JIRBDAI

Journal of Institutional Research, Big Data Analytics and Innovation

One-Way Hashing and Federated Learning for Privacy-Preserving Healthcare Analytics in Resource-Constrained Settings

F. S. Damang, G. I. O. Aimufua, S. E. Chaku and M. O. Adenomon
Center for Cyberspace Studies, Nasarawa State University, Keffi, Nigeria
Corresponding Author: damangfs@gmail.com

Abstract

Machine learning is increasingly applied to Nigerian patient records, yet facility data security is weak, and no lightweight privacy-preserving framework aligned with the Nigeria Data Protection Act (NDPA) 2023 and its General Application and Implementation Directive (GAID) 2025 has been evaluated locally. This study developed and evaluated a framework that combines per-facility salted SHA-256 hashing of patient identifiers with federated learning. Following a design science approach, the study trained logistic regression, random forest, and multilayer perceptron models on 10,000 synthetic records distributed across twelve simulated facilities under centralised, federated, federated-with-hashing, and federated-with-differential-privacy ($\varepsilon = 1.0$) configurations, using repeated stratified five-fold cross-validation. Utility, susceptibility to membership inference and linkage attacks, and computational cost were assessed, and the analysis was repeated on 10,000 de-identified antiretroviral therapy records from 54 facilities in the Federal Capital Territory, Nasarawa State and Katsina State. Hashing left every prediction unchanged in every fold on both datasets. While the salt remained secret, it reduced linkage attack success from 100% to 2.6% on the synthetic data and to 5.0% on the real data, produced no collisions among 110,000 identifiers, and cost about 0.001 ms per record; once the salt was leaked, the protection was lost. Hashing did not affect membership inference, which was driven by model memorisation (centralised Random Forest attack balanced accuracy 0.930 on synthetic and 0.700 on real data). On the real data, where models predicted an administrative 28-day treatment-continuity status with AUC-ROC up to 0.84, partly by exploiting documentation patterns, federation cost 0.015–0.025 in AUC-ROC and differential privacy 0.087–0.195 in accuracy. Salted hashing is an effective, low-cost safeguard against identifier linkage, but it is not a comprehensive privacy guarantee for federated analytics: in deployment it should be implemented as keyed hashing (HMAC-SHA-256) under managed keys and combined with secure aggregation and defences against membership inference and model-update leakage.

KEYWORDS: Privacy-Preserving Machine Learning; Federated Learning; One-Way Hashing; SHA-256; HMAC; Membership Inference; Record Linkage; Nigeria Data Protection Act; HIV

I. INTRODUCTION

Healthcare is the costliest sector for data breaches worldwide; the average cost of a single healthcare breach reached USD 10.93 million in 2023, the thirteenth consecutive year in which the sector headed that ranking (IBM Security, 2023). Nigeria has responded with substantive legislation. The Nigeria Data Protection Regulation (NDPR) 2019, issued by the National Information Technology Development Agency (NITDA, 2019), has been superseded by the Nigeria Data Protection Act (NDPA) 2023 (Federal Republic of Nigeria, 2023), which established the Nigeria Data Protection Commission (NDPC) as the national regulator. The NDPC's General Application and Implementation Directive (GAID) 2025, which took effect

Volume 2, Issue 4, 2026
https://doi.org/10.5281/zenodo.24161095

on 19 September 2025, provides that the NDPR no longer applies as a regulatory instrument (NDPC, 2025). Under this regime, controllers of sensitive personal data, including health data, must observe data minimisation and purpose limitation and apply appropriate technical safeguards, among which pseudonymisation is recognised. Implementation, however, lags legislative intent. A census of health personnel in six hospitals in Bayelsa State recorded low data security behaviour (Biobelemoye et al., 2026), and 57.6% of nurses surveyed at a federal teaching hospital in Nnewi identified data security concerns as a barrier to the use of electronic medical records (Jacob et al., 2026). At the same time, machine learning on Nigerian clinical records has moved beyond feasibility studies. Gradient-boosted models trained on the records of 136,747 people living with HIV have predicted interruption of antiretroviral therapy with 81% sensitivity and 88% specificity and have been piloted in ten facilities (Ogbechie et al., 2023), and an ensemble model trained on 41,394 clients across four states reached an AUC-ROC of 0.843 (Ijaiya et al., 2025). Such models depend on some of the most sensitive categories of personal information, including HIV status, viral load and treatment history, and removing names does not render records anonymous: an estimated 87% of the population of the United States is unique on ZIP code, gender and date of birth (Sweeney, 2000), and 99.98% of Americans could be re-identified in any dataset using 15 demographic attributes (Rocher et al., 2019).

Privacy-preserving machine learning (PPML) offers techniques that allow models to be trained without exposing the underlying records. The strongest of them, such as homomorphic encryption, secure multiparty computation and differential privacy, demand computational resources and specialist expertise that most Nigerian facilities lack (Kairouz et al., 2021). Federated learning, which keeps records within each institution and shares only model parameters, suits distributed health systems, but it does not by itself protect patient identifiers and remains exposed to inference and linkage attacks (Jeong & Chung, 2024). Recent African studies have evaluated federated learning on structured health data (Kovor & Osei, 2026; Meyo et al., 2025; Mirugwe & Nyirenda, 2025), and one Nigerian study has applied it to chest radiographs (Oboti et al., 2026); none, however, has protected patient identifiers through hashing, tested linkage attacks or mapped its mechanisms to Nigerian data protection law. This study addresses that gap. Its aim was to develop and evaluate a PPML framework that integrates one-way hashing of patient identifiers within a federated learning architecture for Nigerian healthcare institutions. The specific objectives were to:

(a). Design a framework that integrates salted SHA-256 hashing within a federated learning pipeline, aligned with NDPA and GAID requirements for pseudonymisation, data minimisation and purpose limitation;
(b). Implement the framework across simulated hospital nodes representing Nigeria's six geopolitical zones;
(c). Evaluate the framework on model utility, privacy resilience and operational efficiency;
(d). Compare the framework with differential privacy and with federated learning alone; and
(e). Derive a compliance mapping that the NDPC, working with NITDA on technical standards, can use in sector-specific guidance.

There is a practical need for such a framework. Hospitals require a tested basis for participating safely in collaborative artificial intelligence initiatives, regulators require a technical reference for translating legislation into auditable standards, and patients, particularly those living with stigmatised conditions such as HIV, require assurance that their records cannot be traced back to them. The study makes four contributions: a reproducible framework in which per-facility salted hashing precedes federated training; empirical evidence, on synthetic data and on real Nigerian HIV treatment records, that salted hashing removes identifier linkage risk at no cost in utility but has no effect on membership inference; a four-configuration experimental design

Volume 2, Issue 4, 2026
https://doi.org/10.5281/zenodo.24161095

that attributes each privacy and utility trade-off to a specific mechanism; and an eight-item mapping of NDPA and GAID obligations to auditable technical mechanisms. Unlike existing privacy-preserving federated learning approaches, which protect what leaves a facility during training (model updates) but leave unprotected the identifiers that link records to patients, the framework adds a separate keyed identifier layer that operates before learning, requires no trusted third party or pseudonym lookup table, and keeps the same patient unlinkable across facilities unless such linkage is expressly authorized (Section 2.5).

II. RELATED WORKS

The literature relevant to this study falls into four core subjects: re-identification and pseudonymisation, federated learning in healthcare, privacy and security threats to federated learning and their defences, and the use of encoded identifiers. Each is reviewed below for what it establishes and where it falls short of the Nigerian setting.

A. Re-identification and Pseudonymisation

It is well established that removing direct identifiers does not, on its own, prevent re-identification. Seminal studies by Sweeney (2000) and Rocher et al. (2019) showed that demographic attributes are often enough to single out an individual. Recent work confirms that the problem persists in contemporary clinical data. Behre et al. (2026) organised the factors that drive the re-identification of medical datasets into data, attacker-knowledge, attacker and technical-organisational perspectives and proposed a procedure for grading them, showing that a quantitative assessment of re-identification risk is feasible for a given dataset. Using approximately 3.95 million hospitalisation records from the Greater Paris University Hospitals, Cohen et al. (2026) found that re-identification success ranged from a median of 0.9% to 99% depending on the attack scenario, with data minimisation accounting for most of the reduction in risk and pseudonymisation contributing a smaller additional reduction. Regulatory guidance has reached the same conclusion: the European Data Protection Board stresses that the protection afforded by pseudonymisation depends on keeping the associated pseudonymisation secrets, such as cryptographic keys, separate from the pseudonymised data and under strict control (EDPB, 2025). These studies justify combining identifier pseudonymisation with minimisation and the generalisation of rare categories, but they were conducted on Western data held in well-resourced environments, and none tested pseudonymisation within a machine learning pipeline.

B. Federated Learning in Healthcare and in Africa

Federated averaging (McMahan et al., 2017) remains the reference aggregation algorithm in recent healthcare benchmarks (Kovor & Osei, 2026; Li et al., 2024) and has been demonstrated at production scale (Bonawitz et al., 2019), and Kairouz et al. (2021) identified diverse infrastructure, communication cost and privacy leakage as its principal open problems. In healthcare, Kaissis et al. (2021) established end-to-end privacy-preserving learning on multi-institutional imaging, and Li et al. (2024) showed that federated algorithms can match centralised prediction on structured emergency department data. Deployment nonetheless remains rare. Of the 612 studies reviewed by Teo et al. (2024), only 5.2% reported a real-life application, imaging accounted for 41.7% of data types, and only 27.7% added any privacy mechanism beyond federation itself.

Evidence from Africa is recent and limited. Fabila et al. (2025) found federated training feasible for tuberculosis chest radiography across eight African countries but constrained by unreliable infrastructure and reluctance to share model updates. Kovor and Osei (2026) showed that federated models retained 97.8% to 98.5% of centralised AUC-PR for malaria prediction

Volume 2, Issue 4, 2026
https://doi.org/10.5281/zenodo.24161095

in Ghana. Mirugwe and Nyirenda (2025) combined federated learning with differential privacy and secure aggregation on routine HIV data from 30 Ugandan facilities, and Meyo et al. (2025) evaluated a comparable design on synthetic records from seven East African hospitals. In Nigeria, the only federated study identified is that of Oboti et al. (2026), who trained a differentially private classifier on 2,600 chest radiographs. These studies demonstrate that federated learning can retain predictive utility on African data, but none protects patient identifiers or evaluates linkage risk.

C. Inference, Reconstruction and Poisoning Threats and their Defences

Shokri et al. (2017) showed that membership inference attacks can determine whether a record was used to train a model, and Yeom et al. (2018) linked such leakage directly to overfitting. These attacks have since become considerably stronger: the robust membership inference attack of Zarifzadeh et al. (2024) attains high true-positive rates at very low false-positive rates with as few as one reference model, and the review by Bai et al. (2025) shows that, in federated settings, an adversary can also exploit the updates exchanged across training rounds. Recent work confirms the threat in clinical settings; Nemecek et al. (2025) detected measurable membership leakage from a language model fine-tuned on electronic health records. Differential privacy offers a formal defence (Abadi et al., 2016; Dwork & Roth, 2014), but it reduces accuracy and requires expertise in calibrating the privacy budget. Ponomareva et al. (2023) further caution that a differential privacy guarantee cannot be interpreted unless the unit of privacy (for example, a record or a participating institution) and the definition of neighbouring datasets are reported alongside ε and δ. An expert consensus has recommended that privacy evaluations measure membership and attribute disclosure directly rather than rely on record-similarity metrics (Pilgram et al., 2025), a recommendation that this study follows.

Membership is not the only information that federated updates disclose. Gradient inversion attacks reconstruct training records from the updates themselves: Hatamizadeh et al. (2023) showed, in a realistic healthcare federation, that the risk of reconstruction rises when a client holds few records or trains with small batches, and Vero et al. (2023) showed that tabular records, the data type of most clinical prediction models, can be reconstructed with high accuracy from both FedSGD and FedAvg updates. Federated learning is also exposed to malicious or compromised participants, which can poison their local data or manipulate their updates to degrade or backdoor the global model (Xia et al., 2023). The principal defences are complementary rather than interchangeable: secure aggregation ensures that the server learns only the sum of the participants' updates (Mansouri et al., 2023), robust aggregation rules limit the influence of anomalous updates (Xia et al., 2023), and differential privacy bounds what a released model or update reveals about any individual contributor. None of these mechanisms protects patient identifiers, which is the gap addressed by the present framework.

D. Encoded Identifiers

Evidence on encoded identifiers comes mainly from record linkage. Lee et al. (2018) proposed federated patient hashing for similarity search across institutions, Ferrante et al. (2024) reported accurate Bloom-filter linkage of Australian health and justice data without releasing identifiers, and Wang et al. (2024) showed that models trained on irreversibly encoded clinical time-series data remained effective while leaking less about their training data. These studies show that encoded identifiers retain their linkage utility, but they treat encoding as a linkage step rather than as a privacy layer within a federated pipeline, and none was conducted in a low-resource setting.

Volume 2, Issue 4, 2026
https://doi.org/10.5281/zenodo.24161095

E. Research Gap

The literature review identifies three gaps. First, studies in Nigeria and across Africa document weak data security in detail (Biobelemoye et al., 2026; Gebeyew et al., 2025; Jacob et al., 2026) but rarely implement and test a technical safeguard. Second, African federated learning studies address utility and, in some cases, membership inference, but not identifier protection or linkage. Third, PPML evaluations typically compare a single private configuration with a non-private baseline, so the contribution of each mechanism to each trade-off cannot be isolated. No study identified in this review has combined one-way identifier hashing with federated learning on structured clinical data in an African setting, quantified the resulting utility, privacy and efficiency trade-offs, and mapped the mechanisms to a national data protection law. The present study addresses each of these shortcomings.

The proposed framework differs from existing privacy-preserving federated learning approaches in what it protects and how. Secure aggregation, differential privacy and homomorphic encryption protect the training signal, that is, the updates and models that leave a facility. They leave untouched the patient identifiers that every facility continues to hold in its analytical extracts, logs and validation sets, and that link a record to a known person. The framework treats identifier protection as a distinct layer that precedes learning, and it offers three practical advantages over established pseudonymisation techniques. First, compared with tokenisation through a trusted third party or a pseudonym lookup table, keyed hashing requires neither a mapping table, which would itself be a high-value target, nor an external service, which suits facilities with limited connectivity. Second, compared with reversible encryption, there is no decryption capability to govern or misuse in routine operation, although, as Section 4.1.2 shows, a leaked key still permits enumeration of a small identifier space. Third, compared with the Bloom-filter encodings used for privacy-preserving record linkage (Ferrante et al., 2024), which are designed to make records linkable across institutions, the per-facility key deliberately keeps the same patient unlinkable across facilities unless such linkage is expressly authorised, which is consistent with purpose limitation. The contribution is therefore not a new cryptographic primitive but an evaluated architecture that assigns each privacy threat to a specific mechanism and measures the residual risk associated with each, on Nigerian data and with reference to Nigerian law.

III. METHODOLOGY

A. Research Design

The study adopted Design Science Research (Hevner et al., 2004; Peffers et al., 2007) because its aim was to build and evaluate an artefact, a PPML framework, in response to a defined problem rather than to describe a phenomenon. The artefact was evaluated computationally in a controlled experiment, first on synthetic records and then on a de-identified extract of real HIV treatment records.

B. Framework Architecture

The framework comprises five layers (Figure 1). Raw records remain in the data layer at each facility. In the privacy layer, every patient identifier is replaced, before any further processing, with a salted SHA-256 hash:

$$h_F(\mathrm{ID}) = \mathrm{SHA\text{-}256}(\mathrm{ID} \parallel \mathrm{salt}_F) \quad (1)$$

where salt_F is a 256-bit value generated for facility F by a cryptographically secure pseudorandom number generator, held in an encrypted local key store and never transmitted. Because salt_F is kept secret rather than stored alongside the hashes, it functions as a secret key (sometimes termed a pepper) rather than as a conventional public salt, and it is managed as a key under the procedures in Section 3.2.1. Hashing was preferred to encryption because it is

Volume 2, Issue 4, 2026
https://doi.org/10.5281/zenodo.24161095

irreversible while the salt remains secret, deterministic within a facility (so that a patient's repeat visits remain linkable locally) and linear in cost. In the learning layer, each facility trains a local model on its hashed records, with normalisation statistics computed locally. In the aggregation layer, a server combines the local parameters by federated averaging (McMahan et al., 2017):

$$\theta_G^{(r+1)} = \Sigma_{F=1}^{K} (|D_F| / |D|)\, \theta_F^{(r)} \quad (2)$$

where θ_F is the parameter vector of facility F after one local epoch in round r, K is the number of facilities and $|D_F| / |D|$ is the facility's share of the training records. Only model parameters cross the institutional boundary. This limits, but does not eliminate, what leaves a facility, because model parameters can themselves carry information about the records that produced them (Section 3.2.2). The evaluation layer monitors utility, privacy and efficiency.

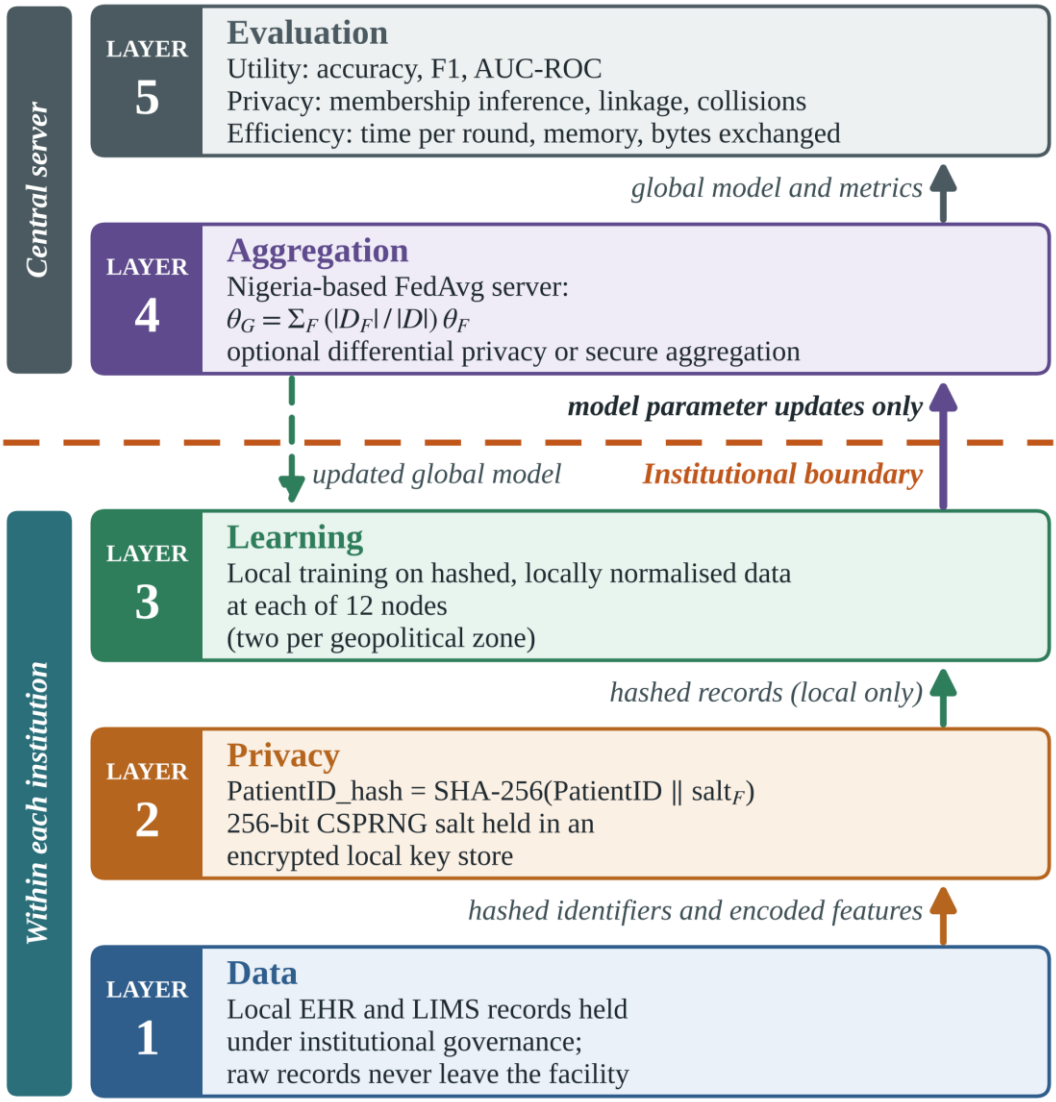


Figure 1. Five-layer Architecture of the Proposed Framework

Raw records and salts remain within each institution; only model parameter updates cross the institutional boundary. The figure depicts the evaluated configuration (12 synthetic nodes; Equation 1); the real-data replication used 54 nodes, and HMAC-SHA-256 (Equation 3) is specified for deployment. Source: Authors' design.

Volume 2, Issue 4, 2026
https://doi.org/10.5281/zenodo.24161095

Keyed Hashing and Key Management

The construction in Equation (1), a hash of the identifier concatenated with a secret value, is the one shown in the privacy layer of Figure 1 and evaluated in this study. For deployment, the framework specifies the standardised keyed-hash construction HMAC-SHA-256 (Krawczyk et al., 1997; National Institute of Standards and Technology [NIST], 2008):

$$h_F(\mathrm{ID}) = \mathrm{HMAC\text{-}SHA\text{-}256}(k_F, \mathrm{ID}) \quad (3)$$

where k_F is the facility key (the salt_F of Equation 1). HMAC computes SHA-256(($k_F \oplus$ opad) ‖ SHA-256(($k_F \oplus$ ipad) ‖ ID)), where ipad and opad are the fixed inner and outer padding constants defined in the standard. HMAC is preferred because its security as a keyed pseudorandom function rests on a formal analysis rather than on ad hoc concatenation, because it is the construction that standard cryptographic libraries and key-management systems expect, and because it removes any ambiguity about how key and message are combined. For the properties evaluated here, the two constructions are equivalent: both are deterministic within a facility, one-way while the key remains secret and computationally negligible, since HMAC adds only a second hash invocation to a per-record cost of about 0.001 ms. The linkage, collision and utility results reported in Section 4 therefore apply equally to HMAC-SHA-256, as does the dependence of protection on the secrecy of the key.

Because protection rests entirely on key secrecy, the framework specifies the following key-management procedures, aligned with NIST guidance on key management (Barker, 2020). The experiments generated keys with Python's secrets module as described above; the access-control, rotation and recovery procedures form part of the deployment specification and were not themselves evaluated.

(a). Generation: Each facility generates a 256-bit key locally with a cryptographically secure generator, such as an operating-system random source or a hardware security module. Keys are never generated centrally, reused across facilities or derived from facility names or codes.

(b). Storage: The key is stored encrypted at rest, preferably in a hardware security module, trusted platform module or operating-system key store, and is kept physically and logically separate from the pseudonymised data. It is never written to analysis extracts, logs, model artefacts or configuration files, and it is never transmitted to the aggregation server.

(c). Access control: Only the hashing service invokes the key; analysts and model developers have no direct access to it. Administrative access requires role-based authorisation under dual control, for example by the data protection officer and the head of health records, and every use is recorded in a tamper-evident audit log that is reviewed during the facility's NDPA compliance audit.

(d). Rotation: Each pseudonym is stored with the version identifier of the key that produced it. Keys are rotated on a fixed schedule (for example, annually), whenever compromise is suspected and whenever staff with administrative access leave. Because rotation changes every pseudonym, the facility re-pseudonymises its retained records from the source identifiers under the new key, retires the old key once re-pseudonymisation is complete and records the transition in the audit log. Models need not be retrained, because the identifier is not a model feature.

(e). Recovery: Loss of the key would sever the facility's ability to link new records to existing pseudonyms. The key is therefore backed up offline in encrypted form and divided among custodians by threshold secret sharing (Shamir, 1979), so that, for example, any two of three designated officers can reconstruct it, but no single custodian can.

(f). Compromise response: A suspected disclosure of the key is handled as a personal data breach under the NDPA: the key is revoked, the records are re-pseudonymised under a new

Volume 2, Issue 4, 2026
https://doi.org/10.5281/zenodo.24161095

key, previously released pseudonymised extracts are treated as identifiable, and the breach is notified to the NDPC as the Act requires.

Threat Model and Scope of the Privacy Claims

The privacy evaluation assumes an honest-but-curious aggregation server, which follows the protocol but may inspect whatever it receives; honest participating facilities; and an external adversary who holds identifiers and quasi-identifiers for some patients and who may obtain a pseudonymised extract or query a trained model. Within this model, the study evaluates two threats: identifier linkage, against which hashing is the defence, and membership inference, against which model choice, federation and differential privacy are compared. Gradient inversion, reconstruction from shared model structures, poisoning by malicious or compromised facilities and a malicious server lie outside the experimental scope. These threats are nonetheless material. With one local epoch of logistic regression, a facility's update is a deterministic function of its records, and under tree pooling each facility shares complete trees whose split thresholds and leaf class frequencies describe its records directly. The privacy claims of this study are therefore limited to identifier protection and to the measured membership leakage, and Section 4.2.6 sets out the additional mechanisms that a deployment requires.

C. Data Source

The following data sources were used in this research:

(a). Synthetic dataset. The primary evaluation used 10,000 synthetic patient records generated with a fixed random seed to reflect the organisation of Nigerian hospital records across twelve facilities, two in each of the six geopolitical zones, spanning tertiary, secondary and primary care (410 to 1,355 records per facility). Each record held a patient identifier, age, gender, diagnosis, treatment type, a laboratory value, facility, zone, tier and a three-class outcome (Improved 42.8%, Stable 38.8%, Deteriorated 18.4%). After one-hot encoding, the models received 28 features. Synthetic data were used for the primary evaluation because assembling 10,000 real records across all tiers and zones would have required multiple data-sharing agreements, and because the cryptographic and aggregation mechanisms under test operate identically on synthetic and real records.

(b). Real replication dataset: The full evaluation was then repeated on an extract of 10,000 routine antiretroviral therapy (ART) records from the Nigeria National Data Repository (Federal Ministry of Health, National AIDS and STIs Control Programme [NASCP], 2026), covering 54 facilities in the Federal Capital Territory (33), Nasarawa State (20) and Katsina State (1). This distribution is an important limitation. Fifty-three of the 54 facilities lie in the North-Central geopolitical zone, a single facility represents the North-West, and the North-East, South-East, South-South and South-West zones are not represented. The real-data experiments are therefore a replication of the framework's mechanisms on real records rather than a nationally representative evaluation, and their estimates should not be generalised to the Nigerian health system. Before analysis, hospital numbers, dates of birth, facility names and codes, local government areas and all dates were removed, and facilities were recoded without retaining a mapping. The facility-assigned patient identifier was retained only so that it could be hashed and attacked; it was never a model feature. The outcome was the dataset's 28-day treatment status (Active 31.0%, Inactive 69.0%), and the ten features were age, age at ART initiation, years on ART, weight, state, sex, WHO clinical stage, TB status, regimen line and transfer-in status. The 28-day status is an administrative measure of treatment continuity rather than a clinical endpoint. It follows the convention used in PEPFAR-supported HIV programme monitoring, under which a client ceases to be

Volume 2, Issue 4, 2026
https://doi.org/10.5281/zenodo.24161095

counted as active once more than 28 days have passed since the last expected clinical contact or antiretroviral drug pick-up without a further contact (UNAIDS, n.d.). The inactive category therefore combines interruption in treatment with documented death, transfer out and stopped treatment, and it records the state of each client's file at the extraction date rather than a prospectively observed event. It is used in this study to provide a non-trivial real-data signal against which the utility cost of hashing, federation and differential privacy can be measured, not to develop a clinical risk-prediction tool. Fields that directly encode treatment status, such as pickup dates and discontinuation indicators, were excluded. The 54 facilities served as the federated nodes.

D. Models and Experimental Configurations

Three models were evaluated: Logistic Regression (LR) as a fast and interpretable baseline, Random Forest (RF) as a non-linear ensemble, and a Multilayer Perceptron (MLP) with two hidden layers of 128 and 64 units. Hyperparameters were fixed before the experiments rather than tuned, so that no difference between configurations could arise from configuration-specific tuning. Because a Random Forest has no parameters that can be averaged, it was federated by tree pooling: each facility trained a forest sized in proportion to its share of the data, and the server averaged the predicted class probabilities of all trees. Under this scheme each facility shares complete trees rather than averaged parameters, which has privacy consequences discussed in Section 4.2.6. Each model was evaluated under the configurations in Table 1.

Table 1. Experimental Configurations

Configuration	Training	Identifier handling	Purpose
Config 1: Centralised	Pooled data on one server	Raw identifiers	Current-practice baseline
Config 2: FL	Federated averaging, up to 50 rounds with early stopping	Raw identifiers	Isolates the effect of federation
Config 3: FL + SHA-256	As Config 2	Per-facility salted SHA-256	The proposed framework; isolates the effect of hashing
FL + DP	As Config 2, with update clipping and Gaussian noise ($\varepsilon = 1.0$, $\delta = 10^{-4}$)	Raw identifiers	Formal defence against membership inference (LR and MLP only)

Comparing Config 1 with Config 2 attributes any change to federation and comparing Config 2 with Config 3 attributes any change to hashing. For differential privacy, each update was clipped to an L2 norm of 1.0, and Gaussian noise was added to the aggregate, with the noise multiplier ($\sigma = 31.15$) chosen by a Rényi differential privacy accountant (Mironov, 2017) composed over all 50 budgeted rounds. Because a privacy budget can be interpreted only in relation to its unit of privacy (Ponomareva et al., 2023), the accounting assumptions are stated explicitly. The unit of privacy was the facility. Clipping bounded each facility's contribution to the record-weighted aggregate, so that the sensitivity of the aggregate to a change in any one facility's update was the clipping norm multiplied by the largest facility weight, and neighbouring datasets were defined as differing in the entire data of one facility. This is client-level differential privacy (McMahan et al., 2018). Because a single record can affect only its own facility's update, the same (ε, δ) bound also holds, conservatively, at record level. All facilities took part in every round, so no amplification by subsampling was claimed, and the budget was composed over all 50 budgeted rounds even when early stopping ended training sooner. Noise was added by the server to the aggregate (central differential privacy), so the

Volume 2, Issue 4, 2026
https://doi.org/10.5281/zenodo.24161095

guarantee protects facilities and patients against any party that observes the released global models, but not against a server that observes the individual clipped updates before noise is added; closing that gap requires secure aggregation combined with distributed noise (Section 4.2.6).

Differential privacy was applied to LR and the MLP but not to the Random Forest, for two reasons. First, the clip-and-noise mechanism requires a fixed-length, real-valued parameter vector whose sensitivity can be bounded, whereas a forest federated by tree pooling exchanges discrete structures, namely split variables, thresholds and leaf class frequencies, that cannot be clipped and perturbed in that way. Second, differentially private forests rely on different mechanisms, such as randomised split selection and noisy leaf counts, whose budgets are spent per tree and composed across trees (Fletcher & Islam, 2019). A guarantee obtained in that way would not be comparable with the $\varepsilon = 1.0$ obtained for LR and the MLP under the same accountant, and including it would confound the mechanism-by-mechanism comparison that the design is intended to isolate. The consequence is that the formal-defence comparison applies to LR and the MLP only. For the Random Forest, privacy evidence rests on the empirical membership inference attacks alone; because the federated forest still leaked measurably on the synthetic data (balanced accuracy 0.588), it should not be deployed within the framework without either a differentially private forest mechanism or constraints on tree depth and minimum leaf size.

E. Evaluation

Utility was measured by stratified five-fold cross-validation repeated twice, giving ten paired test folds of 2,000 records; 15% of each training partition was held out for early stopping, and the same partitions were used for every configuration. Accuracy, macro-averaged precision, recall and F1-score, and AUC-ROC were reported. Configurations were compared by paired t-tests, with Wilcoxon signed-rank tests as a non-parametric check and a Bonferroni correction across 33 comparisons ($\alpha = 0.0015$), and equivalence was assessed against a margin of 1% of Config 2 accuracy. Privacy was assessed with four tests, each repeated with five random seeds. Membership inference used the shadow-model attack of Shokri et al. (2017), with shadow models trained under the same configuration as the target, and the loss-threshold attack of Yeom et al. (2018); leakage was summarised by the attack's balanced accuracy, for which 0.5 corresponds to chance. The linkage attack gave an adversary the identifiers and quasi-identifiers of 10% or 30% of patients and measured the proportion re-identified against a store keyed by raw identifiers, by hashes with a secret salt, and by hashes with a leaked salt. A collision test hashed 110,000 identifiers, and an enumeration test measured the time needed to hash the full identifier space once a salt is known.

Efficiency was measured as hashing time, time per federated round, rounds to convergence, bytes exchanged per round and peak memory. All experiments ran in Python 3.13 (NumPy, pandas, scikit-learn and SciPy; hashlib and secrets for hashing) on a cloud virtual machine with two virtual CPUs, 7 GiB of RAM and no GPU, a configuration more modest than the servers of most Nigerian tertiary hospitals. The facilities were simulated sequentially within one process, so communication cost was measured in bytes rather than transfer time.

Volume 2, Issue 4, 2026
https://doi.org/10.5281/zenodo.24161095

F. Ethical Considerations

The synthetic and real-data experiments were conducted without exposing personal data. Ethical approval was not required for the secondary analysis of the de-identified antiretroviral therapy (ART) extract, as access was authorised by the repository custodian, the Federal Ministry of Health's National AIDS and STIs Control Programme (NASCP), through the National Data Repository (NDR) Analytic Workspace (NAWS). Direct identifiers were removed before analysis, all simulated attacks were conducted within the authorised analytical environment, and no individual-level records were disclosed outside that environment.

IV. RESULTS AND DISCUSSION

The results obtained from this study is presented as follows:

A. Model Utility

On the synthetic data, no model learned a useful mapping from features to outcome. The highest mean accuracy, 0.427 for federated LR, equalled the majority-class rate of 0.428, and macro-averaged AUC-ROC lay between 0.495 and 0.533. A diagnostic analysis showed that the outcome was almost independent of the predictors (largest Cramér's V = 0.043). The real ART data, by contrast, carried a strong signal, and the models predicted 28-day treatment status (an administrative treatment-continuity measure; Section 3.3) with AUC-ROC of up to 0.841 (Table 2).

Table 2: Utility on the Real ART Extract (mean of ten folds; majority-class Accuracy 0.690)

Model	Configuration	Accuracy (mean ± SD)	Macro-F1	AUC-ROC
LR	Config 1: Centralised	0.748 ± 0.008	0.678	0.833
LR	Config 2: FL	0.736 ± 0.007	0.664	0.809
LR	Config 3: FL + SHA-256	0.736 ± 0.007	0.664	0.809
LR	FL + DP (ε = 1.0)	0.541 ± 0.097	0.494	0.539
RF	Config 1: Centralised	0.746 ± 0.012	0.702	0.824
RF	Config 2: FL	0.725 ± 0.007	0.628	0.810
RF	Config 3: FL + SHA-256	0.725 ± 0.007	0.628	0.810
MLP	Config 1: Centralised	0.759 ± 0.008	0.724	0.841
MLP	Config 2: FL	0.739 ± 0.009	0.704	0.816
MLP	Config 3: FL + SHA-256	0.739 ± 0.009	0.704	0.816
MLP	FL + DP (ε = 1.0)	0.652 ± 0.050	0.475	0.531

On both datasets, Config 3 reproduced Config 2 exactly for every model in all ten folds, so the 90% confidence interval of every difference was [0, 0] and equivalence within the 1% margin was established. On the real data, federation reduced AUC-ROC by 0.024 for LR, 0.015 for RF and 0.025 for the MLP, all significant at the Bonferroni threshold ($p \leq 0.0008$). Differential privacy reduced accuracy by 0.195 for LR and 0.087 for the MLP and brought AUC-ROC close to chance; on the synthetic data its accuracy cost was 0.044 to 0.064.

B. Privacy

Table 3 and Figure 2 summarise the attacks. Against a store keyed by raw identifiers, every auxiliary record was re-identified. With hashed identifiers and a secret salt, the adversary was reduced to matching on quasi-identifiers, which re-identified 2.6% of records on the synthetic data and 5.05% on the real data at 30% overlap. When the salt was leaked, linkage success returned to 100% on both datasets, and the identifier space could be enumerated quickly: 10^6 six-digit synthetic identifiers in 0.83 s, and 10^8 eight-digit real identifiers in about 102.5 s. No collisions occurred among 110,000 hashed identifiers.

Volume 2, Issue 4, 2026
https://doi.org/10.5281/zenodo.24161095

Table 3. Privacy Evaluation (means over five attack seeds)

Measure	Synthetic data	Real ART data
Linkage, raw identifiers (30% overlap)	100%	100%
Linkage, hashed with secret salt (30% overlap)	2.6% (95% CI 2.4–2.7%)	5.05% (95% CI 4.59–5.51%)
Linkage, hashed with leaked salt	100%	100%
Time to enumerate identifier space with known salt	0.83 s (10^6 identifiers)	about 102.5 s (10^8 identifiers)
MIA balanced accuracy, centralised RF	0.930	0.700
MIA balanced accuracy, federated RF (Config 2 / Config 3)	0.588 / 0.588	0.502 / 0.505
MIA balanced accuracy, LR and MLP (all configurations)	0.497–0.523	0.496–0.507
Hash collisions among 110,000 identifiers	0	0

Membership inference depended on the model rather than on hashing. LR and the MLP remained at chance in every configuration. The centralised RF memorised its training data (training accuracy 1.000 against test accuracy 0.385 on the synthetic data) and was highly vulnerable. Federation reduced this leakage to 0.588 on the synthetic data, which remained significantly above chance ($p = 4.4 \times 10^{-7}$), and to chance on the real data. Hashing made no difference to membership inference on either dataset (synthetic $p = 0.91$; real $p = 0.48$), and differential privacy kept the attack at chance for LR and the MLP.

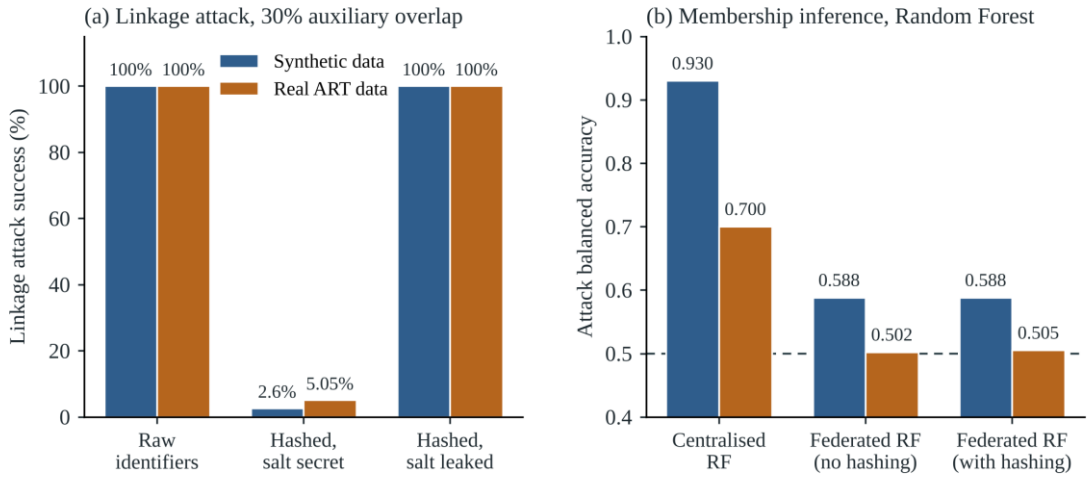


Figure 2. (a) Linkage attack success at 30% auxiliary overlap; (b) membership inference balanced accuracy for the Random Forest,

where 0.5 (dashed line) corresponds to chance. Blue: synthetic data; orange: real ART data. Source: Authors' analysis.

C. Efficiency

Hashing 10,000 identifiers took 10.9 ms on the synthetic data and 12.2 ms on the real data, about 0.001 ms per record. Time per federated round did not differ significantly between Config 2 and Config 3 for any model (for example, MLP 348.2 against 346.2 ms, $p = 0.85$), and rounds to convergence, bytes exchanged and peak memory (at most 3.2 MiB of Python allocations) were identical. Communication was the larger cost, with the MLP exchanging 2.3 MB per round across twelve facilities and 10.1 MB per round across 54.

Volume 2, Issue 4, 2026
https://doi.org/10.5281/zenodo.24161095

D. Sensitivity Analysis and Fairness

On the real data, part of the predictive signal reflected record keeping rather than clinical risk, since almost every record with a missing WHO stage, TB status, regimen line or weight was inactive. Without these fields, centralised AUC-ROC fell to 0.552–0.602, yet the privacy findings were unchanged: hashing again left every prediction unchanged and did not affect membership inference. On the synthetic data, no subgroup by age band, gender, diagnosis, tier, zone or facility differed significantly in accuracy after Bonferroni correction across 36 comparisons.

DISCUSSION

A. What Hashing Does and Does Not Protect

The lack of utility loss attributable to hashing can be explained by the system's architecture. The identifier is used to key and link records but is not a model input, so replacing it with a salted hash changes nothing that the model sees while removing the information an adversary needs to link a record to a known person. The real data make this result more informative than the synthetic data alone, because there the models had predictive value to lose and lost none. The residual linkage of 2.6% to 5.05% reflects matching on quasi-identifiers, which were more distinctive in the real records; this is consistent with the finding of Cohen et al. (2026) that pseudonymisation must be accompanied by minimization of quasi-identifiers. The protection depends entirely on the secrecy of the salt. Because facility identifiers come from small spaces, an adversary who obtains a salt can hash every possible identifier within minutes. Key management, rather than the choice of hash function, is therefore the critical design element, and the salt must be managed as a cryptographic key, preferably within the HMAC-SHA-256 construction and under the generation, storage, access-control, rotation and recovery procedures specified in Section 3.2.1.

B. Membership Inference Requires a Separate Defence

The findings challenge two widely held assumptions: that keeping data local through federated learning is sufficient to ensure privacy, and that pseudonymising identifiers adequately protects against inferences drawn from a trained model. The evidence supports neither assumption. Membership inference exploits a model's behaviour on a record's features and outcome, not its identifier, and its success tracked memorisation, as Yeom et al. (2018) predicted. Where they overlap, the findings agree with recent African studies. Mirugwe and Nyirenda (2025) found membership inference against differentially private federated models to be no better than chance, and Meyo et al. (2025) reported an attack AUC of about 0.50 for a federated Random Forest built by combining site-level probabilities. The measurable leakage of the federated forest on the synthetic data in this study (0.588) may reflect the fully grown trees used here. The present study extends this work by examining identifier protection, which those studies did not address. Because membership inference attacks continue to grow stronger (Bai et al., 2025; Zarifzadeh et al., 2024), the attacks used here should be regarded as providing a lower bound on leakage rather than a certification of safety.

C. The Cost of Federation and Differential Privacy

On the real data, federation cost 0.015 to 0.025 in AUC-ROC, a loss comparable in size to that reported by Kovor and Osei (2026), whose federated models retained 97.8% to 98.5% of centralised performance. Differential privacy was considerably more costly on the real data than on the synthetic data, because one facility held 23.8% of the training records, which raised the sensitivity of the aggregate and hence the noise required for $\varepsilon = 1.0$. The price of a formal guarantee in a small, unbalanced federation is therefore governed by the size distribution of its

Volume 2, Issue 4, 2026
https://doi.org/10.5281/zenodo.24161095

facilities. This is directly relevant to Nigeria, where tertiary hospitals hold far more records than primary health centers.

D. Regulatory Implications

The framework maps each of the eight NDPA and GAID requirements to a mechanism and a verification method: pseudonymisation, data minimisation, purpose limitation, lawful basis, data protection impact assessment, cross-border transfer restriction, data subject rights, and accountability. Pseudonymisation was verified experimentally; four requirements are addressed by design; and three depend on action by each institution at deployment. The mapping therefore makes compliance achievable and auditable, but compliance remains a property of a deployment rather than of a design. It offers the NDPC, working with NITDA on technical standards, a concrete starting point for the sector-specific code of practice for health data that the NDPA anticipates.

E. Interpreting the Real-Data Predictive Performance

The AUC-ROC of up to 0.84 obtained on the real data should be interpreted with care. The outcome is an administrative status recorded at a single extraction date, which combines interruption in treatment with death, transfer and stopped treatment, rather than a prospectively measured clinical event. Much of the signal came from documentation: almost every record with a missing WHO stage, TB status, regimen line or weight was inactive, and without these fields centralised AUC-ROC fell to 0.552–0.602 (Section 4.1.4). The models therefore learned, to a substantial degree, which client files were incompletely maintained, a pattern that may itself be a consequence of disengagement from care rather than a predictor of it. These results are not evidence of clinical risk prediction, and the models are not proposed for clinical use. Their role in this study is methodological: they supply a real-data signal with predictive value to lose, so that the utility cost of each privacy mechanism can be measured. The central utility finding, that hashing changed no prediction, holds whether the documentation fields are included. A clinically useful model would require a prospectively defined outcome, such as interruption in treatment within a defined window after a scheduled visit, longitudinal visit-level data, calibration and external validation, as in the Nigerian studies of Ogbechie et al. (2023) and Ijaiya et al. (2025).

F. Threats Beyond Identifier Linkage and Membership Inference

The results demonstrate protection against identifier linkage more convincingly than they establish comprehensive privacy for federated analytics, and the distinction matters for deployment. Keeping records local and sharing only model parameters does not by itself guarantee patient privacy. Model updates can be inverted to reconstruct training records, and the risk is greatest where updates are computed from few records, as at the smaller facilities in this federation; tabular records have been reconstructed with high accuracy from FedAvg updates (Hatamizadeh et al., 2023; Vero et al., 2023). Under tree pooling, the shared trees disclose split thresholds and leaf class frequencies, and fully grown trees with very small leaves can approach a description of individual records. A malicious or compromised facility could poison its data or updates to degrade or backdoor the global model, and a malicious server could deviate from the protocol, for example by isolating the update of a single facility (Xia et al., 2023). None of these threats was evaluated in this study, and salted hashing mitigates none of them. Table 4 summarises the threats, the extent to which each was evaluated and the mechanisms that a deployment should add.

Volume 2, Issue 4, 2026
https://doi.org/10.5281/zenodo.24161095

Table 4. Threat Coverage and Mechanisms required for Deployment

Threat	Evaluated in this study	Protection provided by the framework	Additional mechanism required for deployment
Identifier linkage	Yes (synthetic and real data)	Per-facility keyed hashing; linkage reduced to 2.6–5.05% while the key is secret	Key management (Section 3.2.1); minimisation and generalisation of quasi-identifiers
Key compromise	Yes (leaked-key scenario)	None once the key is leaked; identifier space enumerated in seconds to minutes	Hardware-backed storage, dual control, rotation and breach response (Section 3.2.1)
Membership inference	Yes (shadow-model and loss-threshold attacks)	None from hashing; leakage reduced by federation and, for LR and MLP, by differential privacy	Differential privacy; regularised models; depth and leaf-size limits for forests; pre-deployment audits with stronger attacks
Gradient inversion and reconstruction from updates	No	None	Secure aggregation; differential privacy; minimum number of records per facility and round
Leakage from shared tree structures (RF)	No	None	Depth and minimum leaf-size constraints; differentially private forests
Malicious or compromised facility (poisoning, backdoors)	No	None	Facility authentication; robust aggregation; anomaly detection on updates; audit logging
Malicious or compromised server	No	Raw records and keys never leave facilities	Secure aggregation with distributed noise; aggregation server located in Nigeria under NDPA governance; verifiable logs

These mechanisms are complementary. Secure aggregation hides individual updates from the server but not the aggregate, and it complicates robust aggregation because the server can no longer inspect the updates that it combines (Mansouri et al., 2023). Differential privacy bounds what the aggregate and the released model reveal, at a cost in accuracy that Section 4.2.3 shows to depend on the size distribution of the facilities. A proportionate design for Nigerian facilities would therefore layer keyed hashing for identifiers, secure aggregation for updates, differential privacy calibrated to the composition of the federation for released models, and authenticated participation with robust aggregation and audit logging for participants. The present study evaluates the first of these layers and the membership-inference component of the third; evaluating the remainder is a priority for future work.

LIMITATIONS

The study has seven limitations. First, the synthetic outcome was almost unpredictable, so clinical utility rests on the real data alone, whose outcome is cross-sectional, combines treatment interruption with death and transfer, and draws much of its signal from documentation patterns. Second, each patient appeared once, so the linkage of repeat visits that deterministic hashing is designed to preserve was not exercised. Third, facilities were simulated in one process, so network latency and node dropout were not modelled, and timings on shared virtual hardware are indicative. Fourth, the threat evaluation was confined to identifier linkage and membership inference: gradient inversion, reconstruction from shared trees, poisoning by

Volume 2, Issue 4, 2026
https://doi.org/10.5281/zenodo.24161095

malicious facilities, a malicious server, model inversion and adaptive attacks were not evaluated, and secure aggregation was not implemented (Section 4.2.6). Fifth, the differential privacy accountant was conservative and, for the reasons given in Section 3.4, was not applied to the Random Forest. Sixth, hyperparameters were fixed rather than tuned, and the HMAC-SHA-256 construction specified for deployment, although equivalent to the evaluated construction for the properties measured, was not separately benchmarked. Seventh, the real data came from 54 facilities in three states, 53 of them in the North-Central zone, so the real-data results cannot be generalised to all six geopolitical zones or to the Nigerian health system.

CONCLUSION

Nigerian health institutions are adopting machine learning faster than they are building the safeguards that the NDPA and the GAID require, and the privacy techniques favoured in the international literature are often too demanding for the infrastructure available. This study set out to design, implement and evaluate a lightweight framework in which patient identifiers are hashed with a per-facility salt before federated training, and to establish what such a framework does and does not protect. The evidence, drawn from 10,000 synthetic records and replicated on 10,000 real antiretroviral therapy records from 54 facilities in three Nigerian states, supports three principal conclusions. First, salted SHA-256 hashing protects identifiers against linkage at no cost in utility and at negligible computational cost, as it changed no prediction on either dataset, reduced linkage success from 100% to between 2.6% and 5.05% and required about a thousandth of a millisecond per record on modest hardware. Second, that protection rests entirely on the secrecy of the salt, since a leaked salt restored linkage to 100% and allowed the identifier space to be enumerated within minutes. Third, neither hashing nor federation prevents membership inference by models that memorise their training data; this threat requires a separate defence through model choice, regularisation or differential privacy, whose cost in accuracy rises sharply when a single facility holds a large share of the data. Taken together, the findings establish that the framework protects patient identifiers against linkage; they do not establish comprehensive privacy for federated analytics, which additionally requires the protection of model updates and of the training process itself. The study contributes a reproducible framework and evaluation pipeline, evidence from Nigerian data on the distinct roles of hashing, federation and differential privacy, a four-configuration design that attributes each trade-off to a specific mechanism, and a mapping of eight NDPA and GAID obligations to auditable technical controls. Based on these findings, three recommendations are made. Nigerian health institutions should apply per-facility keyed hashing (HMAC-SHA-256) to patient identifiers in every dataset used for model development or shared for analysis, manage the keys under the procedures specified in this study, combine hashing with secure aggregation, and test models for membership leakage before deployment. The NDPC, in collaboration with NITDA, should use the compliance mapping as a basis for a sector-specific code of practice for artificial intelligence in healthcare, including requirements on pseudonymisation, key management and the location of aggregation servers. Future research should extend the evaluation to longitudinal visit records, to clinically defined and prospectively measured outcomes such as viral suppression, to gradient inversion, poisoning, model inversion and adaptive attacks, to secure aggregation and differentially private forests, to facilities in all six geopolitical zones, and to aggregation methods designed for heterogeneous facilities, such as FedProx (Li et al., 2020), within a phased pilot on the national health information infrastructure.

Volume 2, Issue 4, 2026
https://doi.org/10.5281/zenodo.24161095

REFERENCES

Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 308-318.

Bai, L., Hu, H., Ye, Q., Li, H., Wang, L., & Xu, J. (2025). Membership inference attacks and defenses in federated learning: A survey. *ACM Computing Surveys, 57*(4), Article 89. https://doi.org/10.1145/3704633.

Barker, E. (2020). *Recommendation for key management: Part 1 – General* (NIST Special Publication 800-57 Part 1, Revision 5). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-57pt1r5.

Behre, S., Kesztyüs, D., Schnabel, S., & Kesztyüs, T. (2026). Proposal of a procedure to stratify the reidentification risk of medical data: RIMEDA. *BMC Medical Informatics and Decision Making, 26*, Article 120. https://doi.org/10.1186/s12911-026-03475-4.

Biobelemoye, J. G., Ebiteinye, D. L., & Idiapho, C. G. N. (2026). Assessment of digital health data procurement and cybersecurity consciousness among health personnel in Bayelsa State, Nigeria. *International Journal of Pharmaceutical Sciences, 4*(5), 4673-4686. https://doi.org/10.5281/zenodo.20279043.

Bonawitz, K., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., Ivanov, V., Kiddon, C., Konečný, J., Mazzocchi, S., McMahan, H. B., Van Overveldt, T., Petrou, D., Ramage, D., & Roselander, J. (2019). Towards federated learning at scale: A system design. *Proceedings of Machine Learning and Systems, 1*, 374-388.

Cohen, A., Jacob, Y., Chatellier, G., Jean, C., Playe, B., Mouchet, A., Audureau, E., Boutet, A., & Bey, R. (2026). Quantifying the effects of pseudonymisation on epidemiological research reliability: A tailored evaluation using a clinical data warehouse. *BMC Medical Informatics and Decision Making, 26*, Article 87. https://doi.org/10.1186/s12911-026-03360-0.

Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science, 9*(3-4), 211-407.

European Data Protection Board. (2025). *Guidelines 01/2025 on pseudonymisation* (Version 1.0, adopted 16 January 2025). EDPB.

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:2505.14217). arXiv. https://doi.org/10.48550/arXiv.2505.14217.

Federal Ministry of Health, National AIDS and STIs Control Programme. (2026). *PHIS3/NDR/Treatment_Dataset* [Unpublished raw data]. National Data Repository, Nigeria.

Federal Republic of Nigeria. (2023). *Nigeria Data Protection Act, 2023*. Federal Government Printer.

Ferrante, A., Brown, A., Randall, S., & Boyd, J. (2024). Linkage for risk-sensitive environments: Australian examples of privacy-preserving record linkage using Bloom filters. *International Journal of Population Data Science, 9*(5). https://doi.org/10.23889/ijpds.v9i5.2636.

Fletcher, S., & Islam, M. Z. (2019). Decision tree classification with differential privacy: A survey. *ACM Computing Surveys, 52*(4), Article 83.

Hatamizadeh, A., Yin, H., Molchanov, P., Myronenko, A., Li, W., Dogra, P., Feng, A., Flores, M. G., Kautz, J., Xu, D., & Roth, H. R. (2023). Do gradient inversion attacks make federated learning unsafe? *IEEE Transactions on Medical Imaging, 42*(7). https://doi.org/10.1109/TMI.2023.3239391.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly, 28*(1), 75-105.

IBM Security. (2023). *Cost of a data breach report 2023*. IBM Corporation.

Ijaiya, M., Troncoso, E., Mutloatse, M., Ifeanyi, D., Obasa, B., Emerenini, F., De Voux, L., Mnguni, T., Parrott, S., Okwor, E., Dare, B., Ogundare, O., Atuma, E., Strachan, M., Fayorsey, R., & Curran, K. (2025). Use of machine learning in predicting continuity of HIV treatment in selected Nigerian states. *PLOS Global Public Health, 5*(4), e0004497. https://doi.org/10.1371/journal.pgph.0004497.

Jacob, J. N., Okeke, U. C., Okwara, B. O., & Ugbala, A. (2026). Infrastructural and training barriers to electronic medical record use among nurses at Nnamdi Azikiwe University Teaching Hospital

(NAUTH), Nnewi, Nigeria. *The Nigerian Health Journal, 26*(2), 886-892. https://doi.org/10.71637/tnhj.v26i2.1404.

Jeong, H., & Chung, T.-M. (2024). *Security and privacy issues and solutions in federated learning for digital healthcare* (arXiv:2401.08458). arXiv. https://doi.org/10.48550/arXiv.2401.08458.

Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R. G. L., Eichner, H., El Rouayheb, S., Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B., Gibbons, P. B., … Zhao, S. (2021). Advances and open problems in federated learning. *Foundations and Trends in Machine Learning, 14*(1-2), 1-210.

Kaissis, G., Ziller, A., Passerat-Palmbach, J., Ryffel, T., Usynin, D., Trask, A., Lima, I., Jr., Mancuso, J., Jungmann, F., Steinborn, M.-M., Saleh, A., Makowski, M., Rueckert, D., & Braren, R. (2021). End-to-end privacy preserving deep learning on multi-institutional medical imaging. *Nature Machine Intelligence, 3*(6), 473-484.

Kovor, D. K., & Osei, E. O. (2026). Multi-scenario evaluation of federated learning for privacy-preserving malaria prediction with Ghana DHS data. *PLOS Digital Health, 5*(7), e0001581. https://doi.org/10.1371/journal.pdig.0001581.

Krawczyk, H., Bellare, M., & Canetti, R. (1997). *HMAC: Keyed-hashing for message authentication* (RFC 2104). Internet Engineering Task Force. https://doi.org/10.17487/RFC2104.

Lee, J., Sun, J., Wang, F., Wang, S., Jun, C.-H., & Jiang, X. (2018). Privacy-preserving patient similarity learning in a federated environment: Development and analysis. *JMIR Medical Informatics, 6*(2), e20. https://doi.org/10.2196/medinform.7744.

Li, S., Miao, D., Wu, Q., Hong, C., D'Agostino, D., Li, X., Ning, Y., Shang, Y., Wang, Z., Liu, M., Fu, H., Ong, M. E. H., Haddadi, H., & Liu, N. (2024). Federated learning in healthcare: A benchmark comparison of engineering and statistical approaches for structured data analysis. *Health Data Science, 4*, Article 0196. https://doi.org/10.34133/hds.0196.

Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., & Smith, V. (2020). Federated optimization in heterogeneous networks. *Proceedings of Machine Learning and Systems, 2*, 429-450.

Mansouri, M., Önen, M., Ben Jaballah, W., & Conti, M. (2023). SoK: Secure aggregation based on cryptographic schemes for federated learning. *Proceedings on Privacy Enhancing Technologies, 2023*(1), 140-157. https://doi.org/10.56553/popets-2023-0009.

McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & Agüera y Arcas, B. (2017). Communication-efficient learning of deep networks from decentralized data. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), 54*, 1273-1282.

McMahan, H. B., Ramage, D., Talwar, K., & Zhang, L. (2018). Learning differentially private recurrent language models. *International Conference on Learning Representations (ICLR 2018)*.

Meyo, M., Ikamari, C., & Mile, A. (2025). A privacy-preserving data governance in cross-border telemedicine using federated learning and differential privacy in Kenya. *Internet of Things and Cloud Computing, 13*(3), 62-76. https://doi.org/10.11648/j.iotcc.20251303.12.

Mironov, I. (2017). Rényi differential privacy. *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, 263-275. https://doi.org/10.1109/CSF.2017.11.

Mirugwe, A., & Nyirenda, J. (2025). *Secure and efficient federated learning for predictive modeling in resource-constrained healthcare systems* [Preprint]. medRxiv. https://doi.org/10.1101/2025.07.27.25332284.

National Information Technology Development Agency. (2019). *Nigeria Data Protection Regulation 2019*. NITDA.

National Institute of Standards and Technology. (2008). *The keyed-hash message authentication code (HMAC)* (FIPS PUB 198-1). https://doi.org/10.6028/NIST.FIPS.198-1.

Nemecek, A., Yun, Z., Rahmani, Z., Harel, Y., Chaudhary, V., Sharif, M., & Ayday, E. (2025). *Exploring membership inference vulnerabilities in clinical large language models* (arXiv:2510.18674). arXiv. https://doi.org/10.48550/arXiv.2510.18674.

Nigeria Data Protection Commission. (2025). *Nigeria Data Protection Act General Application and Implementation Directive (GAID) 2025*. NDPC.

Oboti, N. P., Nwakeze, O. M., Okika, S. S., Okafor, C. M., & Agubosim, C. C. (2026). Privacy-preserving federated learning with MobileViT for chest X-ray classification in Nigerian

hospitals. *Journal of the Nigerian Society of Physical Sciences, 8*(4), Article 3466. https://doi.org/10.46481/jnsps.2026.3466.

Ogbechie, M. D., Fischer Walker, C., Lee, M. T., Abba Gana, A., Oduola, A., Idemudia, A., Edor, M., Harris, E. L., Stephens, J., Gao, X., Chen, P. L., & Persaud, N. E. (2023). Predicting treatment interruption among people living with HIV in Nigeria: Machine learning approach. *JMIR AI, 2*, e44432. https://doi.org/10.2196/44432.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems, 24*(3), 45-77.

Pilgram, L., Dankar, F. K., Drechsler, J., Elliot, M., Domingo-Ferrer, J., Francis, P., Kantarcioglu, M., Kong, L., Malin, B., Muralidhar, K., Myles, P., Prasser, F., Raisaro, J. L., Yan, C., & El Emam, K. (2025). A consensus privacy metrics framework for synthetic data. *Patterns, 6*(10), Article 101320. https://doi.org/10.1016/j.patter.2025.101320.

Ponomareva, N., Hazimeh, H., Kurakin, A., Xu, Z., Denison, C., McMahan, H. B., Vassilvitskii, S., Chien, S., & Thakurta, A. G. (2023). How to DP-fy ML: A practical guide to machine learning with differential privacy. *Journal of Artificial Intelligence Research, 77*, 1113-1201. https://doi.org/10.1613/jair.1.14649.

Rocher, L., Hendrickx, J. M., & de Montjoye, Y.-A. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications, 10*, Article 3069. https://doi.org/10.1038/s41467-019-10933-3.

Shamir, A. (1979). How to share a secret. *Communications of the ACM, 22*(11), 612-613. https://doi.org/10.1145/359168.359176.

Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017). Membership inference attacks against machine learning models. *Proceedings of the 2017 IEEE Symposium on Security and Privacy*, 3-18.

Sweeney, L. (2000). *Simple demographics often identify people uniquely* (Data Privacy Working Paper 3). Carnegie Mellon University.

Teo, Z. L., Jin, L., Li, S., Miao, D., Zhang, X., Ng, W. Y., Tan, T. F., Lee, D. M., Chua, K. J., Heng, J., Liu, Y., Goh, R. S. M., & Ting, D. S. W. (2024). Federated machine learning in healthcare: A systematic review on clinical applications and technical architecture. *Cell Reports Medicine, 5*(2), Article 101419. https://doi.org/10.1016/j.xcrm.2024.101419.

UNAIDS. (n.d.). *TX_ML: Number of ART patients with no clinical contact or ARV pick-up for greater than 28 days since their last expected clinical contact or ARV pick-up* [PEPFAR MER indicator]. Indicator Registry. Retrieved November 11, 2026, from https://indicatorregistry.unaids.org/node/1346.

Vero, M., Balunović, M., Dimitrov, D. I., & Vechev, M. (2023). TabLeak: Tabular data leakage in federated learning. *Proceedings of the 40th International Conference on Machine Learning, PMLR 202*, 35051-35083.

Wang, Y., Armstrong, J., Thakur, A., Zhu, T., Abrol, V., & Clifton, D. A. (2024). Data encoding for healthcare data democratization and information leakage prevention. *Nature Communications, 15*, Article 1582. https://doi.org/10.1038/s41467-024-45777-z.

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

Yeom, S., Giacomelli, I., Fredrikson, M., & Jha, S. (2018). Privacy risk in machine learning: Analyzing the connection to overfitting. *2018 IEEE 31st Computer Security Foundations Symposium (CSF)*, 268-282. https://doi.org/10.1109/CSF.2018.00027.

Zarifzadeh, S., Liu, P., & Shokri, R. (2024). Low-cost high-power membership inference attacks. *Proceedings of the 41st International Conference on Machine Learning, PMLR 235*, 58244-58282.