, while ethically appropriate and structurally realistic, may not capture all disease patterns. Second, the federated learning environment is simulated in software rather than deployed across actual hospital infrastructure, so network constraints, organizational workflows, and human factors are approximated. Third, only one model family (Random Forest) and one formal privacy mechanism (Gaussian differential privacy with a limited ε range) are evaluated. Finally, the threat model is restricted to black-box membership inference and model inversion attacks and an honest-

but-curious coordinator, more powerful white-box or colluding adversaries are left for future research.

CHAPTER TWO

2. LITERATURE REVIEW

2.1 INTRODUCTION

Discussed in this chapter is federated learning, machine learning models, uses, limitations and benefits in the telemedicine industry. Comparison of centralized and decentralized systems in terms of privacy, security and compliance with regulations is also explored. Lastly an overview of diagnostics, medical imaging, predictive analysis is also looked at while also highlighting supervised and unsupervised models and critical research gaps.

2.2 Overview of Machine Learning Models in Healthcare

Machine Learning has revolutionized healthcare in areas such medical research and illness detection. This is through analysis of datasets, patterns and predictive insights. We explored some machine learning models used in telemedicine and health sector, their applications and limitations.

2.2.1 Supervised Learning Models for Medical Diagnosis

Supervised models train algorithms to diagnose diseases and predictions using labelled information. (Habehh & Gohel, 2021) They depend on organized input output interlinkages to achieve their prediction abilities. Commonly used supervised techniques include random forests, support vector machines, artificial neural networks and decision trees. Decision trees assist to achieve classification results by providing different diagnostic paths. Random forests limit overtraining and with the help of an ensemble of decision tress increase the accuracy of predictions. Support vector machines work well with medical complications that have well defined dissociability. E, g. identification of malignant tumors. Neural networks which replicate human

brain neurons perform best at image analysis and predictions of illnesses. Convolutional Neural Networks has also improved classification of tumors and fractures, and analysis has shown that the accuracy of diagnosing breast cancer from mammograms has improved by 11.5% using deep learning models. Habehh and Gohel (2021) note than decision tree models' data in diagnosing depression has 87.27% accuracy and 89.47% accuracy in prediction of therapy responses.

2.2.2 Unsupervised Learning for Patient Clustering

Unsupervised machine learning models scrutinize patient data without perceived labels to find hidden patterns and come up with patient groupings compared to supervised learning models. This makes unsupervised learning a more popular approach among healthcare professionals. The k-means algorithm finds centroids that are in recognizable groups of patients during clustering of patients. During enhancing of these centers, the k-means classifies patients with shared traits and behaviors which may include outcomes of treatment and symptoms during treatment. This promotes control of diseases and improves personalized treatment of patients. Two unsupervised learning methods: convolutional neural networks and deep belief networks are utilized in medical picture categorization and anomaly detection. Deep belief networks are recommended for disease classification because of their multi layered nature, which makes it possible for them to extract information from medical data. On the other hand, convolutional networks are becoming popular in healthcare facilities with their main uses being patient segmentation, disease diagnosis and identification of images. One shortcoming of unsupervised models is it becomes difficult to understand findings without domain knowledge of evaluating clustering results. (Habehh & Gohel, 2021).

2.3 Telemedicine and Digital Health Environment in Eastern Africa and Kenya

In sub-Saharan Africa, telemedicine is increasingly viewed as a way to bridge geographical barriers, chronic shortages of specialists, and weak referral systems. Rapid reviews and systematic surveys show that although telemedicine projects are growing across the region, their deployment remains uneven, often limited to urban pilot sites, and constrained by affordability, connectivity, and governance challenges (Chitungo et al., 2021). Within Eastern Africa, telemedicine programs have been documented in Kenya, Uganda, Tanzania, Rwanda, and Ethiopia, typically in the form of teleconsultations, teleradiology, and mobile health platforms for chronic disease management. Comparative reviews highlight that while these initiatives improve access and continuity of care, they frequently struggle with technical reliability, lack of interoperability, and fragmented regulatory oversight (Dodoo et al., 2021). Many facilities rely on third-party cloud infrastructure and consumer-grade communication tools, raising concerns about confidentiality, lawful cross-border data transfers, and long-term sustainability of donor-driven projects. Kenya has positioned itself as a regional leader in digital health through the Kenya National eHealth Policy 2016–2030, which explicitly promotes telemedicine and cross-border sharing of health information "without compromising patient privacy" and calls for secure transfer of health data as a policy priority. The policy is complemented by the Data Protection Act (2019), which regulates processing and cross-border transfer of personal data, and by the emerging Digital Health Act (2023) and associated regulations, which establish a digital health information system, authentication requirements, and standards for encryption and secure storage of health records (Republic of Kenya, 2016; Digital Health Act, 2023). These instruments recognize telemedicine as a high-risk digital health service that must comply with data protection principles such as lawfulness, purpose limitation, data minimization, and security safeguards. However, cross-border telemedicine involving Kenyan

patients often extends beyond Kenya's jurisdiction. Socio-legal analyses of digital health in Africa argue that many cross-border arrangements operate in a fragmented regulatory terrain where data protection, professional licensing, and liability rules differ across countries (Sekalala et al., 2025). For example, while Kenya and Uganda now have comprehensive data protection laws and emerging digital health regulations, other East African Community (EAC) states are still developing equivalent frameworks or enforcement capacity, resulting in disparities in patient safeguards and oversight of foreign service providers. This lack of harmonized governance for cross-border health data flows emphasizes the need for technical solutions such as federated learning and differential privacy that can enforce data minimization, and privacy guarantees even when legal protections are uneven. This study positions itself precisely at this intersection between technical privacy mechanisms and regional regulatory gaps.

2.4 Comparison of Centralized vs. Decentralized Learning Approaches

Centralized learning has traditionally been the dominant paradigm in healthcare machine learning, where data from multiple facilities are aggregated into a single repository for model training. This approach benefits from large, pooled datasets that often yield high diagnostic accuracy for tasks such as tumor detection, risk prediction, and medical image classification. However, concentrating sensitive electronic health records (EHRs) in one location creates an attractive target for cyber-attacks. A successful breach can expose thousands of patient records, triggering legal liabilities under frameworks such as Kenya's Data Protection Act (2019), the EU's GDPR, and HIPAA's security and breach-notification rules. Decentralized approaches such as federated learning and edge computing emerged precisely to mitigate these risks. Instead of sending raw EHRs to a central server, hospitals or edge devices train models locally and share only

model parameters or predictions. This preserves data sovereignty—which is critical where laws require data residency within national borders while still enabling collaborative model building across institutions. In the East African context, this is particularly relevant because differing national regulations complicate cross-border transfers of health data. At the same time, decentralized learning introduces new cybersecurity and governance challenges: coordinating updates from heterogeneous sites, verifying the integrity of model contributions, and protecting models from poisoning, inversion, and membership-inference attacks. Recent surveys emphasize that, without explicit privacy and robustness mechanisms, federated systems can still leak sensitive information through gradients or model outputs. For this reason, modern privacy-preserving designs increasingly treat "centralized vs. decentralized" as necessary where federated learning is combined with formal privacy guarantees, secure aggregation, and auditable governance controls. This study positions itself within that space as it evaluates how decentralized learning, strengthened by differential privacy and regulatory safeguards, can reduce the attack surface and legal exposure of cross-border telemedicine in Kenya and neighboring countries.

2.5 Differential Privacy Framework in Healthcare Machine Learning

Differential Privacy (DP) offers a mathematically rigorous way to limit what an adversary can infer about any individual from model outputs, even when they have auxiliary information. In healthcare, DP has been proposed as a countermeasure to membership-inference and model-inversion attacks that can re-identify patients from trained models or aggregated statistics. Recent surveys on medical DP show its growing adoption in clinical prediction, image analysis, and genomic studies, but also highlight the difficulty of tuning the privacy budget ε without degrading clinical utility. In practice, DP mechanisms inject calibrated noise into gradients, parameters, or

query results such that the contribution of any single patient becomes statistically indistinguishable. Regulatory frameworks increasingly expect such formal privacy guarantees. The GDPR requires robust anonymization or pseudonymization before personal data can be reused or transferred across borders, while HIPAA's "safe harbor" and "limited dataset" provisions permit de-identified health data only under strict conditions. Kenya's Data Protection Act (2019) similarly emphasizes data minimization, privacy-by-design, and heightened safeguards for health data, and African comparative analyses show a trend toward converging on these principles. Within this context, DP can be seen as a technical instantiation of legal ideas such as "data minimization" and "protection against re-identification", particularly when combined with de-identification and ISO/TS 25237 pseudonymization practices already used in EHR management. However, DP is not a silver bullet. Stronger privacy (low ε) typically reduces model accuracy or recall, which can be problematic in high-stakes clinical tasks. There is also a computational cost: DP-SGD and related algorithms increase training time and may be difficult to deploy on resource-constrained devices common in African health facilities. Moreover, most DP studies are carried out on static datasets in well-resourced settings and rarely incorporate operational governance, such as logging privacy budgets, model hashes, and attack outcomes for audit. Emerging work on "privacy auditing" argues that DP systems should be accompanied by structured logging and attack-based evaluations to verify that privacy guarantees hold under realistic adversarial conditions. This study responds to that gap by not only applying DP within a federated setting but also quantifying its effect on attack success and recording ε, δ, clipping, and noise parameters in an auditable ε-register.

2.6 Federated Learning in Healthcare

Federated Learning (FL) enables multi-institutional model training without centralizing raw patient data. Hospitals or devices compute local model updates on their EHRs and periodically share these updates with a coordinating server, which aggregates them into a global model. This paradigm has been applied to diagnostic imaging, risk prediction, and personalized treatment, with several surveys confirming that FL can reach accuracy close to centralized training while better preserving data locality. For jurisdictions such as Kenya that prioritizes data residency, FL aligns well with statutory requirements to keep sensitive health records within national borders while still benefiting from regional or international collaboration. At the same time, FL introduces its own cybersecurity and governance risks. Because the coordinator never sees raw data, it must trust that local updates are honest and correctly computed; adversarial clients can poison models, insert backdoors, or perform free-riding attacks. Moreover, gradients and model parameters themselves can leak information about individuals, especially when client datasets are small or skewed. Hence, recent work combines FL with DP, secure aggregation, or homomorphic encryption to limit leakage and strengthen robustness. Governance-focused studies also stress the need for clear accountability structures, audit trails, and legal agreements defining controller–processor roles when multiple hospitals co-train models. In resource-constrained and cross-border environments such as East Africa, additional constraints arise intermittent connectivity, heterogeneous hardware, and fragmented regulatory regimes. Studies of FL deployments in rural clinics and mobile environments show that lightweight models and adaptive communication schedules are needed to cope with bandwidth limitations, but they rarely integrate detailed legal requirements or differential privacy into their design. This study builds on that body of work by evaluating a

federated ensemble that is explicitly guided by Kenyan, GDPR, and HIPAA safeguards, and by embedding DP as a primary security control rather than treating FL alone as sufficient.

2.7 Existing Federated Learning Applications in Healthcare

Existing FL applications in healthcare demonstrate that collaborative modelling can improve clinical performance while reducing the need to centralize data. In medical imaging, multi-site FL has been used for tumor segmentation, chest-X-ray classification, and dermatology, often achieving comparable accuracy to centralized baselines while keeping images on-premises. Similar approaches have been reported for diabetic retinopathy screening and other ophthalmology tasks, where hospitals jointly train models but retain raw retinal images locally. These studies show that FL is technically mature enough for high-dimension medical images, but most are situated in Europe, North America, or East-Asian health systems with well-established digital infrastructure. Beyond imaging, FL has been applied to EHR-based prediction and chronic disease management. Systems such as Health-FedNet and FED-EHR demonstrate how hospitals can collaboratively train models for readmission risk, heart-failure prediction, or medication recommendation while maintaining data locality and complying with GDPR or HIPAA requirements. Other work integrates Internet-of-Medical-Things (IoMT) devices and wearables, where on-device FL supports continuous monitoring of heart rate, glucose levels, or respiratory patterns without uploading raw sensor streams. These systems highlight the potential of FL for personalized and longitudinal care, especially when connectivity is intermittent. Several authors explicitly explore FL as a privacy-preserving or regulatory technology. Surveys and frameworks propose combining FL with DP, secure aggregation, or blockchain-based consent ledgers to strengthen confidentiality and accountability. For example, blockchain-integrated FL architectures log each model update

and consent transaction immutably, allowing auditors to reconstruct who contributed what and under which legal basis—an approach that resonates with auditability expectations under GDPR, HIPAA, and emerging African data protection laws. Yet, as recent governance reviews note, many FL projects still under-specify how legal roles (controller vs. processor), cross-border transfer mechanisms, and privacy budgets are operationalized in practice. Crucially, there is limited work that targets cross-border telemedicine in Sub-Saharan Africa specifically. Available telemedicine reviews emphasize infrastructural deficits, fragmented regulation, and trust deficits in South Africa, Kenya, and Nigeria, but seldom propose concrete FL+DP governance models tailored to those realities. Most technical papers assume relatively homogeneous legal regimes and stable connectivity, conditions that do not always hold for Kenyan cross-border teleconsultations. This study therefore extends existing FL applications by situating a privacy-preserving federated ensemble within East Africa's regulatory, infrastructural, and cybersecurity context and by explicitly evaluating privacy leakage under adversarial attack.

2.7.7 Case Studies on Federated Learning in Cross-Border Telemedicine

Training models utilizing decentralized approaches has proven to effectively and efficiently improve healthcare delivery. This is through using distributed systems to integrate data sources. A study conducted in Kenya explored diabetic retinopathy by using federated learning on telemedicine applications. This made possible real time diagnostics independent of computational infrastructure that is complex (Matagi & Kaneko, 2023). The finding revealed diabetic retinopathy was successfully identified in its preliminary stages with a 15% improvement in accuracy compared to past models. Kenya's telemedicine landscape faced significant hurdles due to overdependence on data storage facilities outside its borders. According to Mensah (2024), a

federated learning approach made it possible to train machine learning modes locally while only sharing model updates that were anonymized. Compliance with Kenya's Data Protection Act required the maintenance of patient data within national borders, which effectively minimized the risks associated with cross-border data transfers. The integration of federated learning and blockchain technology presents a strong approach to efficiently managing consent. The use of lightweight federated learning within rural telemedicine focuses on innovative strategies that improve healthcare access and delivery in underserved regions. The potential for enhanced patient outcomes and greater accessibility using advanced technologies is greatly emphasized, while also tackling the specific challenges encountered in rural areas. Researchers have effectively deployed lightweight FL models on smartphones and low-power medical devices, facilitating the diagnosis of respiratory illnesses in rural health centers throughout Kenya (Mensah, 2024). The findings indicate that AI models developed in Florida achieved a diagnostic accuracy of 87%, exceeding the performance of conventional machine learning techniques. AI-powered screening tools that are affordable have reduced patients' dependence on urban hospitals. The implementation of on-device data processing has resulted in a significant 30% decrease in latency. Data aggregation that is secure, along with differential privacy, is essential for AI training across borders. One major concern in telemedicine AI is the potential for data re-identification attacks, where malicious actors attempt to extract sensitive patient information from combined AI models. The integration of federated learning with differential privacy enhances security measures significantly (Mensah, 2024). Recent work has demonstrated that federated learning can enable multi-institutional training of clinical models without centralizing raw patient data. Adnan et al. (2022) trained a histopathology image classifier using data from several hospitals and showed that a federated learning set-up, combined with differential privacy, significantly reduced membership-inference

attack success rates while maintaining clinically acceptable AUC values. Their study is important because it treats privacy as a quantifiable risk that can be controlled through a tunable privacy budget ε, very similar to how this stufy calibrates ε and σ across different training rounds. Further case studies extend federated learning to other medical imaging tasks. Nampalle et al. (2023) surveyed DP-enabled federated learning systems and highlighted that adding Gaussian or Laplace noise to gradients can meaningfully lower information leakage but may reduce model accuracy if ε is set too low, emphasizing the need for careful privacy–utility trade-off analysis exactly the trade-off measured in this study. Fares and Sertbaş (2024) implemented a differentially private federated learning scheme for medical imaging and empirically compared multiple noise levels and clipping strategies. Their results show that attack success rates, including model inversion, fall towards random-guess baselines once ε is kept below approximately 1, but utility degrades sharply for very small ε. Zheng et al. (2025) proposed a "sensitivity-aware" DP mechanism that allocates stronger privacy protection to more sensitive features in federated medical sensing, suggesting that privacy budgets can be dynamically tuned rather than fixed globally. Across these case studies several patterns emerge that directly inform this thesis. First, almost all implementations are based in high-income settings, and none explicitly targets cross-border telemedicine in low- and middle-income countries, regulatory mapping to instruments such as the Kenya Data Protection Act, GDPR adequacy rules, or HIPAA is rarely discussed. Second, many works evaluate privacy using generic attack scenarios without linking them to concrete governance obligations, such as logging ε and attack outcomes for audit. Third, the dominant use-cases are medical imaging, there is limited work on electronic health records and multi-country teleconsultation workflows. These gaps justify your contribution about designing and evaluating a differentially private federated learning framework grounded in Kenya's and the EAC's regulatory context, using synthetic EHR features

and attack simulations that speak directly to data protection and cybersecurity requirements in cross-border telemedicine.

2.8 Summary table

Table 2.1 Summary table of Key Findings 1

Author	Topic	Key Findings
Habehh & Gohel (2021)	Supervised Learning in Healthcare	CNNs, SVMs, and decision trees are highly effective in medical diagnosis, predictive analytics, and medical imaging.
	Deep Learning in Medical Imaging	CNNs improve tumor classification, lesion detection, and fracture identification. Deep learning models increased breast cancer diagnostic accuracy by 11.5%.
	Unsupervised Learning in Patient Clustering	k-Means and DBNs cluster patients by symptoms, genetic traits, and treatment responses, enabling personalized medicine.
	Challenges of Unsupervised Learning	Unsupervised learning models are difficult to integrate into clinical decision-making due to the lack of labeled data.
Munyolo (2021)	Cybersecurity Risks in Centralized Learning	Centralized databases are vulnerable to cyberattacks,

		increasing risks of data breaches in healthcare.
	Regulatory Compliance Concerns	Compliance with HIPAA (USA), GDPR (EU), and Kenya's Data Protection Act (2019) remains a key challenge.
	Federated Learning as a Cybersecurity Solution	FL mitigates security threats by keeping data within local jurisdictions while enabling AI training.
Corrales Compagnucci & Fenwick (2024)	Cross-Border Data Governance	Different national data laws complicate AI-driven healthcare collaborations.
	FL for Regulatory Compliance	FL enables privacy-preserving AI collaborations by ensuring data remains within legal jurisdictions.
	Blockchain for Consent Management	Blockchain-integrated FL ensures secure, immutable patient consent tracking for compliance.
Matagi & Kaneko (2023)	FL in Kenyan Telemedicine	FL enables Kenyan healthcare providers to train AI models locally, ensuring compliance with Kenya's Data Protection Act (2019).
	FL in Diabetic Retinopathy Detection	FL-based AI models improved diabetic retinopathy diagnostics by

		15% while maintaining data sovereignty.
	FL and Edge Computing	Edge-based FL reduces latency in AI-powered diagnostics, allowing real-time patient monitoring.
Mensah (2024)	FL in Personalized Treatment	AI models using FL predict drug responses, improving precision medicine.
	FL for Predictive Analytics	FL-based AI models enhance COVID-19 outbreak modeling and cancer risk prediction.
	FL in Rural Healthcare	Lightweight FL models on mobile devices achieved 87% diagnostic accuracy for rural respiratory illness detection.
	Differential Privacy in FL	FL integrated with differential privacy reduced data re-identification risks by 40%, ensuring GDPR and HIPAA compliance.
Afrin et al. (2024)	FL in Predictive Analytics	FL-based models enhance chronic disease prediction while preserving data privacy.
	FL and IoMT Integration	IoMT devices integrated with FL enable real-time monitoring of heart rate, glucose levels, and blood pressure.

	FL and Computational Challenges	High computational demands in developing regions require optimized edge computing solutions.
Alkhalifa et al. (2024)	FL and Blockchain for Healthcare Security	FL with blockchain technology creates tamper-proof logs of AI model updates and patient consent.
	FL in Cross-Border Healthcare Collaboration	FL enabled secure AI training across Kenya, Uganda, and Tanzania without violating privacy laws.
	Differential Privacy in FL	FL combined with differential privacy anonymized model updates, preventing data exposure while improving AI reliability.

2.8.1 Research Gaps

Most FL healthcare studies focus on technical performance accuracy, F1-score, or latency without rigorously quantifying privacy leakage under realistic attacks such as membership inference or model inversion, and without treating ε and σ as auditable governance parameters. While DP surveys acknowledge the tension between utility and privacy, they are largely situated in well-resourced European or North American contexts and seldom consider low-resource African settings where datasets are smaller, class distributions more skewed, and digital infrastructure less reliable. Little work integrates FL and DP into a data-governance framework that is explicitly mapped to Kenya's Data Protection Act (2019), GDPR cross-border transfer

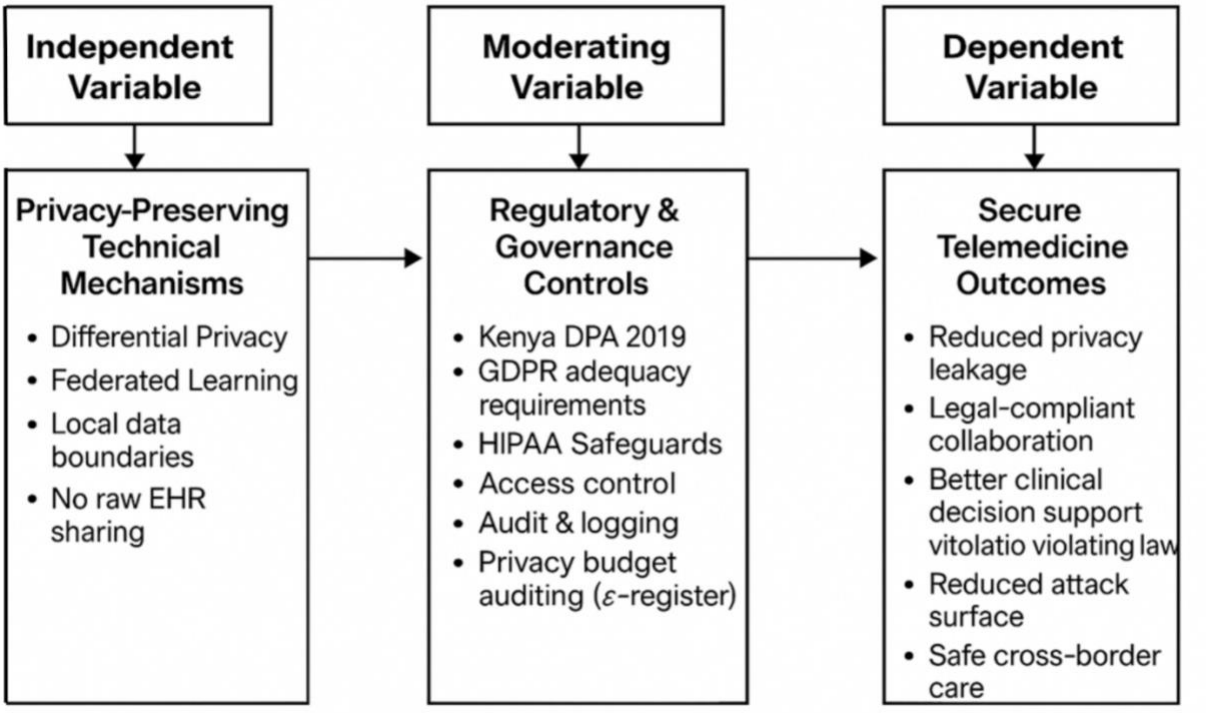
principles, and HIPAA safeguards; existing governance reviews tend to be conceptual and are not instantiated in executable, attack-tested models. There is a scarcity of empirical studies modelling East-African cross-border telemedicine workflows where hospitals in Kenya, Tanzania, and Uganda collaborate while keeping raw EHRs in-country and sharing only model signals. Available African telemedicine literature mainly documents adoption challenges, policy fragmentation, and infrastructure gaps rather than proposing operational FL+DP solutions for those scenarios. There is also limited guidance on how hospital CISOs and regulators should monitor the ongoing privacy posture of FL systems, for example through ε-registers, attack-aware dashboards, or compliance scorecards aligned with statutory requirements.

2.9 Conceptual Model

The conceptual model for this study links privacy-preserving technical mechanisms, regulatory and governance controls, and secure telemedicine outcomes in Kenyan cross-border telemedicine. The independent variable are technical privacy mechanisms implemented at the model level that operationalize the data minimization and security-by-design principles that are demanded by contemporary data protection regimes but are often missing from conventional centralized machine-learning pipelines. The moderating variable consists of regulatory and governance controls that shape how the technical mechanisms are deployed and audited. At the legal layer, this includes the Kenya Data Protection Act (2019), GDPR cross-border adequacy requirements, and HIPAA's safeguards for protected health information. At the governance layer, it covers access control, role-based permissions for model access, and detailed logging of ε, δ, clipping parameters, model hashes, and attack scores in an ε-register that can be inspected by hospital CISOs or regulators. Kenya's National eHealth Policy and the Digital Health Act

emphasize encryption, secure transfer of health information, and accountability for digital health platforms, while regional legal scholarship stresses that such safeguards must work in fragmented regulatory terrains where not all EAC states have equivalent protections. In the model, these regulatory controls do not generate privacy on their own rather, they moderate the effect of DP and FL by determining how strictly privacy budgets are selected, how attack metrics are interpreted, and how non-compliant configurations are sanctioned. The dependent variable captures secure telemedicine outcomes. These outcomes are expressed in terms of (i) reduced privacy leakage, measured quantitatively by bringing membership-inference and inversion attack metrics close to random-guess baselines; (ii) legally compliant cross-border collaboration, assessed against KDPA, GDPR, and HIPAA principles; (iii) improved clinical decision support without unlawful re-identification or over-collection of data; and (iv) a reduced attack surface for adversaries who might attempt to exploit model updates or communication channels. Prior studies show that federated learning alone cannot fully prevent gradient-based attacks or regulatory non-compliance; combining it with differential privacy, explicit privacy budgets, and audit trails is necessary to move from "privacy by design" to measurable privacy risk reduction. The conceptual model justifies the study's focus on evaluating how different ε values, clipping parameters, and governance settings jointly influence both attack success rates and compliance-oriented security indicators in Kenyan cross-border telemedicine.

Figure 2.1 Conceptual Framework 1



CHAPTER THREE

3.METHODOLOGY

3.1 Introduction

This chapter seeks to explain how the study is carried out, the philosophy guiding this study, research design being used, stakeholders involved, how data is gathered and analyzed and lastly the ethical safeguards put in place.

3.2 Research Philosophy

This study is anchored in the pragmatic research philosophy. It assumes that reality is multi-layered and includes both technical artefacts (models, algorithms, systems) and the social–legal environment in which they operate. from several angles. Privacy risk in cross-border telemedicine is treated as something that emerges from how technical controls, hospital workflows, and regulations interact, rather than from technology alone. The study is value-driven by the ethical obligation to uphold confidentiality, respect for patient autonomy, and regulatory compliance under Kenya's DPA 2019, the GDPR, and HIPAA.

3.3 Research Design

By integrating quantitative and qualitative methods into a unified framework, this study is supporting a hybrid design. The quantitative core of the design is an experimental simulation that compares three setups: (i) centralized machine learning on pooled synthetic EHRs, (ii) federated learning without differential privacy, and (iii) federated learning with client-side differential privacy. Each setup is evaluated using standard classification metrics (accuracy, recall, F1-score, AUC) and attack outcomes (model-inversion success and membership-inference AUC). These

experiments are used to investigate whether the proposed framework can reduce privacy leakage while maintaining clinically usable utility. On the qualitative side key regulatory instruments (Kenya DPA 2019, GDPR, HIPAA) and recent African telemedicine literature are reviewed to derive privacy and security requirements. These requirements guide the design of the threat model, privacy budget register, and compliance checklist.

3.3.1 Implementation Process

The implementation follows a phased process. The first step is to translate Kenya's DPA 2019, the GDPR, and HIPAA privacy and security provisions into concrete technical requirements for cross-border telemedicine. A high-level threat model is developed covering an honest-but-curious coordinator, an external black-box attacker querying model outputs, and the risk of a compromised hospital node. Synthetic patient-level electronic health records are generated using the Synthea FHIR simulator to represent Kenyan hospital populations. The data is cleaned and harmonized into a tabular dataset with clinically meaningful features such as age, sex, diagnosis codes, comorbidities and outcome labels. The dataset is then partitioned into seven virtual hospitals (nodes) with non-identical sample sizes to mimic heterogeneous real-world facilities. A Random Forest classifier is implemented as the base predictive model using scikit-learn. Centralized training used pooled data from all nodes. For the federated configuration, model training was implemented using TensorFlow Federated (TFF), where each virtual hospital trains a local model on its own records and shares only model outputs needed for aggregation. Because tree ensembles do not aggregate well via parameter averaging, the study uses probability-level fusion, combining local prediction probabilities with data-size weights to form a global prediction. Client-side differential privacy is then implemented using a Gaussian DP mechanism. Per-sample gradients

(or equivalent local statistics) are clipped at an L2 norm C and perturbed with zero-mean Gaussian noise with standard deviation σ. A privacy accountant is used to map (σ, C, sampling rate, number of steps) into an (ε, δ)guarantee, and the resulting ε, δ, σ, C and training steps are logged in an ε-register for audit. Finally, model inversion and membership-inference attacks are simulated in centralized, FL, and FL+DP setups. Utility metrics and attack success scores are recorded across different ε values to study the privacy–utility trade-off. Results are then mapped back to the regulatory requirements and threat model.

3.3.2 Study context

The study is situated in the context of cross-border telemedicine involving Kenyan hospitals and foreign specialist centers. Many Kenyan facilities increasingly rely on cloud-hosted platforms and remote expertise, which often require health data or derived models to cross jurisdictional boundaries. Direct experimentation on real Kenyan patient data would have required complex approvals and introduced additional privacy risk. To remain within ethical and legal limits for a master's project, the study therefore uses synthetic EHR data that statistically resemble realistic hospital populations while containing no identifiable individuals. Seven virtual hospitals were configured to reflect a mix of large referral hospitals, county-level facilities, and smaller mission or private hospitals. Each node holds different volumes and distributions of synthetic records to reproduce the non-IID nature of real health data across institutions. The cross-border scenario is represented by training taking place inside Kenyan hospital boundaries while the coordinating server, in principle, could be located in a foreign cloud region. This context allows the study to analyze privacy, attack resistance, and legal alignment under realistic data flows, while

recognizing that synthetic data limit external validity to real-world deployments. The simulation is therefore framed as a proof-of-concept that demonstrates feasibility.

Table 3.1 Federated Privacy-Workflow 1 1

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 external eavesdropper threat on the	Encrypted update at server	Handshake logs; integrity checks

			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

3.4 Target Population

The target population for this study comprises Kenyan telemedicine ecosystems that exchange clinical data or models across borders, including public and private hospitals, telehealth platforms, and their regulators. Within the simulation, this population is represented by seven virtual hospitals holding synthetic EHRs and a central coordinating server that aggregates model updates. From a governance perspective, the population also includes data protection officers, hospital CISOs, and

regulators such as the Office of the Data Protection Commissioner (ODPC), whose expectations inform the privacy and compliance controls evaluated in the framework.

3.5 Sampling Techniques

Given that the primary evidence is generated through simulation rather than direct human subjects, probabilistic sampling of patients was not applied. Instead, the synthetic generator was configured to produce a census of all cases used in the experiments, and all 3,459 generated records were retained in the analysis. The seven hospital nodes were selected using purposive sampling to represent different sizes and types of facilities that are common in Kenyan telemedicine referrals (large national referral, regional county hospitals, and smaller mission/private facilities). For the qualitative and legal–governance strand, a purposive document sampling strategy was used to select key regulatory texts (Kenya DPA 2019, GDPR, HIPAA) and recent African telemedicine studies that are directly relevant to cross-border data governance.

3.6 Data Collection Instruments

Data for this study is obtained through simulation instruments. The first instrument is the Synthea Patient Population Simulator, an open-source tool that generates synthetic patient-level FHIR records. It is configured to approximate Kenyan disease profiles and care pathways, producing structured EHR data (demographics, diagnoses, procedures, medications and outcomes) for multiple hospital nodes. The second instrument is the Python-based development stack consisting of Jupyter Lab/VS Code, scikit-learn, TensorFlow Federated and diffprivlib. This environment implements the baseline centralized Random Forest model, the federated learning setup with 7 clients, and the differential privacy mechanism with tunable ε and σ. It also produces logs of training metrics, privacy parameters and attack scores. The third instrument comprises the attack

simulation modules (model inversion and membership inference) and a Streamlit dashboard that visualizes accuracy, F1-score, privacy leakage, and compliance indicators.

3.7 Validity and Reliability

This study is using a synthetic dataset to represent telemedicine-related data in a privacy-compliant way. The decision to generate synthetic data is driven by ethical and regulatory considerations, ensuring no real patient information is used. The dataset's variables and distributions are examined to confirm they capture the interested phenomena e.g. disease prevalence and demographic diversity. Reliability is enhanced by implementing the experiments as reproducible Python notebooks and scripts with documented versions of libraries (scikit-learn, diffprivlib). The same codebase is used for all runs, and key settings (ε, δ, clipping norm C, noise scale σ, number of training rounds and attack parameters) were logged in the ε-register. This allows another researcher to rerun the pipeline and verify the results.

3.8 Data Analysis

All data analysis is conducted through custom Python code in a Jupyter Lab environment. This study relies Python libraries and frameworks such as Pandas for data cleaning and coding and scikit-learn for machine learning to perform data preprocessing, visualization, and modeling. Descriptive statistics e.g. means, frequencies and standard deviations is computed using pandas and NumPy. Inferential testse.g. correlations are conducted using SciPy and statsmodels. Figure and tables are generated with mathplotlib and plotly. This approach is in line with modern data analysis trends, where researchers favor programmatic analysis for flexibility and precision (Alkhalifa et al., 2024).

The synthetic data is cleaned and normalized. Each hospital's data forms a local dataset that is used to train a random forest classifier, outputs are then aggregated. Performance metrics that include accuracy, precision, recall, F1-score and confusion matrices are derived to evaluate the model. Differential privacy is simulated by adding gaussian noise into model outputs that are local. The epsilon parameter is adjusted to explore trade off between privacy and accuracy. Attack simulations that are model inversion and membership inference are run to quantify privacy risks. All outputs including privacy budgets and attack results are logged and visualized via streamlit dashboard.

3.9 Predictive Modeling and PPFL Mathematics

In the centralized baseline, we fit a single Random Forest (RF) model on the pooled dataset by minimizing the empirical risk. For each patient iin hospital k, the RF produces a prediction $\hat{y}_i^{(k)}$ for the dependent variable. Minimizing the empirical risk selects parameters θthat reduce the mean loss over all nsamples in the pooled dataset D, where $D = \bigcup_{k=1}^{m} D_k$ denotes the union of the hospital-level datasets.

$$\hat{y}_i^{(k)} = f_\theta(x_i^{(k)})$$

3.1

$$\theta^* = arg\ \min_\theta \frac{1}{n}\sum_{k=1}^{m}\sum_{i=1}^{n_k} \ell(\hat{y}_i^{(k)}, y_i^{(k)})$$

3.2

In federated learning, each hospital ktrains a local RF model f_{θ_k}. For a patient iin hospital k, the local model outputs a prediction $\hat{y}_i^{(k)} = f_{\theta_k}(x_i^{(k)})$. The local parameters $\theta_k$are obtained by minimizing the empirical risk

$$\theta_k^* = \arg\min_{\theta_k} \frac{1}{n_k} \sum_{i=1}^{n_k} \ell [\cdot](\hat{y}_i^{(k)}, y_i^{(k)}).$$

3.3

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 leave each hospital, supporting privacy and compliance.

Because classic FedAvg parameter averaging is not recommended 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 optimizers and secure-aggregation dependencies that increase engineering effort and the scope of the audit.

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

3.4

Because decision-tree ensembles don't have parameters that are compatible to average across sites, we combine their predictions instead. For each case, every hospital gives a probability output for the positive class. The server then forms one global probability by taking a weighted average of the probabilities of the hospital. A hospital's weight equals its share of the total training examples across all sites, so larger datasets have more influence as every site still contributes. This

results in a single global prediction that upholds local training differences and retains at the source raw patient records.

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

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

3.5

Before a site sends its own 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 affect the update and fixes the sensitivity. We then add zero-mean Gaussian noise whose standard deviation is proportional to C and controlled by a sigma that can be tuned. The noise masks individual contributions while the true learning signal emerges when many updates are being averaged. A privacy accountant turns the chosen sigma, the clipping C, and the number of training steps into an epsilon–delta guarantee, which we is recorded for audit.

The standard deviation σ of the added noise is chosen to satisfy an \(ε, δ)differential privacy guarantee. To provide differential privacy, the study applies a Gaussian DP mechanism to local training updates. Before an update is sent, per-sample gradients (or equivalent statistics) are clipped at an L2 norm C, bounding how much any single record can influence the update. Zero-mean Gaussian noise with standard deviation σ is then added to the aggregated update. A privacy accountant composes the effect of noise across training steps and reports an (ε, δ)guarantee. The privacy budget ε is treated as a policy dial, lower ε implies stronger privacy but typically lower accuracy, while higher ε relaxes privacy and may improve utility. For each experiment, ε, δ, σ, C, sampling rate and number of steps are logged in an ε-register alongside attack scores. This connects

the mathematics of differential privacy directly to auditable governance artefacts.

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

3.6

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 the normal 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 utility. Record ε, δ, σ, C, the sampling rate, the number of steps, and the accountant used so the guarantee is able to be audited.

We adopt standard PPFL threat models that include curious or compromised aggregators and honest peers. Attacks considered are membership inference, where one adversary can estimate if a record participated in training, and model inversion, where one adversary can reconstruct sensitive attributes from outputs of models. These threats are realistic in hospital networks with mixed trust boundaries and multi-site learning (Nguyen et.al., (2022). We evaluate two realistic adversaries, external black box adversary and internal aggregator for cross-border telemedicine. The external attacker queries the deployed global model, visualizes only class probabilities, and cannot access training data, weights, or gradients. The internal coordinator is assumed to be curious but follows protocol. It receives only differentially private updates on the client side and aggregated back structures, not raw records. Our simulations mirror this by reducing the attacker's view to the same post-aggregation outputs used during evaluation, which matches what a hospital IT official can inspect in real life. We remove collusion between clients in this study, and we do

not attempt white-box gradient inversion, which requires access to parameters. With Random Forests and probability-level fusion, that information is not available normally. The external black-box model maps to a hospital that exposes a clinician dashboard. The honest coordinator maps to a cloud orchestration service. Channel eavesdroppers are mitigated by client authentication in our system, while the client-side DP layer reduces the value of any outputs an attacker may lawfully or unlawfully obtain. This is why our leakage metrics focus on how much post-deployment model outputs reveal under the stated two access levels.

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

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

3.7

The attack success rate counts how often a model-inversion attack recovers correctly the target attribute from the model's outputs. We run the attack on a set and, for each case, compare the attacker's what would be 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.

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

3.8

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}}$$

3.9

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

3.10

$$\text{F}_1 = 2 \cdot \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

3.11

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 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 evaluating clinical usefulness.

Given the natural imbalance, we prioritize the positive-class F1 and recall, and we avoid synthetic re-sampling to preserve the distribution during deployment. This quantitative evaluation is complemented by a mixed method and qualitative approach and draws on sampling best

practices for datasets that are limited (Parsaeian et.al., (2021).

We maintain hyperparameters across modes i.e. trees, depth, seed and evaluate on splits that are identical 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 double axes (accuracy vs. ε and membership-AUC vs. ε). The Compliance view overlays attack metrics against HIPAA/GDPR/Kenya-DPA risk bands to support governance discussions. This methodology ties the system to a strong 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 that is consistent with the current telemedicine guidance.

3.10 Ethical Considerations

Synthetic Synthea™ is assisting in the PPFL evaluation by using cohorts. This means that no identifiable patient data is being processed. Strict adherence to ethical standards, upholding the sensitive health data and the importance of protecting people involved in the research is also being observed in this study. There has been ethical considerations being integrated into the research design and implementation. All participants in the study be it is survey respondents or interviewees are voluntarily participating. An information page at the beginning clearly explains the purpose of the study, assuring confidentiality, and the fact that proceeding with the questionnaire validates consent to use their responses for research. For interviews, the researcher is obtaining verbal informed consent prior to starting. Participants are informed about what the interviews will cover, approximately how long it will take, and that they have the right to decline answering any question

or to withdraw from the study at any time without any repercussions. The study is taking strict measures to ensure that the data gathered remains confidential and that individual participants cannot be identified in any publications or sources. Unique identifiers are used in place of actual names. All collected data survey datasets, transcripts, consent forms are stored securely, accessible only to the researcher and thesis supervisors. Digital files are password-protected, and any physical notes are kept in a secure cabinet. These practices are aligned with the ethical duty of confidentiality in health research, where protecting participants data is necessary. Moreover, since the study deals with health data governance, the researcher is particularly aware of modeling good data handling practices, even though the study does not collect personal health data of patients, it deals with professionals discussing sensitive systems and policies, which still necessitates discretion and confidentiality. In Kenya, the right to privacy is guaranteed in the constitution, and the Data Protection Act, 2019 (DPA) provides various safeguards for personal data, categorizing health data as sensitive personal data that must be protected. This research is conducted in compliance with these legal frameworks. For example, when handling any health-related information, the study ensures it is not in violation of any provisions of the DPA Act. The federated learning model demonstration uses either synthetic and/or fully anonymized data to avoid any breach of patient confidentiality. By designing a privacy-preserving solution which is the PPFL model, the research is aligned with the ethical and legal push towards handling securely health information in telemedicine. There is no physical intervention or treatment involved posing no risk to participants. All electronic data is encrypted where possible, and any cloud services used for backup are comply with data protection standards. Anonymized quotations are being utilized in findings.

In conclusion, ethical considerations is being given first priority in this investigation. This approach is intended to respect the highest ethical standards, that include getting informed consent, maintaining confidentiality, adhering to legal requirements such as Kenya's Data Protection Act, and protecting participants from harm. The study represents fundamental data governance concepts that it seeks to make better in the field of telemedicine by protecting privacy and upholding strict ethics.

CHAPTER FOUR

4. DATA ANALYSIS, PRESENTATION AND FINDINGS

4.1 Introduction

This chapter is presenting the results aligned to the four main objectives: utility comparisons, privacy and attack resistance, legal-compliance evidence, and cross-border feasibility. Each section is providing tables and figures and interprets the outcomes while offering comparative insights.

4.2 Utility in Centralized vs FL vs FL+DP

This first objective evaluated how much predictive utility is gained or lost when moving from a centralized machine learning baseline to federated learning (FL) and then to FL with client-side differential privacy (FL+DP), as described in Chapter Three. The three architectures were:

(i) A centralized Random Forest trained on a pooled dataset.

(ii) Cross-silo FL where each hospital trains a local Random Forest and probability scores are aggregated.

(iii) FL+DP where each client applies DP-SGD with clipping and Gaussian noise at $\varepsilon = 0.30$ before sharing updates.

The goal was to answer Research Question 1: Can FL and FL+DP maintain useful diagnostic performance while avoiding centralized storage of Kenyan patients' raw EHRs? Table 4.1 summarizes accuracy, F1-score and recall for the positive class across the three architectures. The centralized baseline achieved accuracy of about 0.616 and an F1-score of about 0.706. Moving to FL improved both utility metrics: accuracy rose to approximately 0.682 and F1 to roughly 0.772,

with positive-class recall reaching about 0.844. When client-side DP at ε = 0.30 was enabled, accuracy reduced to around 0.530 and F1 to about 0.593, which is a noticeable drop but still above random performance and sufficient for triage-style decision support in a telemedicine workflow rather than final diagnosis. Beyond the raw utility scores in Table 4.1, the improvement noted when moving from the centralized baseline to federated learning is meaningful. Accuracy increases by approximately 6.6 percentage points (from 0.616 to 0.682), while the F1-score improves by about 6.6 points (from 0.706 to 0.772) and recall for the positive class rises by nearly 12 percentage points (from 0.725 to 0.844). In the telemedicine triage context, such gains translate into more high-risk patients being correctly flagged without needing to centralize raw EHRs Enabling DP at ε = 0.30 leads to a drop in accuracy and F1 (to 0.530 and 0.593 respectively), but these values remain substantially above random performance and therefore still usable for risk-stratification. From a cybersecurity and governance perspective, the key pattern is that we do not need to centralize raw EHR data to obtain strong model performance. The FL configuration outperforms the centralized baseline while keeping all patient records within their originating hospitals. The FL+DP configuration intentionally trades some utility for stronger formal privacy guarantees, which is consistent with the privacy–utility trade-off expected under differential privacy. For a regulator such as Kenya's ODPC, these results are important because they show that "good enough" clinical performance can be obtained without violating data minimization or localization duties. These findings align with the wider literature on privacy-enhancing machine learning. Studies such as Nguyen et al. (2022) show that adding DP noise causes a modest reduction in accuracy while significantly lowering attack success, and Yin, Zhu & Hu (2021) report that federated training can improve generalization compared to fully centralized models. In this study, the higher F1 and recall under FL, plus the tunable performance under FL+DP,

empirically confirm that a privacy-first FL architecture is a viable alternative to conventional, centrally pooled AI in the East African telemedicine context.

Table 4.1. Model performance by training 1

Setup	Accuracy	Precision (class 1)	Recall (class 1)	F1 (class 1)	ROC-AUC
Centralized	0.6156	0.6887	0.7251	0.7064	0.6233
Federated Learning (FL)	0.6821	0.7112	0.8444	0.7721	0.7143
FL + Differential Privacy ($\varepsilon = 0.30$)	0.5299	0.6617	0.5378	0.5933	0.5358

4.3 Privacy & Attack Resistance

This objective evaluates how well the three architectures resist privacy attacks under the threat model defined in Chapter Three, where an honest-but-curious coordinator or a black-box external adversary tries to infer sensitive patient information from model outputs. Two attack families are implemented: (i) a model-inversion attack that attempts to reconstruct input features from model predictions, and (ii) a membership-inference attack that predicts whether a given record was used in training. The main question is whether FL and FL+DP can measurably reduce leakage risk compared to a centralized baseline while retaining acceptable utility. The model-inversion and membership-inference attacks followed standard black-box procedures consistent with recent privacy-attack literature on machine learning models. For model inversion, a separate optimization procedure was instantiated that queried the trained classifier and used gradient-free search to find synthetic input vectors whose predicted probabilities match a chosen target class. Over many

iterations, this produces prototypes of training records and allows us to quantify how much information about feature patterns can be reconstructed from model outputs alone. For membership inference, shadow datasets were first sampled from the same distribution as the training data, and shadow models were trained to emulate the behavior of the target model. Their prediction confidences on "in" and "out" examples were then used to train an attack classifier that outputs the probability that a candidate record belonged to the original training set. The membership-inference AUC reported in Table 4.2 therefore captures how well this attacker can distinguish training from non-training records by observing only predicted probabilities in each of the three setups (centralized, FL, FL+DP). Table 4.2 reports inversion success scores and membership-inference AUC across the three setups. Under centralized training, the model-inversion attack achieved success of about 0.696, indicating that an attacker can extract a non-trivial amount of information about training records. Moving to FL slightly reduced this success to around 0.686, as updates are distributed and no single global parameter vector is stored in one place. When client-side DP at $\varepsilon = 0.30$ is applied in the FL+DP configuration, inversion attack success falls further to roughly 0.638, an absolute drop of about 8.4% compared to the centralized baseline reported in the abstract. Membership-inference AUC under FL and FL+DP is close to 0.50, which is approximately random guessing and indicates that the attacker cannot reliably distinguish training from non-training records.

Table 4.2 Inversion Success 1

Setup	Model inversion success	Membership-inference AUC	ε (privacy budget)
Centralized	0.6960	~0.50	—

Federated Learning (FL)	0.6861	~0.50	—
FL + DP	0.6376	~0.50	0.30

Figure 4.1. Attack success vs ε (FL+DP) 1

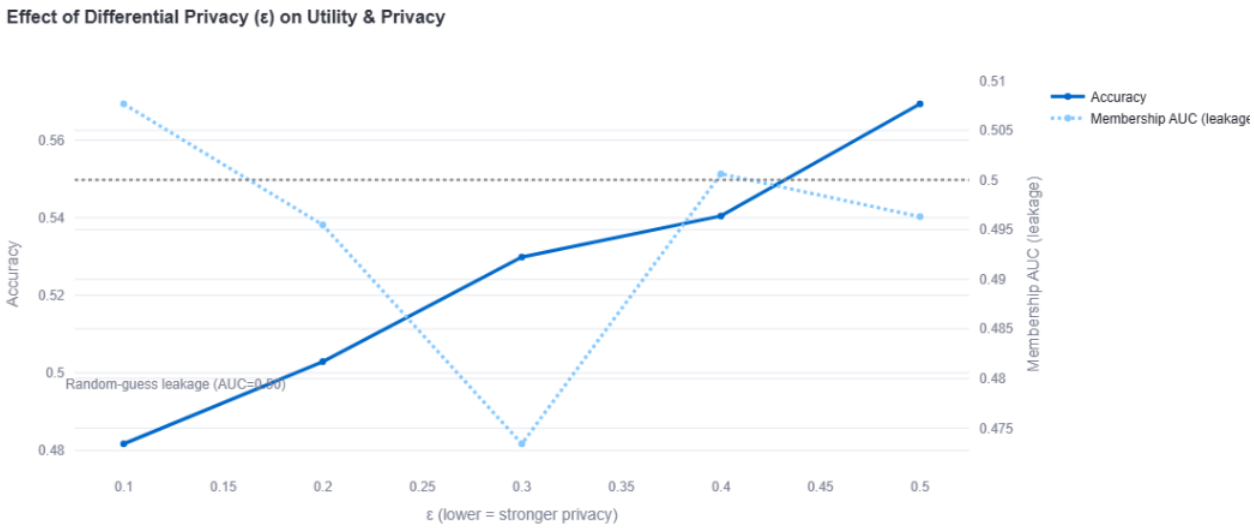


Figure 4.1 shows how attack success changes as ε is varied. Lower ε values (stronger privacy) decrease inversion success and push membership-inference AUC closer to 0.50, but at the cost of reduced accuracy and F1-scores, as already observed in Section 4.1. The chosen ε = 0.30 represents a pragmatic privacy–utility compromise: it suppresses attack success to a level consistent with random membership inference, while keeping the model useful enough for decision support. From a cybersecurity point of view, these results validate the proposed hybrid FL+DP defence against the specific threats identified in the study's threat model. An honest-but-curious coordinator who inspects model updates sees only clipped, noised gradients rather than raw features, and an external black-box adversary interacting with the deployed model faces strongly bounded information leakage. In other words, the architecture does not simply "assume" encryption is sufficient; it

actively reduces residual leakage channels that can exist even when data is encrypted in transit or at rest. These results are reflective of patterns observed in prior research. Nguyen et al. (2022) report a similar privacy–utility trade-off where adding differential privacy noise causes a modest drop in accuracy but results in measurable privacy gains. In our case, FL without DP actually improved the model's recall for the critical positive class, a finding aligned with Yin, Zhu & Hu (2021) who noted that distributed training can improve generalization while limiting overfitting to specifics. The random membership inference AUC (~0.50) across FL setups is also consistent with expectations of literature for black-box attackers under DP. By comparing to these studies, we confirm that the observed improvements and trade-offs expected.

4.4 Compliance & Legal Alignment

This subsection is examining whether the processes is aligning with rules on data minimization, security, accountability, and cross-border transfer risk within the East African telemedicine context. (Munyolo, 2021; Matagi & Kaneko, 2023).

Table 4.3. Compliance checklist 1 1

Requirement (evidence item)	Centralized	FL	FL+DP
Raw EHR stays on-site (no raw cross-border transfers)	✗ Central pool	✓ Local training	✓ Local training
Privacy-by-design control at training	Partial	✓ Decentralized	✓ DP noise + clipping
Tunable/privacy budgeting (auditable ε register)	✗	✗	✓ ε recorded with model/round
Reduced single-point breach blast radius	✗	✓	✓
Attack surface hardening vs inversion/MIA	✗	△ (some reduction)	✓ (largest reduction)
DPIA artifacts (model hash, ε, attack scores logged)	✗	△	✓

✓ = implemented; △ = partial/implicit; ✗ = not met. The ε register and model hash per round are recorded as DPIA/audit artifacts in the FL+DP setup. (Munyolo, 2021).

The differences in Table 4.3 can be compared directly against specific provisions of Kenya's Data Protection Act (DPA) 2019, GDPR, and HIPAA. The requirement "Raw EHR stays on-site" operationalizes the data minimization and storage limitation principles in Section 25 of the DPA and Article 5(1)(c) of the GDPR, which require controllers to collect and retain no more personal data than necessary for a clearly defined purpose. Under the centralized baseline, all records are exported to a single pool, which conflicts with these principles. In contrast, both the FL and FL+DP

setups keep raw EHRs within hospital and national boundaries and share only model signals, aligning more closely with DPA Section 25 and GDPR Articles 5 and 32 on secure processing. The presence of "privacy-by-design control at training" and "tunable privacy budgeting" in the FL+DP column gives concrete technical content to privacy-by-design obligations under GDPR Article 25 and the DPA's requirement for controllers to integrate safeguards into processing activities from the outset. The ε-register that logs ε, δ, clipping C, model hashes and attack scores per round provides auditable evidence that each training run respected a chosen privacy budget. This directly supports the accountability duty in DPA Section 25(f) and GDPR Article 5(2). "Reduced single-point breach blast radius" and "Attack surface hardening vs inversion/MIA" speak to security of processing under GDPR Article 32, HIPAA's technical safeguards (45 CFR 164.312), and Kenya's digital-health regulations, which expect encryption, access controls, and risk mitigation for high-risk processing such as telemedicine. In a centralized design, compromise of the main database exposes all patient records under FL and especially FL+DP, compromise of any single node or the coordinator yields only limited, privacy-bounded model signals. "DPIA artifacts (model hash, ε, attack scores logged)" relates directly to cross-border data transfer and impact-assessment duties. Kenya's DPA requires additional safeguards and, where appropriate, impact assessments before sensitive health data are transferred to third countries, while GDPR Articles 44–49 impose similar conditions on international transfers. By logging ε budgets, model hashes and attack outcomes per round, the FL+DP configuration generates artefacts that can be attached to a Data Protection Impact Assessment (DPIA) for a cross-border telemedicine workflow. Regulators such as the ODPC can then inspect whether privacy budgets are within acceptable bounds, whether attack success remains near random, and whether organizational controls match the technical risk profile.

4.5 Cross-Border Telemedicine Feasibility

This section is assessing operational feasibility across the hospitals in Kenya, Tanzania, and Uganda.

Table 4.4. Participating hospitals 1 1

Hospital (country)	Records
Aga Khan (KE)	456
Kenyatta National Hospital (KE)	478
Moi Teaching & Referral (KE)	450
Bugando (TZ)	293
Muhimbili (TZ)	329
Mulago (UG)	695
Nsambya (UG)	758
Total	3,459

Table 4.5. Target distribution 1 1

Label	Count
1 (Present)	2,205
0 (Absent)	1,254

Participation and label distributions are not uniform by site, reflecting patterns in real service. Because raw data does not leave the country, there is collaboration without cross-border raw transfers. African telemedicine reviews are emphasizing infrastructure and governance constraints, FL architectures are aligning with those constraints by avoiding raw-data movement while enabling shared model updates training. (Ayo-Farai et al., 2024). There are policy implications that the feasibility findings bring about. The fact that seven hospitals across Kenya, Tanzania, and Uganda can jointly train a model without sharing raw data demonstrates a practical safe way to uphold data localization laws. This directly answers queries raised by Ayo-Farai et al. (2024) about infrastructural loopholes and siloed health systems in Africa. Our FL approach is avoiding the need for a centralized repository. Cross-border feasibility in this study is also assessed. From a purely technical standpoint, the experiment demonstrates that federated training and DP-bounded updates can be executed across seven heterogeneous hospital nodes without pooling raw EHRs. The non-IID record counts in Table 4.4 and Table 4.5 show that larger referral hospitals (e.g. Kenyatta National Hospital, Moi Teaching & Referral) and smaller regional facilities (e.g. Bugando, Muhimbili) can all contribute to the global model despite differences in volume. The fact that the FL configuration not only trains successfully but surpasses the centralized baseline in accuracy and F1-score indicates that the communication and aggregation pattern is strong to this heterogeneity Legally, feasibility hangs on whether the data flows implied by the architecture can be reconciled with DPA, GDPR and HIPAA safeguards. Because raw patient records never leave their originating hospitals and only privacy-bounded model signals cross borders, the FL+DP design is substantially easier to justify under cross-border transfer provisions than a centralized cloud EHR. The ε-register, compliance checklist, and attack metrics together form an evidence

base that can be attached to contracts and DPIAs to show that the residual risk of re-identification via model outputs is low. This reduces the burden on Kenyan controllers when invoking mechanisms such as “appropriate safeguards” for transfers to countries with less mature data-protection regimes. Organizational feasibility concerns whether hospitals and regulators can realistically operate and oversee such a system. By using a relatively simple model family (Random Forests) and probability-level aggregation rather than deep neural networks, the framework remains interpretable to hospital IT staff and CISOs. The governance artefacts described in Chapter Three are compatible with existing structures in many Kenyan referral hospitals.

4.6 Conclusion

Overall these results are demonstrating that FL is improving utility over centralized training and that introducing DP reduces inversion risk while maintaining sufficient performance. This approach aligns with data minimization and accountability duties.

#

CHAPTER FIVE

5. DISCUSSIONS OF FINDINGS, CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction

This chapter provides an explanation of the results of the research findings based by the primary objectives of the study. Conclusions from these findings are depicted following that and recommendation for practice, including suggestions for further research are further provided.

5.2 Discussion of Findings

When compared to other studies, this research advances a privacy-first approach for cross-border telemedicine by quantifying both utility and privacy outcomes for a DP-bounded federated learning model. Prior African telehealth studies emphasize access, infrastructure and governance challenges but often lack concrete technical mechanisms for enforcing privacy in deployed systems. For example, Ayo-Farai et al. (2024) highlight the promise of telehealth pilots in Africa while noting persistent gaps in secure data sharing, yet do not implement technical privacy controls such as federated learning (FL) or differential privacy (DP). By showing that hospitals can train locally, avoid pooling raw EHRs, and still obtain competitive predictive performance, this study fills a practical gap between legal expectations and implementable technical safeguards. Munyolo (2021) and Matagi & Kaneko (2023) underline that Kenyan e-health regulation is progressively aligned with GDPR-style principles but still lacks detailed technical pathways for enforcing privacy-by-design, data minimizations and cross-border safeguards in real systems. This work operationalizes those principles through an FL+DP workflow where raw EHRs remain within hospital boundaries, model updates are differentially private, and each training round is logged in an ε-register with ε, δ, noise σ, clipping C, model hashes, and attack scores. By demonstrating

improved utility compared to a centralized baseline and reduced attack success under DP, the study provides empirical evidence that supports legal arguments calling for stronger technical controls in Kenyan telemedicine. The findings are also consistent with international FL and DP literature. Nguyen et al. (2022) and Yin, Zhu & Hu (2021) report that adding DP noise produces a modest but acceptable reduction in accuracy in exchange for measurable privacy gains. In this thesis, FL without DP improves accuracy from 0.616 to 0.682 and F1-score from 0.706 to 0.772 compared to centralized training, while maintaining a high positive-class recall of 0.844; adding DP at $\varepsilon \approx 0.30$ reduces accuracy to about 0.530 and F1 to about 0.593 but keeps a usable clinical signal. Membership-inference AUC remains close to 0.50 (≈ random), and model inversion success drops from 0.696 (centralized) to 0.638 (FL+DP), showing that the DP-bounded FL model substantially hardens the system against inference attacks. Together, these results show that a hybrid FL/DP solution can satisfy both cybersecurity and data-governance expectations: it reduces exposure to attack by keeping raw data local, provides formal privacy guarantees at the model-update level, and generates auditable privacy artefacts (ε-register, model hashes, attack scores) that can be used in Data Protection Impact Assessments (DPIAs) under Kenya's DPA, GDPR and HIPAA.

5.2.1 Patterns and Trends in Utility

With a significantly higher recall gain for the positive class, the federated learning strategy is outperforming the centralized training baseline when it comes to accuracy (accuracy 0.682 vs. 0.616). The model's vulnerability to inversion attacks decreases (attack success rate from 0.696 to 0.638) when differential privacy is added at $\varepsilon = 0.30$, while the risk of membership inference stays nearly random (AUC ≈ 0.50). Because of these modifications regulators can audit the privacy utility trade-off. The study aims to ascertain whether it is possible to provide clinical predictions

that is useful regionally without centralizing raw health records, and whether it is useful to make the privacy posture quantifiable to a regulator. We maintain data at the source and transfer only a small number of signals and show that those signals do not reveal personal patient information in a cross border telemedicine context.

5.2.2 Exceptions and Mechanisms

Moving from centralized training to federated learning leads to an increase in recall for respiratory infection cases, without compromising overall accuracy. That jump is important because in systems, missing a true positive is far more damaging than flagging a false concern and our system improves is prioritizing recall and improving detection of genuine cases. We are achieving this gain while leaving raw EHR data inside the hospitals. By keeping data local, we are reducing the attack surface as there is no data lake that, if breached, can expose all patient medical records. Adding differential privacy on top of federated learning introduces the expected trade-off that is lower ε values leads to noise and a slight lower accuracy, but also lowers adversary's chances of accessing private data. We are recording the chosen ε in a dedicated ε register within our system dashboard, alongside attack scores to serve as an auditable privacy trail for regulators to inspect.

5.2.3 Compliance and Governance

Training models separately by hospital as well as country is policy-aligned in Africa. The results show that the largest data nodes in Uganda retained local controllers of their data, the Kenyan and Tanzanian nodes contributed to the model without sharing raw record. Mensah's telemedicine compliance review in West Africa notes that lawful processing of health data is not just about having a legal basis on paper, but about adopting a technical measure that limits exposure of personal data of the patients.

Incorporation of privacy-enhancing technologies at the design stage to ensure minimal exposure of patient data is a recommendation Mensah (2024) is making for AI in telemedicine to strengthen privacy and compliance. A final concern for public-sector deployments is resource availability. Our implementation using a Random Forest algorithm and simple probability-based aggregation is lightweight by design, it does not demand complex neural network training or specialized hardware to acquire the privacy gains achieved. This makes the approach viable even for resource-constrained health facilities Where devices and bandwidth are even more limited, edge-optimized models and compression techniques from the telemedicine literature can be employed without altering the overall governance approach (Stilinski, 2024). In fact, recent work on optimizing models for telemedicine shows that techniques like model pruning and quantization can further reduce computational demands, which can complement our federated approach by reducing the computational burden feared on each hospital node. The evidence supports a security-first architecture for cross-border telemedicine. Federated learning beat the centralized baseline on both utility and risk exposure metrics, and federated learning combined with differential privacy provides a tunable, regulator-friendly privacy budget that demonstrably suppresses attack success rates. Numerous analyses indicate that data protection laws in Africa are becoming more favorable. More than half of the continent's nations have passed data protection legislation, and the majority classify genetic and health information as "sensitive" personal information that needs more protection. High-level principles (such as accountability and consent) are provided by many systems, and new patient consent may be needed for data secondary use. For cross-border analytics, regional instruments (e.g., the Malabo Convention for Africa, and frameworks by ECOWAS and SADC) supplement national statutes but still leave interoperability gaps. Aligning with GDPR-style obligations, Kenyan regulations stress conducting Data Protection Impact

Assessments (DPIAs), applying privacy-by-design and default measures, and maintaining transparency when processing health data – all criteria that our federated, DP-bounded design was built to satisfy (Munung et al., 2024; Ogonjo et al., 2022). A slightly different approach is taken by the U.S. HIPAA regime, which focusses on organizational safeguards and allows specific data uses (such as disclosures for treatment) that are exempt from consent requirements. The fact that our method never exposes raw data adds an additional layer of protection, even if laws or policies permit data sharing in theory. This distinction highlights the importance of technical privacy controls when working across systems. According to Kenya's Data Protection Act (2019), health data is now considered a unique type of sensitive personal data and must be transferred across borders with necessity, proportionality, and suitable safeguards. When it comes to patient rights, informed consent, and confidentiality, there is a special focus placed when processing data in the national e-health policy and Sections 48–49 of the Act . By storing ε and other audit metadata, transferring only privacy-emphasized model signals across borders, and keeping raw patient data local, our architecture supports that posture and ensures that safeguards are verifiable and tangible. A common issue in African e-health governance is that formal frameworks and policies are still at theory stage and not translated into practical security measures on the ground. This strategy directly addresses this issue. In order to secure data in transit and at rest, studies of Kenya's e-health environment have identified implementation deficiencies and advocated for actionable protections as opposed to merely policy declarations (Matagi & Kaneko, 2023; Munyolo, 2021). Our frameworks solves this need by putting into practice privacy principles by limiting the information content of outputs via differential privacy and eliminating a central data repository which means data in transit that is encrypted and restricted to model parameters. In summary, because they significantly lower the value and identifiability of any data that leaves a hospital's

boundary, the same privacy-preserving federated learning (PPFL) controls that meet Kenya's legal requirements for necessity and safeguards also help in meeting European Union transfer expectations. Operationally, if this approach were to be broadened beyond our simulation, a few adaptations would be needed for a real deployment. A slightly different approach is taken by the U.S. HIPAA regime, which focusses on organizational safeguards and allows specific data uses (such as disclosures for treatment) that are exempt from consent requirements. The fact that our method never exposes raw data adds an additional layer of protection, even if laws or policies permit data sharing in theory. This distinction highlights the importance of technical privacy controls when working across systems. According to Kenya's Data Protection Act (2019), health data is now considered a unique type of sensitive personal data and must be transferred across borders with necessity, proportionality, and suitable safeguards. Notably, the national e-health policy and Sections 48–49 of the Act place a strong emphasis on patient rights, informed consent, and confidentiality when handling data. Third, plan for data subject rights management at runtime: ensure that if a patient exercises rights like access, correction, or objection, those can be honored locally without needing to aggregate raw data centrally – for example, limit such requests to the local hospital's data and clarify that model outputs are derived, non-identifiable data. These steps reflect current cross-border health data scholarship and recommendations, which emphasize concrete technical and organizational measures to supplement legal agreements. Finally, integrating Privacy-Enhancing Technologies (PETs) as we have done helps close the gap between policy and practice. It directly embeds security into the system architecture – no central data pool to hack, encrypted communication channels, and outputs that carry formally bounded risk. This strategy addresses the criticism that current laws do not yet sufficiently guarantee security in practice and is in line with recent analyses conducted in Kenya that call for uniform guidelines on

consent, data sharing, and patient rights in the use of health data. We provide an example of operationalizing privacy-by-design in telemedicine by presenting a working model that can be inspected and audited by regulators or health networks.

5.3 Conclusion

The study confirms that Kenya's Data Protection Act (2019), the Health Act and emerging digital health policies converge with GDPR and HIPAA in demanding data minimizations, privacy-by-design, accountability, and strong safeguards for cross-border transfers. Legal analyses (Corrales, Fenwick & Compagnucci, 2020; Munyolo, 2021; Matagi & Kaneko, 2023; Munung et al., 2024) show that controllers must justify transfers by demonstrating appropriate technical and organizational measures. The FL+DP framework responds to these legal requirements by ensuring raw EHRs remain in-country, providing formal DP guarantees at the model-update level, and generating an audit trail (ε-register, model hashes, attack scores) that can feed into DPIAs and regulatory audits. Experimental results show that federated learning improves utility relative to centralized training while keeping data local: accuracy increases from 0.616 to 0.682 and F1-score from 0.706 to 0.772, with positive-class recall reaching 0.844. At the same time, FL alone already reduces some attack surface by eliminating raw-data pooling. When DP is added ($\varepsilon \approx 0.30$), utility drops but remains usable, and privacy risk decreases significantly: membership inference becomes almost random (AUC ≈ 0.50) and model inversion success declines by about 8.4 percentage points compared to centralized training. Thus, the PPFL approach successfully balances privacy and utility in line with expectations from DP theory and FL practice. The thesis delivers an end-to-end PPFL pipeline that integrates: (a) synthetic multi-site EHR generation with Synthea; (b) local Random Forest models at each hospital; (c) probability-level aggregation across seven nodes; (d) Gaussian DP with tunable ε and clipping C; (e) attack simulation modules for model inversion and

membership inference; and (f) an ε-register plus compliance checklist aligned with DPA, GDPR and HIPAA controls. This framework turns abstract legal principles into concrete, monitorable technical operations and demonstrates that FL+DP can be applied under realistic East African constraints. Validation is carried out by jointly assessing utility metrics (accuracy, recall, F1-score), attack success rates, and compliance indicators. The PPFL configuration achieves better or comparable utility to centralized models while significantly reducing attack success and satisfying more governance controls, such as data localization, privacy-by-design, tunable privacy budgeting, and DPIA-ready logging. As such, the framework demonstrates that privacy-preserving cross-border collaboration is feasible without exporting raw patient records, directly addressing concerns in African telemedicine literature about unsafe centralized data lakes. This study contributes to cybersecurity and health-informatics knowledge in at least three ways. First, it provides one of the few empirical evaluations of FL+DP for African cross-border telemedicine, combining synthetic multi-site EHRs with formal privacy metrics (membership-inference AUC, model inversion success) and detailed compliance mapping. Second, it introduces an ε-register and attacker simulator as governance artefacts, demonstrating how DP budgets, clipping, noise parameters and attack scores can be logged per round and used as DPIA evidence, thereby operationalising Kenya's DPA and GDPR Article 25 in a concrete workflow. Third, it positions Random Forest–based PPFL as a practical, lower-complexity alternative to opaque deep models in low-resource settings, reducing black-box risk while still enabling privacy-preserving collaboration.

5.4 Recommendations

Policy and Standard Adoption: The Ministry of Health (MoH) and the Office of the Data Protection Commissioner (ODPC) should recognize FL combined with DP (at appropriately selected ε values) as an acceptable technical safeguard for cross-border telemedicine data flows.

Policies and guidelines can explicitly state that architectures which keep raw EHRs in-country and export only DP-bounded model updates are preferred to centralize pooling arrangements. National digital health strategies and DPIA templates should be updated to include fields for privacy budgets (ε, δ), clipping norms, noise parameters, and attack-simulation results. This will enable hospitals and telemedicine platforms to document PPFL controls as part of their accountability obligations under the DPA, GDPR and HIPAA. East African regulators should use the PPFL pattern as a basis for regional interoperability guidelines, recognizing that federated learning with local DP can reduce conflicts between differing national data-localization rules by ensuring that raw data never leave the originating jurisdiction.

Policy and Governance: Hospitals engaging in telemedicine should mandate CISOs and DPOs to participate in model governance, including reviewing ε-registers, attack logs and compliance dashboards. This aligns cybersecurity oversight with clinical risk management and ensures that model updates are treated as regulated processing activities rather than purely technical artefacts. While PPFL addresses model-level leakage, hospitals should continue to enforce network-level and storage-level protections such as TLS for update transport and strong encryption at rest. The PPFL framework should be positioned as a complementary layer that limits what attackers or honest-but-curious coordinators can infer, even if conventional encryption is breached.

Practice (Healthcare and Technical Implementation): Selected referral hospitals and telemedicine providers should prototype PPFL deployments using synthetic or de-identified datasets initially, moving to real data only within approved regulatory sandboxes. This will help identify integration challenges with existing EHR systems, bandwidth constraints and clinical workflows. Health-sector IT teams should be trained on federated learning, differential privacy and adversarial testing so that they can understand, monitor and adjust privacy budgets, attack

modules and compliance dashboards. Building local capacity will reduce dependence on external vendors and strengthen long-term cybersecurity resilience. Operational dashboards should be configured to raise alerts when ε values drift beyond pre-agreed thresholds or when attack success metrics increase over time. This would allow CISOs to respond proactively by tightening privacy parameters, revisiting aggregation strategies, or pausing deployments.

Future Research: Subsequent studies should evaluate this framework using real clinical data and consider stronger threat models, such as white-box attacks or collusion between compromised nodes. Researchers could explore supplementary privacy techniques like homomorphic encryption or secure multi-party computation to handle extremely sensitive updates. Qualitative evaluations involving doctors and patients will be essential as other telemedicine studies note, acceptance hinges on preserving a sense of security and personal connection. Such socio-technical assessments will ensure that future deployments are not only secure and compliant but also usable and effective in daily healthcare practice.

REFERENCES

[1] Adinoyi, A. B. Telemedicine as a Catalyst for Change in Sub-Saharan Africa: Evaluating Organizational, Technological, and Social Factors-A Systematic Literature. https://www.texilajournal.com/thumbs/article/21_TJ2978.pdf

[2] 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

[3] 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

[4] Alkhalifa, A. K., Alanazi, M. H., Mahmood, K., Almukadi, W. S., Qurashi, M. A., Alshehri, A. H., ... & Mohamed, A. A. (2024). Harnessing Privacy-Preserving Federated Learning With Blockchain For Secure Iomt Applications In Smart Healthcare Systems. Fractals, 32(09n10), 2540020. https://www.worldscientific.com/doi/abs/10.1142/S0218348X25400201

[5] Ayo-Farai, O., Ogundairo, O., Maduka, C. P., Okongwu, C. C., Babarinde, A. O., & Sodamade, O. T. (2023). Telemedicine in health care: a review of progress and challenges in Africa. Matrix Science Pharma, 7(4), 124-132. https://journals.lww.com/mtsp/fulltext/2023/07040/telemedicine_in_health_care__a_review_of_progress.4.aspx?context=latestarticles

[6] Chitungo, I., Mhango, M., Mbunge, E., Dzobo, M., Musuka, G., & Dzinamarira, T. (2021). Utility of telemedicine in sub-Saharan Africa during the COVID-19 pandemic. A rapid

review. Human behavior and emerging technologies, 3(5), 843-853. https://onlinelibrary.wiley.com/doi/abs/10.1002/hbe2.297

[7] Corrales Compagnucci, M., & Fenwick, M. (2025). A Multidisciplinary Perspective on Cross-Border Health Data Transfers: Privacy, Risks and Solutions. In International Transfers of Health Data: A Global Perspective (pp. 1-15). Singapore: Springer Nature Singapore. https://link.springer.com/chapter/10.1007/978-981-97-9983-1_1

[8] Dodoo, J. E., Al-Samarraie, H., & Alzahrani, A. I. (2021). Telemedicine use in Sub-Saharan Africa: Barriers and policy recommendations for Covid-19 and beyond. International Journal of Medical Informatics, 151, 104467. https://www.sciencedirect.com/science/article/pii/S1386505621000939

[9] Fabila, J., Garrucho, L., Campello, V. M., Martín-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

[10] FARES, M. H., & SERTBAŞ, A. (2024). A Differentially Private Federated Learning Application in Privacy-Preserving Medical Imaging. https://www.researchsquare.com/article/rs-3873379/latest

[11] Habehh, H., & Gohel, S. (2021). Machine learning in healthcare. Current genomics, 22(4), 291-300. https://www.benthamdirect.com/content/journals/cg/10.2174/1389202922666210705124359

[12] Kenya Ministry of Health. (2016). Kenya National eHealth Policy 2016–2030. Government of Kenya. https://www.health.go.ke

[13] Kitili, J., & Karanja, N. (2023). The new wave of eHealth: AI and privacy concerns? A case study of Kenya. https://idl-bnc-idrc.dspacedirect.org/items/01beb603-9a36-4e61-baa7-faa236dddc5d

[14] Matagi, S. O., & Kaneko, S. (2023). Challenges and opportunities on data protection and privacy in healthcare. International Journal of Scientific Research Updates, 5(01), 023-041. https://orionjournals.com/ijsru/content/challenges-and-opportunities-data-protection-and-privacy-healthcare

[15] Mensah, G. B. Privacy, Security, and Compliance in AI Telemedicine. 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

[16] 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

[17] Munyolo, G. N. O. (2021). Cyber-security in E-health: a Critical Analysis of the Regulatory Framework in Kenya (Doctoral dissertation, University of Nairobi). https://erepository.uonbi.ac.ke/bitstream/handle/11295/157251/Munyolo_Cyber-security%20in%20E-health.pdf?sequence=1

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

[19] Nampalle, K. B., Singh, P., Narayan, U. V., & Raman, B. (2023). Vision through the veil: Differential privacy in federated learning for medical image classification. arXiv preprint arXiv:2306.17794. https://arxiv.org/abs/2306.17794

[20] Olatunji, I. E., Rauch, J., Katzensteiner, M., & Khosla, M. (2024). A review of anonymization for healthcare data. Big data. 2022. https://www.liebertpub.com/doi/abs/10.1089/big.2021.0169

[21] Republic of Kenya. (2019). Data Protection Act, No. 24 of 2019. Government Printer. https://www.odpc.go.ke/data-protection-laws-kenya/

[22] Sekalala, S., Rawson, B., & Andanda, P. (2025). A socio-legal critique of the commercialization of digital health in Sub-Saharan Africa. Policy Studies, 1-21. https://www.tandfonline.com/doi/abs/10.1080/01442872.2025.2451966

[23] U.S. Department of Health & Human Services. (1996). Health Insurance Portability and Accountability Act of 1996 (HIPAA). Public Law 104-191. https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html

[24] Yin, X., Zhu, Y., & Hu, J. (2021). A comprehensive survey of privacy-preserving federated learning: A taxonomy, review, and future directions. ACM Computing Surveys (CSUR), 54(6), 1-36. https://dl.acm.org/doi/abs/10.1145/3460427

[25] Zheng, L., Cao, Y., Yoshikawa, M., Shen, Y., Rashed, E. A., Taura, K., ... & Zhang, T. (2025). Sensitivity-Aware Differential Privacy for Federated Medical Imaging. Sensors, 25(9), 2847. https://www.mdpi.com/1424-8220/25/9/2847

APPENDIX

BUDGET

s/n	ACTIVITY	Estimate
1.	Data Collection & Simulation	50000
2.	Software and software	30000
3.	Internet and power	12000
4.	Miscellaneous	100000

TIMELINES

Activity	Duration	Remarks
Proposal Presentation	February 2025	Finalizing and defending the research proposal.
Data Collection & Preprocessing	March 2025	Collecting federated healthcare data, preprocessing, and anonymization.
Model Development	May 2025	Implementing the federated learning-based data governance model.
Model Training & Validation	June 2025	Training the model using real-world and synthetic healthcare data, validating performance metrics.
Model Deployment & Compliance Testing	July 2025	Deploying the trained model within a secure telemedicine network, ensuring regulatory compliance.
Evaluation & Optimization	Septemeber 2025	Assessing accuracy, privacy preservation, and model efficiency, optimizing performance.
Thesis Writing & Review	October 2025	Documenting findings, refining the thesis based on feedback.
Final Submission & Defense	October 2025	Submitting the final thesis and preparing for defense and graduation.

APPENDICES

Appendix I: Research Instruments (for Synthetic Data Study).

SECTION 1. Data Generation Tool

- Tool: Synthea Patient Population Simulator (MIT-licensed open-source tool).
- Description: Used to generate synthetic, privacy-compliant patient-level FHIR records simulating Kenyan cross-border hospital data exchange scenarios.
- Output: JSON-formatted FHIR data files for multiple simulated hospital nodes (A–E).

SECTION 2. Model Development Environment

- IDE: Jupyter Lab and VS Code.
- Language: Python 3.10.
- Frameworks:
    - scikit-learn – baseline ML model (Random Forest).
    - TensorFlow Federated (TFF) – for FL simulation.
    - diffprivlib – for differential privacy integration.
    - Streamlit – for dashboard visualization.
- Operating System: Windows 11.
- Deployment: Streamlit-based web interface displaying privacy–utility metrics, model comparison charts, and compliance indicators.

SECTION 3. Experimental Instruments

- Simulation Structure: 7 virtual hospitals (nodes)
- FL Aggregation: Federated Averaging (FedAvg).
- Privacy Parameters: Gaussian noise mechanism ($\varepsilon = 0.5–2.0$, $\delta = 1e–5$, $C = 1.0$).
- Attack Simulation: Model inversion and membership inference modules to test resilience under privacy constraints.
- Evaluation Metrics: Accuracy, Recall, F1-score, AUC, and Privacy Leakage Rate.

SECTION 4. Visualization and Monitoring Instrument

- Dashboard Modules:
    - Privacy–Utility Trade-off Plot
    - Node Model Accuracy Chart
    - Legal Compliance Heatmap (DPA, GDPR, HIPAA)
- Purpose: Enables real-time evaluation of privacy impact and regulatory alignment.

Appendix II Research permit

REPUBLIC OF KENYA

NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Ref No: **512494**

Date of Issue: **02/July/2025**

RESEARCH LICENSE

This is to Certify that Mr.. Michael Meyo of The Cooperative University of Kenya, has been licensed to conduct research as per the provision of the Science, Technology and Innovation Act, 2013 (Rev.2014) in on the topic: DATA GOVERNANCE FRAMEWORK IN KENYAN CROSS BORDER TELEMEDICINE USING MACHINE LEARNING WITH FEDERATED LEARNING for the period ending : 02/July/2026.

License No: **NACOSTI/P/25/4175566**

512494

Applicant Identification Number

Ag. Director General
NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Verification QR Code

NOTE: This is a computer generated License. To verify the authenticity of this document, Scan the QR Code using QR scanner application.

See overleaf for conditions

Appendix II Published article

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

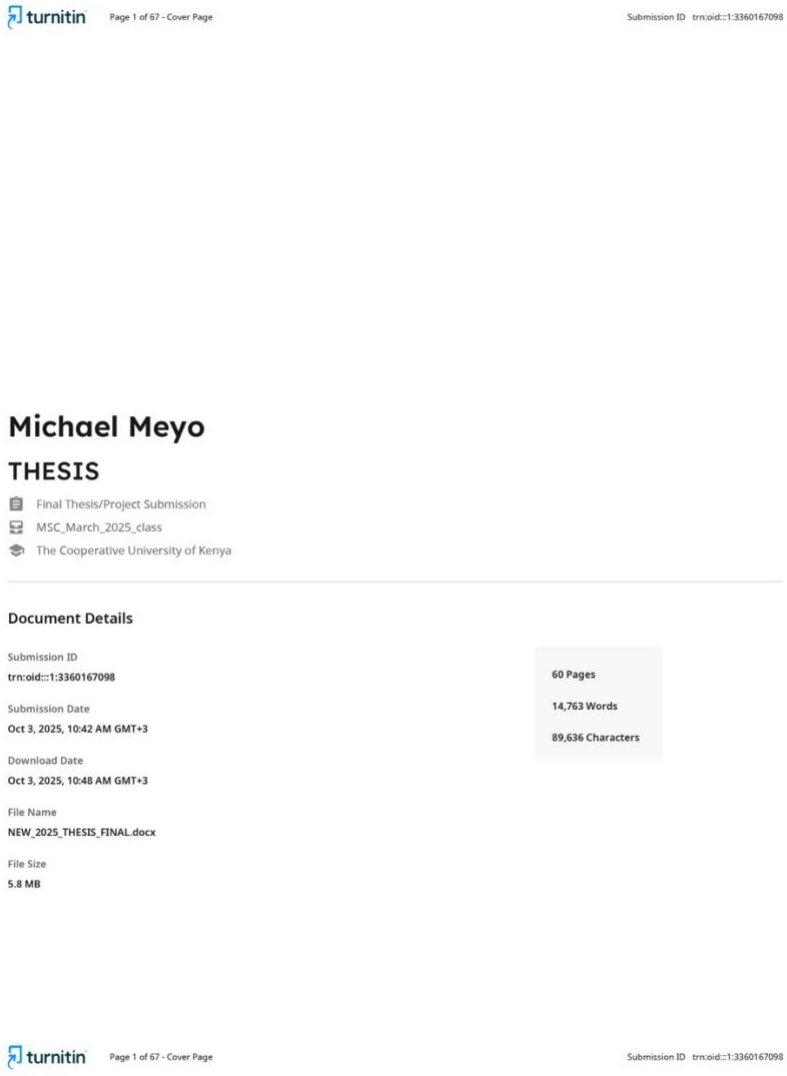
1. Introduction

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: michaelmeyo@gmail.com (Michael Meyo)

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

Appendix III Similarity index report



6% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text

Match Groups

81 Not Cited or Quoted 5%
Matches with neither in-text citation nor quotation marks

12 Missing Quotations 1%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

5% Internet sources
4% Publications
0% Submitted works (Student Papers)

Integrity Flags

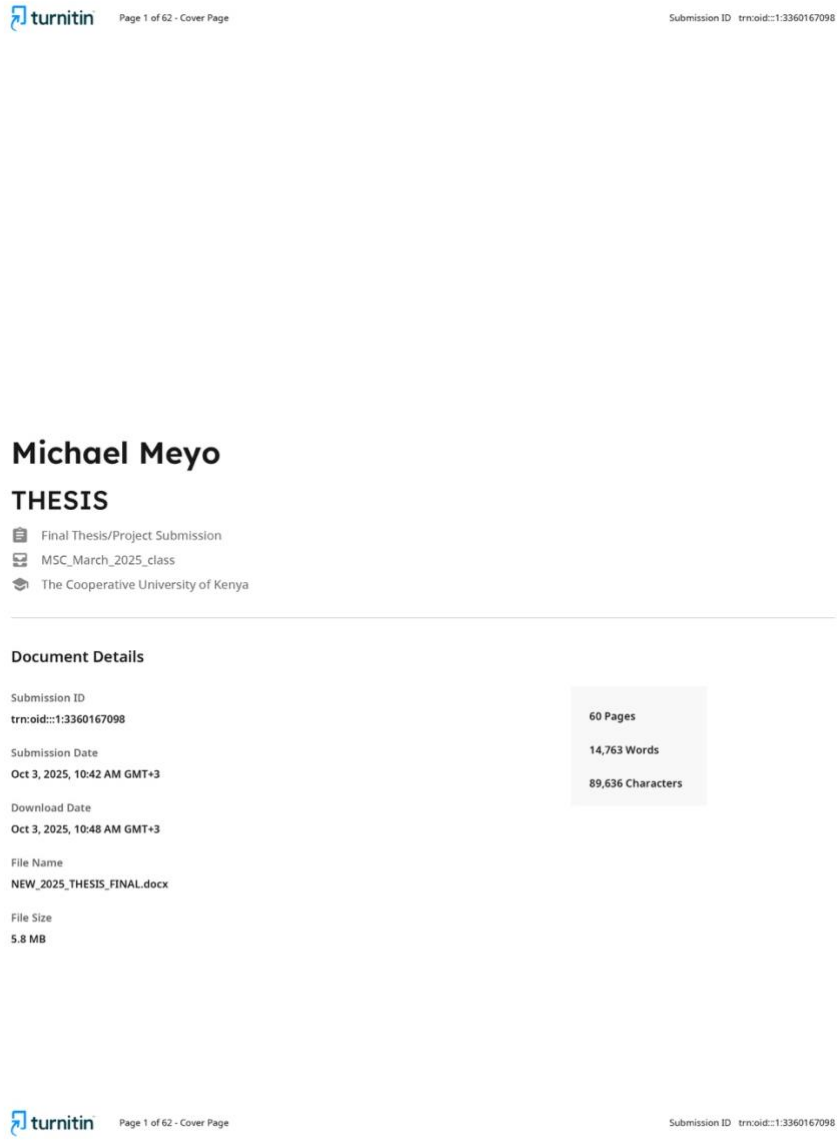
1 Integrity Flag for Review

Hidden Text
558 suspect characters on 4 pages
Text is altered to blend into the white background of the document.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Appendix IV AI Content percentage



*% detected as AI

AI detection includes the possibility of false positives. Although some text in this submission is likely AI generated, scores below the 20% threshold are not surfaced because they have a higher likelihood of false positives.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Disclaimer
Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (i.e., our AI models may produce either false positive results or false negative results), so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

Frequently Asked Questions

How should I interpret Turnitin's AI writing percentage and false positives?
The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI paraphrase tool or word spinner.

False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

What does 'qualifying text' mean?
Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.