Received: 9 September 2025 | Revised: 3 November 2025 | Accepted: 11 November 2025 | Published online: 1 December 2025

RESEARCH ARTICLE

Medinformatics
2025, Vol. 00(00) 1–7
DOI: 10.47852/bonviewMEDIN52027621

BON VIEW PUBLISHING

Secure and Efficient Federated Learning for Predictive Modeling in Resource-Constrained Healthcare Systems

Alex Mirugwe[1,*] **and Juwa Nyirenda**[2]

[1] *School of Public Health, Makerere University, Uganda*

[2] *Department of Statistical Science, University of Cape Town, South Africa*

Abstract: Predictive modeling in healthcare holds promise for improving clinical outcomes, but in many low-resource settings, data fragmentation, privacy concerns, and infrastructural limitations hinder centralized machine learning approaches. These barriers are especially critical in human immunodeficiency virus (HIV) care, where privacy concerns are heightened, and any use case involving patient-level data raises significant ethical, regulatory, and confidentiality concerns. We developed a privacy-preserving federated learning (FL) framework to predict HIV viral load suppression using retrospective data from 50,000 patients and over one million visits across 30 health facilities in Uganda. The framework utilizes federated averaging for distributed training, secure multiparty aggregation, and differential privacy to ensure data confidentiality. To address cross-site heterogeneity, we integrated domain-adversarial neural networks to promote domain-invariant feature learning. A multilayer perceptron model was trained collaboratively across facilities using only local data. The federated model achieved an area under the ROC curve (AUC) of 0.874, nearly matching a centralized baseline (AUC 0.881) and substantially outperforming site-specific models (average AUC 0.758). Sensitivity (89.6%) and specificity (66.8%) demonstrate strong capability in identifying both suppressed and unsuppressed cases. Domain adaptation reduced inter-facility performance variability, and differential privacy imposed minimal degradation in accuracy. Training was completed within one hour using modest hardware, which supported feasibility in low-resource settings. Our study demonstrates that FL can deliver robust, privacy-preserving predictive performance in HIV care without requiring the centralization of sensitive patient data. The proposed architecture is adaptable to other clinical prediction tasks and represents a practical pathway for scaling ethical AI across decentralized healthcare systems in low- and middle-income countries.

Keywords: federated learning, HIV treatment prediction, differential privacy, domain adaptation

1. Introduction

Over the past decade, machine learning applications in healthcare have expanded significantly, demonstrating potential across various domains, including disease diagnosis, prognosis, clinical trials, genomics sequencing, and personalized treatment [1, 2]. However, data fragmentation remains a major challenge, particularly in low-resource settings, where health information systems are often siloed and lack interoperability [3]. This fragmentation complicates the development and deployment of robust machine learning models, as institutions are often reluctant to share patient-level data due to privacy, security, and ethical concerns [4]. Consequently, the inability to access diverse and sufficiently large datasets limits model generalizability and hinders real-world implementation, ultimately preventing the realization of machine learning's full potential in healthcare [5].

To address the challenge of fragmented and inaccessible healthcare data while preserving patient privacy, federated learning (FL) has emerged as a viable solution [4, 6]. FL enables collaborative model training across multiple institutions without requiring direct data sharing, ensuring that sensitive patient information remains localized and secure [7]. By allowing machine learning models to be trained on decentralized datasets, this approach mitigates privacy risks and aligns with ethical and regulatory requirements [8]. This also improves the trustworthiness and sovereignty of health data by giving institutions control over their datasets while still contributing to the development of more robust and generalizable machine learning models [5].

Despite its potential, FL in healthcare faces several challenges that hinder its widespread adoption. One major concern is the heterogeneity of healthcare data, as different institutions use varying electronic medical record (EMR) systems, data formats, and collection standards, which can introduce biases and inconsistencies in model training [9, 10]. The computational and communication demands of FL also create barriers, especially in resource-limited settings where institutions may lack the necessary infrastructure to support decentralized training [8]. While FL enhances privacy by keeping data local, it remains vulnerable to adversarial attacks, including model inversion and poisoning, which can compromise sensitive patient information even without direct data access [11, 12]. These challenges must be addressed to ensure the robust-

*Corresponding author: Alex Mirugwe, School of Public Health, Makerere University, Uganda. Email: amirugwe@musph.ac.ug ness, security, and scalability of FL in real-world healthcare applications.

This study aims to develop a secure and efficient FL framework for predictive modeling in resource-constrained healthcare settings. The focus is on predicting the human immunodeficiency virus (HIV) viral load (VL) suppression status of clients using decentralized data from multiple health facilities with different EMR systems. To ensure privacy and security, the framework incorporates robust aggregation protocols and lightweight encryption techniques, minimizing communication overhead to account for infrastructure limitations. The approach also addresses data heterogeneity by integrating domain adaptation methods that improve model generalizability across different healthcare institutions. The effectiveness of this framework will be evaluated by comparing its performance to centralized machine learning approaches and assessing its feasibility in environments where data sharing is restricted due to privacy and ethical concerns. By tackling these challenges, this work advances privacy-preserving predictive analytics in low-resource healthcare settings. Existing studies typically address privacy, efficiency, or heterogeneity in isolation. This study combines all three: secure aggregation, differential privacy (DP), and domain adaptation, within a unified FL framework designed for low-resource health systems. Using large-scale HIV data from 30 health facilities, we show that the framework matches centralized performance while preserving data privacy and institutional autonomy. This is the first demonstration of such an integrated approach applied in a real-world, resource-constrained clinical setting.

2. Literature Review

FL has increasingly become a significant research area due to its promise of addressing critical privacy and data-sharing issues in healthcare. Originating from distributed learning paradigms, FL allows multiple institutions to collaboratively train machine learning models without centralizing patient data, therefore mitigating privacy risks [7]. Several studies have demonstrated successful applications of FL in diverse healthcare domains, including medical imaging, disease diagnosis, and clinical prediction tasks [4, 13].

Kaissis et al. [5] proposed privacy-preserving FL architectures for medical imaging, highlighting their ability to train robust and generalizable models without compromising patient confidentiality. Sheller et al. [9] similarly applied FL for multi-institutional collaboration in brain tumor segmentation, illustrating its feasibility to achieve comparable accuracy to centrally trained models. Despite these promising demonstrations, many FL frameworks remain resource-intensive, making them challenging to implement effectively in low-resource healthcare systems where infrastructure and computational capacities are limited [8].

Several researchers have proposed methods aimed explicitly at improving the efficiency and practicality of FL for resource-constrained settings. Bonawitz et al. [14] introduced secure aggregation protocols designed to reduce communication overhead and protect against adversarial attacks, thus improving both the security and feasibility of FL systems. In the same line, lightweight encryption schemes and optimized communication strategies have been explored to lower computational demands, making FL more adaptable to settings with limited network bandwidth and processing capabilities [6].

In the context of HIV care, previous studies have utilized centralized machine learning approaches to predict VL suppression, treatment adherence, and retention in care [15, 16]. However, privacy and data sovereignty concerns significantly limit the scalability and real-world deployment of such models. Recent studies highlight the need for federated methodologies that can harness decentralized HIV clinical data without breaching patient confidentiality, ensuring wider adoption and implementation [17].

Domain adaptation techniques have emerged as essential in addressing data heterogeneity, particularly within FL frameworks involving multiple institutions with varied clinical workflows, EMR systems, and data collection standards [18]. Shi and Xu [19] demonstrated that integrating domain adaptation into federated settings substantially improves predictive accuracy and generalizability by harmonizing diverse data sources, effectively reducing model bias and enhancing overall performance across heterogeneous environments.

Building upon these existing studies, our work uniquely integrates robust aggregation methods, lightweight encryption, and domain adaptation within a unified FL framework tailored explicitly to the challenges faced by Uganda's resource-constrained healthcare environment. This targeted approach aims to bridge existing methodological gaps, ensuring secure, privacy-preserving predictive modeling of HIV VL suppression status from decentralized datasets across multiple healthcare institutions.

3. Research Methodology

3.1. Data description and preprocessing

3.1.1. Dataset overview

This study uses a retrospective multi-health facility dataset comprising clinical records from 30 HIV treatment facilities throughout Uganda. The dataset includes 50,000 unique patients and over 1,000,000 individual clinical visit entries. Each entry represents a distinct patient encounter and is labeled with a binary outcome indicating whether the patient's VL was suppressed during that visit, defined according to standard clinical thresholds as VL <200 copies/mL [20]. Predictor variables include demographic and clinical features routinely collected at all sites, that is, sex, age, duration on antiretroviral therapy (ART), number of prior VL tests, prior suppression history, tuberculosis (TB) co-infection status, WHO clinical stage (I–IV), and ART regimen line categorized as first line, second line, or salvage. These features were selected for their documented association with virologic outcomes in prior HIV studies [21, 22] and their availability across all facilities.

3.1.2. Data preprocessing

All data remained in situ at each facility to comply with FL protocols. Preprocessing steps were applied locally under a unified schema. Clinical records with missing outcome labels were excluded. For missing predictor fields, imputation strategies were applied. For example, unrecorded WHO stages were encoded using a placeholder value (e.g., 0), preserving their presence in the dataset while signaling unknown status to the model.

Categorical features were numerically encoded: sex as binary (male = 1, female = 0), ART duration as ordinal values ranging from 0 ("< 6 months") to 4 ("> 5 years"), and ART regimen line as integers representing first line (1), second line (2), or third line/salvage (3) treatment. TB status was similarly binarized (1 for active TB, 0 otherwise), and WHO stages were represented as integers from 1 to 4, with missing values retained as 0. Continuous variables such as age and the number of prior VL tests were normalized using zero-mean, unit-variance normalization, based on statistics calculated from each site's local training subset.

The binary target variable $y_i \in \{0, 1\}$ was defined such that $y_i = 1$ if the VL for visit i was suppressed, and $y_i = 0$ otherwise. To incorporate temporal context, a derived feature was included for each visit indicating whether the patient's most recent VL result prior to the visit was suppressed. Another feature recorded the cumulative number of VL tests prior to the current encounter. These temporal indicators offer proxy insights into longitudinal adherence and clinical monitoring intensity.

To address class imbalance, since approximately 80% of visits resulted in viral suppression, training incorporated class weighting. For each facility k, the weight w applied to the minority (unsuppressed) class was computed such that:

$$\mathrm{w.N_0} = \mathrm{N_1}, \tag{1}$$

where N0 and N1 denote the number of unsuppressed and suppressed samples, respectively, within that facility's dataset. This weighting strategy ensured that the model placed sufficient emphasis on detecting unsuppressed cases, which are clinically critical.

Following preprocessing, each facility $k \in \{1, ..., 30\}$ maintained a local dataset $D_k = \{(x_i, y_i)\}_{i=1}^{nk}$ with $\sum_{k=1}^{30} n_k \approx 1,000,000$. No raw data were transferred or centralized; only model parameters were shared during federated training.

3.2. Federated learning framework

3.2.1. Federated learning approach

We utilized a Federated Averaging (FedAvg) algorithm [7] to train a global model collaboratively across 30 distinct healthcare facilities, each with locally stored data. Let K = 30 represent the number of participating sites, and θ denote the model parameters. The training commenced with a global initialization of θ0, which was either randomly set or pre-trained on publicly available data. In each global round t = 1, 2, ..., T, the central server dispatched the current global model θt−1 to all facilities. Each site k trained its local model using its dataset Dk for E epochs of mini-batch stochastic gradient descent (SGD) and returned updated weights $\theta_t^{(k)}$. The local training objective was to minimize the class-weighted binary cross-entropy loss:

$$L_k(\theta) = -\frac{1}{n_k}\sum_{i=1}^{n_k}\left[y_i \log f_\theta(x_i) + (1 - y_i)\log(1 - f_\theta(x_i))\right], \tag{2}$$

where fθ(x) represents the predicted probability of viral suppression. The global server then aggregated these updates using a weighted average based on data size:

$$\theta_t = \frac{1}{N}\sum_{k=1}^{K} n_k \theta_t^{(k)}, \tag{3}$$

Where $N = \sum_{k=1}^{K} n_k$. This process was iterated until convergence, typically achieved in 20 rounds. To prevent overfitting, each site was limited to E = 1 epoch per round. The entire process was orchestrated using TensorFlow Federated, with secure cryptographic enhancements and compliance with national data governance requirements.

3.2.2. Model architecture

The predictive model was a compact multilayer perceptron (MLP), suitable for tabular data and constrained computational environments. The architecture comprised three hidden layers with 64, 32, and 16 neurons, respectively, each followed by ReLU activations and dropout (rate = 0.2) to prevent overfitting. The input layer accepted 10 features (e.g., age, ART duration, WHO stage; see Section 3.1.1), and the output layer used a sigmoid activation to produce the probability of VL suppression ($P(y = 1/x)$). The model's design balanced expressive power and computational efficiency, ensuring viability on devices with limited processing capacity. Figure 1 summarizes the architectural flow of the MLP used in this study.

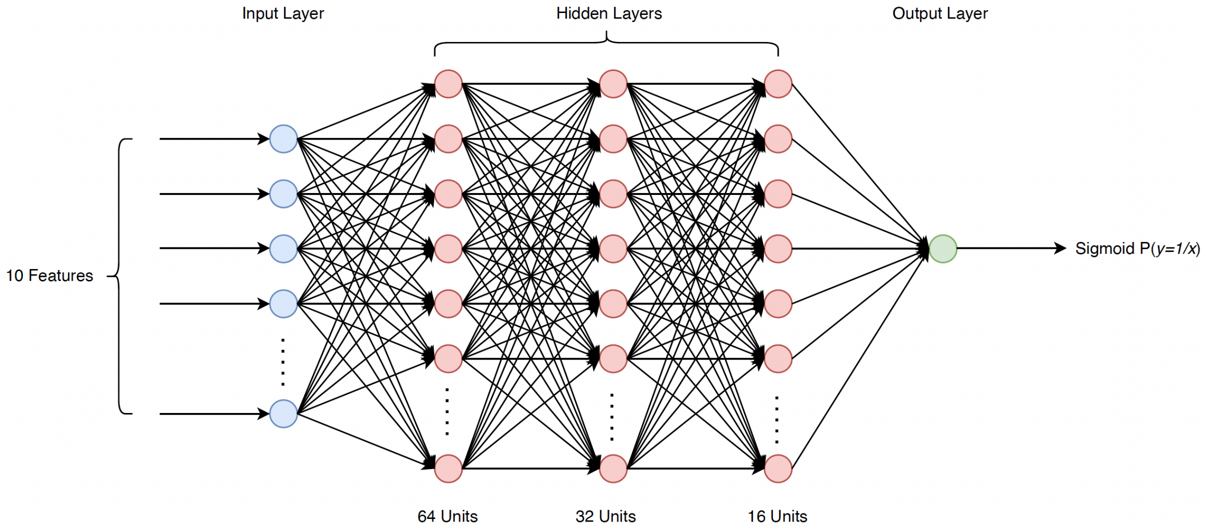

Figure 1. Architectural structure of the study MLP network

3.2.3. Local training strategy

Each site utilized mini-batch SGD (batch size = 128, momentum = 0.9) with an initial learning rate of 0.01, decayed every 5 global rounds. A 10% local validation split was used for early stopping. Model training was conducted in Python using TensorFlow Federated, PySyft for encryption, and TensorFlow Privacy for DP. All operations were executed within a secure national data infrastructure to satisfy residency requirements.

3.2.4. Secure aggregation

To ensure confidentiality during model training, we implemented secure aggregation following the multiparty encryption scheme proposed by Bonawitz et al. [23]. Each facility encrypted its model updates $\theta_t^{(k)}$ with randomly generated masks, ensuring that individual updates remained inaccessible to the server. When summed, these encrypted updates canceled out the noise, yielding $\sum_k \theta_t^{(k)}$ without revealing any single update. This process was realized using PySyft, with robustness to client dropout and minimal network overhead, making it practical in low-bandwidth settings. The approach significantly improves the privacy guarantees beyond what standard FL offers.

3.2.5. Differential privacy

Although FL and secure aggregation limit direct data exposure, model memorization can still pose privacy risks. To address this, we utilized the DP-SGD algorithm [24], which enforces DP by clipping gradients to a fixed norm and adding Gaussian noise. Privacy guarantees were quantified using parameters (ε, δ), with $\varepsilon \approx 1$ and $\delta = 10-5$ achieved across the training rounds, calculated using composition theorems. We adopted user-level privacy, treating all visits from the same patient as a single entity. Despite the noise injection, model performance remained strong, showing only a 2% drop in AUC under strict privacy settings, an acceptable trade-off for ensuring patient confidentiality.

3.2.6. Handling data heterogeneity with domain adaptation

To mitigate the effects of non-identically distributed (non-IID) data across facilities, we integrated a domain adaptation mechanism based on the Domain-Adversarial Neural Network (DANN) approach [25]. The model was extended with a secondary domain classifier branch trained adversarially to discourage domain-specific feature learning. This was achieved using a gradient reversal layer and a composite loss function:

$$\mathscr{L}_{\text{total}} = \mathscr{L}_{\text{prediction}} - \lambda \mathscr{L}_{\text{domain}}, \tag{4}$$

where λ was annealed from 0.5 to 0 during training to shift focus from domain-invariance to task performance gradually. Facilities used anonymized average feature representations exchanged via the server to simulate negative domain samples in a privacy-preserving manner. This method harmonized internal representations across disparate clinical contexts, improving generalization, especially for outlier sites such as pediatric clinics and TB co-treatment centers. Figure 2 illustrates the FL pipeline, integrating local training, secure aggregation, and domain adaptation.

3.2.7. Baseline models for comparison

To evaluate the FL framework, we compared it against two baselines: a centralized model trained on pooled data from all facilities, and independent local models trained separately at each site. All models used the same MLP architecture and training configuration. The centralized model served as an upper-bound reference, while local models reflected isolated training on smaller datasets without inter-site collaboration.

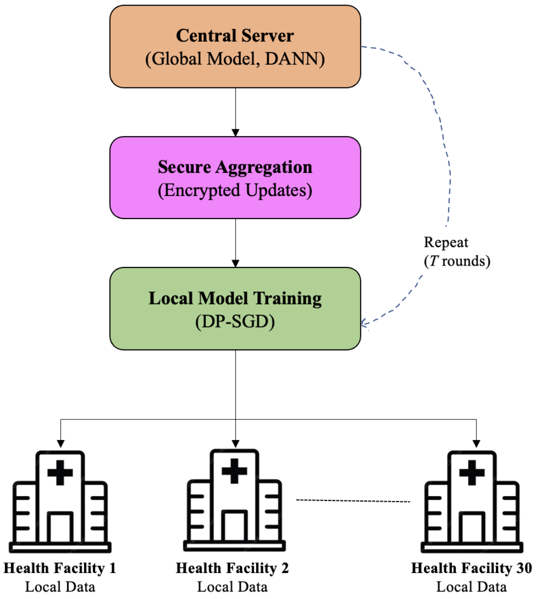

Figure 2. Federated learning framework was used in this study. Local models are trained independently at each health facility using DP-SGD on site-specific data. Encrypted model updates are aggregated securely and used to update a global model on the central server, which incorporates domain-adversarial training (DANN) to mitigate heterogeneity. The process is repeated over multiple rounds without sharing raw data.

3.3. Evaluation methods

Model evaluation was conducted using held-out test data to compare the FL framework against two baselines: centralized and local models. At each of the 30 facilities, 20% of records were set aside as a test set. From the remaining 80%, 10% was used for validation during training. The FL model was trained without transferring data, while the centralized model was trained on pooled data (as a benchmark), and local models were trained independently per site using the same architecture.

All models were assessed on the combined test set (200,000 records), with local models evaluated only on their respective sites' data. Metrics included accuracy, sensitivity, specificity, precision, recall, and area under the curve receiver operating characteristics (AUC-ROC). Sensitivity (TPR) and specificity (TNR) were computed with respect to viral suppression (VL < 200 copies/mL). Macro-averaging was used to aggregate per-site results.

ROC curves were plotted by sweeping classification thresholds, and AUCs were computed as threshold-independent measures. To test statistical significance, McNemar's test and DeLong's test were applied to accuracy and AUC differences, respectively [26].

We also conducted ablation studies to isolate the effects of DP and domain adaptation. Each model variant was trained under identical conditions on a CPU-only server, with an average training time of two minutes per global round and full convergence in under an hour.

4. Results

4.1. Model performance overview

We compared the predictive performance of the FL model against a centralized model trained on pooled data and individual local models trained independently at each facility. Table 1 summarizes the key evaluation metrics across these approaches.

The FL model achieved an AUC of 0.874, closely matching the centralized model (AUC = 0.881), with a non-significant difference according to DeLong's test ($p = 0.37$). Importantly, FL substantially outperformed the average local model (AUC = 0.758), highlighting the value of collaborative training even in the absence of centralized data sharing.

Figure 3 presents the ROC curve of the FL model compared to a representative local model, showing superior discrimination with consistently higher true positive rates across all thresholds. The accompanying bar graph illustrates the performance metrics, accuracy, sensitivity, specificity, precision, and AUC, demonstrating the FL model's overall advantage in predictive power and robustness across key evaluation measures.

4.2. Per-facility analysis

The federated model consistently outperformed local models across 28 of 30 facilities. Gains were particularly notable at smaller clinics (<1000 patients), where local models often failed to generalize due to limited data. For example, at a Health Centre II (lower-level facility) with low unsuppressed cases, the local model achieved only 60% accuracy and AUC $\approx$ 0.60, whereas the FL model reached AUC $\approx$ 0.85. Calibration analysis showed the FL model was better aligned with true probabilities across sites, unlike local models, which were often overconfident due to skewed training distributions.

4.3. Impact of privacy and security measures

Secure aggregation introduced negligible overhead and no measurable performance degradation. Incorporating differential privacy (DP) at a stringent budget of $\varepsilon \approx 1$ resulted in a small AUC reduction (~0.03) and a 2% drop in accuracy. However, the DP-enabled model (AUC = 0.844) still outperformed all local models, providing a strong privacy-utility trade-off. Simulated membership inference attacks confirmed that with DP, an adversary's ability to detect if a patient's data was used was no better than chance, validating the privacy protection.

4.4. Domain adaptation and heterogeneity mitigation

To address data heterogeneity caused by differences in EMR systems and clinical workflows across facilities, we applied adversarial domain adaptation. Its effectiveness was assessed by comparing model performance with and without domain adaptation. When applied, per-facility AUC range narrowed from [0.80, 0.89] to [0.84, 0.88], indicating more consistent performance. Facilities with distributions showed notable AUC gains (e.g., from 0.80 to 0.86). These results demonstrate that domain adaptation improved generalization and reduced performance variability across heterogeneous sites.

5. Discussion

This study presents compelling evidence that FL can serve as a viable and privacy-preserving alternative to centralized machine learning for clinical predictive modeling in low-resource healthcare settings. Our findings show that the federated model achieved predictive accuracy nearly indistinguishable from a centrally trained

Table 1. Performance comparison across model types on the held-out test set

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	AUC
Federated (Global FL)	83.1	89.6	66.8	84.7	0.874
Centralized	84.0	92.1	60.5	82.9	0.881
Local (Average)	75.2	78.4	55.3	79.8	0.758
Local (Best)	80.5	87.9	54.8	81.5	0.812
Local (Worst)	60.8	50.0	80.0	66.7	0.645

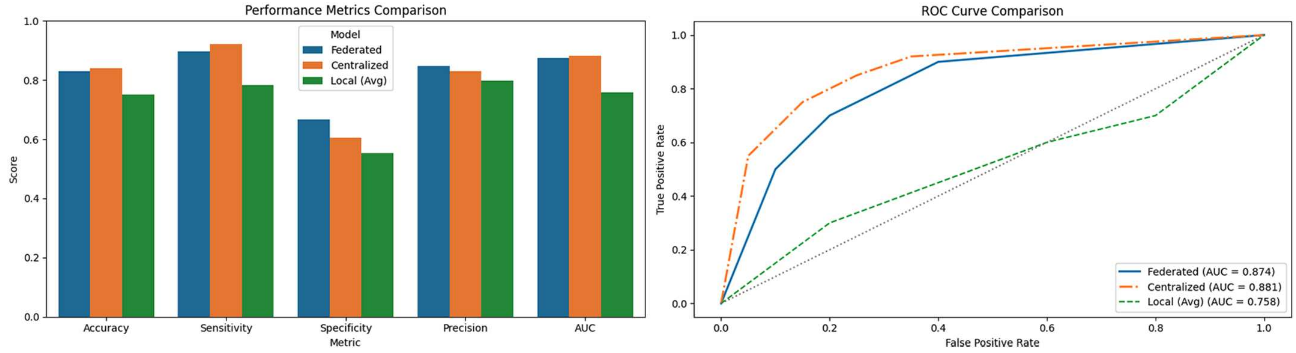

Figure 3. (Left): ROC curves comparing FL (solid) with a representative local model (dashed). (Right): Model comparison across metrics

model (AUC 0.874 vs. 0.881), despite never accessing raw patient data. This result highlights the maturity of FL as a practical machine learning paradigm capable of high-performance learning under stringent data governance constraints. Particularly in sub-Saharan Africa, where data fragmentation, privacy laws, and limited digital infrastructure hinder centralized artificial intelligence (AI) approaches, FL offers a pathway to unlock the value of distributed health records for population-level insights and individual-level care improvement [4, 5].

Beyond privacy, FL demonstrated superior generalizability compared to models trained locally at individual facilities. Smaller clinics with limited data, especially those in rural areas or serving special populations, saw marked gains in model performance through participation in the federated system. For example, facilities with fewer than 1,000 patients, which produced local models with AUCs as low as 0.60, benefited substantially from the shared model, which improved performance to ~0.85. This finding illustrates how FL can improve equity in clinical AI, allowing under-resourced sites to access the predictive strength of models trained on much larger datasets without compromising ownership or control over their data. Furthermore, the integration of adversarial domain adaptation allowed the model to learn representations that generalized across heterogeneous site distributions, narrowing per-site AUC ranges and correcting underperformance on specialized facilities like pediatric or TB/HIV co-treatment clinics [5, 25].

Our study also emphasizes the importance of embedding robust privacy techniques into FL pipelines. Secure aggregation prevented the server from accessing individual model updates, while DP provided strong protections against inference attacks from the final model. The application of user-level DP ensured that even patients with multiple records were protected as a unit, a critical feature for chronic disease datasets with longitudinal follow-ups. Despite these protections, performance degradation was minimal. The AUC declined by only ~0.03 under strong privacy settings ($\epsilon \approx 1$). This outcome challenges the common assumption that privacy must come at the expense of utility and suggests that, at least for structured clinical data, well-tuned DP mechanisms can coexist with high model fidelity [7, 24]. These privacy guarantees not only enable legal compliance under frameworks like Uganda's Data Protection and Privacy Act and the general data protection regulation (GDPR) but also strengthen stakeholder trust, which is vital for sustained deployment.

While these findings are encouraging, our evaluation was limited to Uganda's health system. The framework's performance in other countries or disease areas may vary depending on data standards, governance, and infrastructure. Future work should assess transferability across regions and explore integration with clinical workflows. Beyond geography, the other limitations include, the model was evaluated on retrospective data and not tested in real-time clinical workflows. All test sites participated in training, so external generalizability to unseen institutions remains unknown. The framework also assumes full client participation in each round, which may not hold in practice due to intermittent connectivity or system constraints. Finally, calibration metrics were not evaluated, limiting the assessment of the model's reliability for clinical risk estimations.

6. Conclusion

This study presents a secure, efficient FL framework tailored for predictive modeling in resource-constrained healthcare systems, with a focus on HIV VL suppression across multiple facilities in Uganda. By combining federated averaging, DP, secure aggregation, and domain adaptation, we achieved predictive performance nearly equivalent to centralized learning (AUC ~0.87 vs. ~0.88) while preserving patient data privacy and local data sovereignty. The model consistently outperformed locally trained models, particularly in smaller clinics, demonstrating the power of collaborative learning across disparate and heterogeneous sources. Furthermore, the adversarial domain adaptation mechanism effectively mitigated inter-facility variability, resulting in more equitable performance across settings and reducing model bias associated with data distribution shifts.

Beyond HIV care, the architecture developed in this study is widely applicable to other domains where sensitive, siloed data and limited infrastructure challenge the deployment of centralized AI. Our approach, based on compact neural networks, fault-tolerant secure aggregation, and privacy-aware training, ran efficiently on commodity hardware without sacrificing accuracy. It is compatible with existing health information systems and amenable to enhancements such as explainable AI (e.g., Shapley Additive exPlanations) and per-site fine-tuning [27]. In demonstrating that high-performance AI can be achieved without centralized data collection, this work sets a precedent for ethical, scalable, and equitable machine learning in global health, paving the way for cross-institutional AI collaborations in low-resource settings.

Acknowledgements

The authors acknowledge that an earlier version of this manuscript was published as a preprint at medRxiv (DOI: https://doi.org/10.1101/2025.07.27.25332284) on 28 July 2025 and SSRN (DOI: http://dx.doi.org/10.2139/ssrn.5404037) on 25 August 2025. The preprint has not been peer reviewed. This published article incorporates revisions made during the peer review process for Medinformatics.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support this work are available upon reasonable request to the corresponding author.

Author Contribution Statement

Alex Mirugwe: Conceptualization, Methodology, Software, Formal analysis, Resources, Data curation, Writing – original draft, Visualization. **Juwa Nyirenda:** Methodology, Validation, Resources, Writing – review & editing.

References

[1] Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., . . . , & Liu, Y. (2021). Deep learning-enabled medical computer vision. *npj Digital Medicine*, *4*(1), 5. https://doi.org/10.1038/s41746-020-00376-2

[2] Huerta, E. A., Khan, A., Davis, E., Bushell, C., Gropp, W. D., Katz, D. S., ..., & Saxton, A. (2020). Convergence of

artificial intelligence and high performance computing on NSF-supported cyberinfrastructure. *Journal of Big Data*, *7*(1), 88. https://doi.org/10.1186/s40537-020-00361-2

[3] Ngiam, K. Y., & Khor, I. W. (2019). Big data and machine learning algorithms for health-care delivery. *The Lancet Oncology*, *20*(5), e262–e273. https://doi.org/10.1016/S1470-2045(19)30149-4

[4] Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., ..., & Cardoso, M. J. (2020). The future of digital health with federated learning. *npj Digital Medicine*, *3*(1), 119. https://doi.org/10.1038/s41746-020-00323-1

[5] Kaissis, G. A., Makowski, M. R., Rückert, D., & Braren, R. F. (2020). Secure, privacy-preserving and federated machine learning in medical imaging. *Nature Machine Intelligence*, *2*(6), 305–311. https://doi.org/10.1038/s42256-020-0186-1

[6] Xu, J., Glicksberg, B. S., Su, C., Walker, P., Bian, J., & Wang, F. (2021). Federated learning for healthcare informatics. *Journal of Healthcare Informatics Research*, *5*(1), 1–19. https://doi.org/10.1007/s41666-020-00082-4

[7] Sun, Y., Ochiai, H., & Esaki, H. (2022). Decentralized deep learning for multi-access edge computing: A survey on communication efficiency and trustworthiness. *IEEE Transactions on Artificial Intelligence*, *3*(6), 963–972. https://doi.org/10.1109/TAI.2021.3133819

[8] Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine*, *37*(3), 50–60. https://doi.org/10.1109/MSP.2020.2975749

[9] Sheller, M. J., Reina, G. A., Edwards, B., Martin, J., & Bakas, S. (2019). Multi-institutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, 92-104.* https://doi:10.1007/978-3-030-11723-8_9

[10] Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology*, *10*(2), 12. https://doi.org/10.1145/3298981

[11] Nasr, M., Shokri, R., & Houmansadr, A. (2019). Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning. In *2019 IEEE Symposium on Security and Privacy,* 739–753. https://doi.org/10.1109/SP.2019.00065

[12] Bagdasaryan, E., Veit, A., Hua, Y., Estrin, D., & Shmatikov, V. (2020). How to backdoor federated learning. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, 2938–2948. https://proceedings.mlr.press/v108/bagdasaryan20a.html

[13] Amin, M. S., Ahmad, S., & Loh, W.-K. (2025). Federated learning for Healthcare 5.0: A comprehensive survey, taxonomy, challenges, and solutions. *Soft Computing*, *29*(2), 673–700. https://doi.org/10.1007/s00500-025-10508-z

[14] Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., ..., & Seth, K. (2016). Practical secure aggregation for federated learning on user-held data. In *30th Conference on Neural Information Processing Systems.*

[15] Marcus, J. L., Sewell, W. C., Balzer, L. B., & Krakower, D. S. (2020). Artificial intelligence and machine learning for HIV prevention: Emerging approaches to ending the epidemic. *Current HIV/AIDS Reports*, *17*(3), 171–179. https://doi.org/10.1007/s11904-020-00490-6

[16] Chiramba, N. W., Ndlovu, B., Dube, S., Kiwa, F. J., & Muduva, M. (2024). Optimizing antiretroviral therapy (ART) adherence through predictive analytics using machine learning techniques. In *5th South American Industrial Engineering and Operations Management Conference.* https://doi.org/10.46254/SA05.20240195

[17] Brisimi, T. S., Chen, R., Mela, T., Olshevsky, A., Paschalidis, I. Ch., & Shi, W. (2018). Federated learning of predictive models from federated Electronic Health Records. *International Journal of Medical Informatics*, *112*, 59–67. https://doi.org/10.1016/j.ijmedinf.2018.01.007

[18] Peng, X., Huang, Z., Zhu, Y., & Saenko, K. (2020). Federated adversarial domain adaptation. *In International Conference on Learning Representations.*

[19] Shi, Y., & Xu, X. (2022). Deep federated adaptation: An adaptative residential load forecasting approach with federated learning. *Sensors*, *22*(9), 3264. https://doi.org/10.3390/s22093264

[20] World Health Organization. (2021). Consolidated guidelines on HIV prevention, testing, treatment, service delivery and monitoring: Recommendations for a public health approach. https://www.who.int/publications/i/item/9789240031593

[21] Pyngottu, A., Scherrer, A. U., Kouyos, R., Huber, M., Hirsch, H., Perreau, M., ..., & Günthard, H. F. (2021). Predictors of virological failure and time to viral suppression of first-line integrase inhibitor–based antiretroviral treatment. *Clinical Infectious Diseases*, *73*(7), e2134–e2141. https://doi.org/10.1093/cid/ciaa1614

[22] Maskew, M., Sharpey-Schafer, K., de Voux, L., Crompton, T., Bor, J., Rennick, M., ..., & Pisa, P. (2022). Applying machine learning and predictive modeling to retention and viral suppression in South African HIV treatment cohorts. *Scientific Reports*, *12*(1), 12715. https://doi.org/10.1038/s41598-022-16062-0

[23] Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., ..., & Seth, K. (2017). Practical secure aggregation for privacy-preserving machine learning. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 1175–1191. https://doi.org/10.1145/3133956.3133982

[24] Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 308–318. https://doi.org/10.1145/2976749.2978318

[25] Liu, X., Yoo, C., Xing, F., Oh, H., El, Fakhri, G., Kang, J.-W., & Woo, J. (2022). Deep unsupervised domain adaptation: A review of recent advances and perspectives. *APSIPA Transactions on Signal and Information Processing*, *11*(1), e25. https://doi.org/10.1561/116.00000192

[26] Nahm, F. S. (2022). Receiver operating characteristic curve: Overview and practical use for clinicians. *Korean Journal of Anesthesiology*, *75*(1), 25–36. https://doi.org/10.4097/kja.21209

[27] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems,* 4768–4777.

How to Cite: Mirugwe, A., & Nyirenda, J. (2025). Secure and Efficient Federated Learning for Predictive Modeling in Resource-Constrained Healthcare Systems. *Medinformatics*. https://doi.org/10.47852/bonviewMEDIN52027621