Makerere University

College of Computing and Information Sciences

Department of Computer Science

A Blockchain-Secured Federated Learning System for Detection and Early Warning of Foot-and-Mouth Disease in Ugandan Cattle Farms

Research Proposal

By

Muhindo Mubaraka

Student Number: 2400725633

Registration Number: 2024/HD05/25633U

muhindo.mubaraka@students.mak.ac.ug

Supervisor

Dr. Chongomweru Halimu

Department of Computer Science

A Research Proposal Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Computer Science

January 25, 2026

Abstract

Training machine learning models traditionally requires collecting all data in one central location. While this works in well-resourced environments with reliable internet, it creates problems in resource-limited settings. Privacy concerns prevent farmers and health facilities from sharing sensitive data, and poor connectivity makes uploading large datasets impractical. Centralizing data also creates security risks and trust issues when multiple organizations need to collaborate.

Federated learning offers an alternative by training models on distributed data without moving it from its source. Each participant trains locally and shares only model updates, which are aggregated to create a global model. However, this introduces a trust problem: participants cannot verify whether aggregation was done correctly or whether some contributed manipulated updates.

This research addresses the trust problem by integrating blockchain technology with federated learning. The blockchain provides a transparent, tamper-proof record of the training process, allowing participants to verify that all steps were executed correctly. This study designs and evaluates such a system specifically for resource-limited environments, using Foot-and-Mouth Disease prediction in Ugandan cattle farms as a practical test case.

The system demonstrates that collaborative machine learning can work in settings with limited resources and unreliable connectivity while maintaining privacy and trust. The architecture balances security with computational efficiency, making blockchain-verified federated learning practical for developing regions. This has broader implications for privacy-preserving AI in agriculture, healthcare, and other domains where data sharing faces technical and trust barriers.

Keywords: Federated Learning, Blockchain, Privacy-Preserving Machine Learning, Disease Prediction, Distributed Systems, Differential Privacy, Byzantine Fault Tolerance

Contents

1 Introduction

1.1 Background of the Study

Traditional machine learning requires collecting all data in one central location for training. In many real-world situations, especially in low-resource regions, this approach is not practical. There are several reasons why this does not work well.

First, privacy concerns make people reluctant to share their data. For example, a livestock farmer may fear to share photos or health records of their animals because they worry that government auditors or the ministry of agriculture will identify their farm and penalize them for any violations or diseases found. Similarly, hospitals are hesitant to share patient data because of confidentiality laws and fear of exposing sensitive medical information that could be traced back to individual patients or institutions.

Second, many organizations do not have the network and computing resources to send large amounts of data to a central server. For example, a rural health clinic with slow internet connection cannot afford to upload thousands of high-resolution medical images every day. The upload would take too long, consume too much bandwidth, and may fail frequently due to unreliable connectivity. This makes centralized machine learning not practical for such settings [2].

Federated learning (FL) solves these problems by keeping data on local devices and only sharing model updates instead of raw data. Each participant trains a model on their own data, then sends only the learned patterns (model updates) to a central server. The server combines these updates to create a global model without ever accessing the actual data [1, 3].

However, FL has a critical trust issue. When participants send their model updates, there is no easy way to verify if someone is sending corrupted or fake updates that could damage the model. This research investigates how blockchain technology can provide transparent verification of the collaborative learning process, ensuring that all contributions are recorded immutably and the aggregation process is auditable, while keeping the system light enough to work in places with limited resources.

To demonstrate and evaluate the proposed system, this research uses Foot-and-Mouth Disease (FMD) prediction in Uganda as a practical test case. FMD is a highly contagious livestock disease that causes significant economic losses in Uganda through reduced milk production, trade restrictions, and animal morbidity. The disease data is naturally distributed across different districts, making it an ideal scenario to test a privacy-preserving federated learning system. This application context allows the research to validate the technical contributions under real-world conditions typical of resource-limited environments.

1.2 Problem Statement

Federated learning has a trust problem. When multiple participants train a model together, there is no reliable way to verify if someone is sending honest updates or corrupted data. A malicious participant could poison the entire model by sending fake updates. Similarly, participants cannot verify if the central server is actually using their contributions or ignoring them.

Federated learning was designed to work in low-resource environments where participants have limited computing power and unreliable network connections. However, there is no lightweight verification mechanism that can provide trust and accountability in these resource-limited settings.

The core problem is: how can a verification mechanism be designed that provides trust and accountability in federated learning, while remaining lightweight enough to work in resource-constrained environments? This research addresses this gap by designing a practical blockchain-based verification system for federated learning in resource-limited settings.

1.3 Research Objectives

1.3.1 Main Objective

To design and evaluate a blockchain-based federated learning system with early warning capabilities for Foot-and-Mouth Disease in Ugandan cattle farms.

1.3.2 Specific Objectives

1. To collect, clean, and preprocess Foot-and-Mouth Disease data from the Uganda Livestock Information Tracking System (ULITS) for use in federated learning experiments.

2. To design a system architecture that integrates federated learning with blockchain for secure and verifiable model aggregation.

3. To implement a lightweight prototype that works efficiently in resource-limited environments typical of Ugandan cattle farms.

4. To evaluate the system performance in terms of prediction accuracy, computational efficiency, and security against malicious attacks.

5. To assess the system's practical effectiveness through field testing with real Foot-and-Mouth Disease data from Ugandan cattle farms.

1.4 Research Questions

1. How can data quality be ensured through appropriate collection, cleaning, and preprocessing of FMD records for use in federated learning?

2. How can blockchain technology be integrated with federated learning to provide secure and verifiable model aggregation for early warning of Foot-and-Mouth Disease?

3. What system architecture will work efficiently in resource-limited environments typical of Ugandan cattle farms while maintaining prediction accuracy?

4. How does the system perform in terms of FMD prediction accuracy, computational cost, and security against malicious participants?

5. What is the practical effectiveness of the system when tested with real Foot-and-Mouth Disease data from Ugandan cattle farms?

1.5 Significance of the Study

This research addresses the trust problem in federated learning by combining it with blockchain technology to create a verifiable and secure system for collaborative machine learning in resource-limited environments.

The significance of this study extends across multiple dimensions:

Scientific Contribution: This research provides a practical solution to the trust problem in federated learning by integrating blockchain verification in a way that works efficiently in resource-limited settings. It shows how to balance security with computational efficiency.

Technical Innovation: The system demonstrates that blockchain-based federated learning can work in environments with limited resources and unreliable connectivity. This opens possibilities for similar applications in other low-resource settings.

Practical Application: By addressing Foot-and-Mouth Disease prediction in Ugandan cattle farms, this work solves a real problem in the livestock sector. Early warning can help farmers and veterinary officers respond quickly to disease outbreaks, reducing economic losses.

Data Privacy and Trust: The system allows multiple districts to collaborate on disease prediction without sharing raw data. The blockchain layer provides proof that the learning process was not manipulated, addressing the trust issue that prevents data collaboration.

Broader Applicability: While tested with FMD data in Uganda, the architecture can be adapted to other diseases and other countries facing similar challenges with distributed data and limited resources.

1.6 Scope and Limitations

1.6.1 Scope

This research focuses on:

- **Geographic Focus:** Uganda, specifically districts that participate in the Uganda Livestock Information Tracking System (ULITS) which covers FMD data collection
- **Disease Focus:** Foot-and-Mouth Disease as the test case for the early warning system
- **Data Period:** Historical FMD outbreak data from 2011-2024 for training and testing the system
- **Technical Focus:** Design and evaluation of a system that integrates federated learning with blockchain verification, specifically for resource-limited environments

1.6.2 Limitations

- **Data Quality:** The accuracy of predictions depends on the quality and completeness of FMD data collected by veterinary officers in different districts. Incomplete or inaccurate data will affect the system's performance.
- **Network Requirements:** While designed for low-connectivity environments, the system still needs periodic internet access for participants to share model updates and synchronize with the blockchain.
- **Scale of Testing:** The evaluation will involve up to 50 participating nodes (representing districts). Testing at much larger scales is beyond the scope of this research.
- **Trust Assumption:** The security mechanisms assume that most participants (at least 51%) act honestly. If the majority of participants are malicious, the system's security guarantees may not hold.
- **Single Disease:** While the system architecture can potentially work for other diseases, this research focuses only on Foot-and-Mouth Disease. Testing with other diseases is not included.

2 Literature Review

Existing research on federated learning, blockchain technology, and their integration for trustworthy distributed machine learning is examined. The review is organized around key themes: privacy-preserving machine learning, trust verification in distributed systems, blockchain integration, and practical applications in resource-limited settings.

2.1 Federated Learning and Privacy Preservation

Federated learning has gained attention as a way to train machine learning models without collecting data in one central location [1]. The basic idea is simple: instead of sending data to a server, each participant trains a model locally and only shares the model updates. The server aggregates these updates to improve the global model [2].

This approach addresses privacy concerns because raw data never leaves its source. For example, hospitals can collaborate on disease prediction without sharing patient records, and farms can work together on pest detection without exposing their operational data. This is particularly valuable in sectors where data is sensitive or regulations prevent data sharing.

However, while federated learning protects privacy during training, it does not solve the trust problem. Participants cannot verify whether the aggregation was done correctly or whether some participants sent manipulated updates. The system relies on trusting the central server, which becomes a single point of failure. Several researchers have pointed out this limitation [3], noting that federated learning alone does not provide proof that the process was fair or that the final model was not compromised.

2.2 Trust and Verification Challenges

The trust problem in federated learning becomes critical in real-world applications. When multiple organizations collaborate on model training, they need assurance that the process is transparent and verifiable. Without this, participants may hesitate to join the system, especially when decisions based on the model have significant consequences.

Recent work has explored different approaches to address trust. Some researchers propose cryptographic methods like secure multi-party computation and homomorphic encryption to verify computations without revealing data. Others suggest reputation systems where participants build trust over time based on their behavior. However, these approaches either add significant computational overhead or take time to establish trust, neither of which works well in resource-limited environments that need quick deployment.

The agricultural sector in developing countries presents a particular challenge. Veterinary officers collecting disease data in different districts may not trust a centralized

system controlled by one authority. They need transparency about how their data contributes to predictions and confidence that their contributions are properly acknowledged. This trust gap prevents many potential collaborative systems from being adopted.

2.3 Blockchain Integration for Verifiable Federated Learning

Blockchain technology has emerged as a potential solution to the trust problem in federated learning. The idea is to use blockchain as a transparent, tamper-proof record of the training process. Each model update, aggregation step, and participant contribution can be recorded on the blockchain, creating an auditable history.

Several recent studies have explored this integration. Zhang et al. [4] proposed combining blockchain with federated learning using cryptographic methods like verifiable random functions and zero-knowledge proofs. Their system aims to ensure fairness and privacy during model training. However, their work focuses on the theoretical framework and does not address practical deployment in resource-limited environments.

Chen et al. [5] developed a system called FLock that uses blockchain state channels to prevent model poisoning attacks. Their approach shows that blockchain can help detect malicious participants who try to corrupt the model. The system is robust against attacks, but it is designed for environments with good network connectivity and sufficient computational resources.

What emerges from these studies is that blockchain can indeed provide transparency and verification for federated learning. The challenge is making it work efficiently in settings where resources are limited. Most proposed systems assume stable internet connections and powerful devices, which does not match the reality in many developing regions.

2.4 Applications in Disease Prediction and Healthcare

Federated learning has been applied to healthcare and disease prediction with promising results. Teo et al. [6] reviewed 612 studies on federated learning in healthcare and found that it enables hospitals to collaborate on diagnostic models without violating patient privacy. The review highlights successes in medical image analysis, disease outbreak prediction, and drug discovery.

However, most healthcare applications focus on well-resourced medical institutions in developed countries. The infrastructure requirements, including high-speed networks and powerful computing devices, are often not available in rural clinics or agricultural settings in developing regions.

In the livestock sector, Kapalaga et al. [7] created a unified dataset for Foot-and-Mouth Disease prediction in Uganda. Their work addresses data fragmentation by combining

records from multiple sources into one national dataset. They show that machine learning models can predict FMD outbreaks based on factors like rainfall and temperature. However, their approach is centralized, requiring all data to be collected in one location. This creates privacy concerns and network bottlenecks, especially in areas with poor connectivity.

What is missing from the literature is a system that combines the privacy benefits of federated learning with the trust guarantees of blockchain, specifically designed for disease prediction in resource-limited agricultural settings. The existing work either focuses on healthcare in developed countries or uses centralized approaches that do not address privacy and trust issues.

2.5 Resource-Constrained Environments and Computational Efficiency

An important consideration for any system deployed in developing regions is computational efficiency. Many proposed blockchain-federated learning systems use complex cryptographic protocols that require significant processing power and memory. This works in data centers but not on the mobile devices or low-power computers typically available in rural areas.

Some researchers have started exploring lightweight approaches. Edge computing techniques show promise for running machine learning models on resource-limited devices. Model compression and quantization can reduce the size and computational requirements of neural networks. However, most of this work focuses on inference (using trained models) rather than training, and little attention has been paid to making the blockchain verification layer lightweight.

The gap here is clear: a system architecture is needed that balances security and verifiability with computational efficiency. The system must work with intermittent internet connectivity, limited processing power, and minimal storage capacity while still providing the transparency and trust benefits of blockchain.

2.6 Summary of Reviewed Literature

Table 1 summarizes the key studies reviewed, highlighting their main contributions and limitations relevant to this research.

2.7 Summary of Research Gaps

Through this literature review, three critical gaps that this research addresses are identified:

Table 1: Summary of Key Literature

Study	Main Contribution	Limitation
McMahan et al. [1]	Introduced federated averaging algorithm for decentralized training	No trust verification mechanism
Kairouz et al. [2]	Comprehensive review of FL challenges and open problems	Does not address blockchain integration
Bonawitz et al. [3]	Secure aggregation protocol for privacy preservation	High computational overhead
Zhang et al. [4]	Blockchain-FL integration with cryptographic verification	Not designed for resource-limited settings
Chen et al. [5]	State channels for attack prevention in FL	Requires stable network connectivity
Teo et al. [6]	Review of FL in healthcare applications	Focus on well-resourced institutions
Kapalaga et al. [7]	Unified FMD dataset for Uganda	Centralized approach, privacy concerns

1. **Trust Verification Gap:** While federated learning provides privacy, existing systems lack built-in mechanisms for participants to verify that model aggregation was done correctly and that all participants followed the protocol. Current solutions either rely on trust in a central authority or use cryptographic methods too heavy for resource-limited settings.

2. **Practical Deployment Gap:** Most proposed blockchain-federated learning systems are designed for well-resourced environments. There is no system specifically designed for resource-limited settings with poor connectivity, limited devices, and distributed data sources typical of developing regions.

3. **Application Validation Gap:** Existing blockchain-FL integrations lack real-world validation in challenging environments. This research uses Foot-and-Mouth Disease prediction in Uganda as a test case to validate the proposed system under realistic conditions, demonstrating that the technical contributions work in practice.

This review shows that both federated learning and blockchain technologies have advanced significantly, but their integration for practical systems in resource-constrained environments remains unexplored. The trust problem in federated learning is well recognized but not adequately solved for resource-limited settings.

3 Research Methodology

This section describes the methodology for designing, implementing, and evaluating a blockchain-secured federated learning system for resource-limited environments. The primary contribution is to the science of privacy-preserving distributed machine learning, specifically addressing the trust verification problem in federated learning. Foot-and-Mouth Disease prediction in Uganda serves as a practical test case to validate the system under real-world conditions.

3.1 Research Design

This study adopts a design science research methodology, which is appropriate for creating and evaluating new artifacts in information systems. The research follows an iterative process of problem identification, objectives definition, design and implementation, demonstration, evaluation, and communication.

3.2 Data Collection, Cleaning, and Preprocessing

This involves obtaining, cleaning, and preparing FMD data for federated learning experiments.

3.2.1 Data Source

FMD outbreak records will be obtained from the Uganda Livestock Information Tracking System (ULITS), which contains surveillance data collected by district veterinary officers across Uganda from 2011 to 2024. The dataset includes information on outbreak locations, dates, herd sizes, clinical signs observed, vaccination history, and environmental factors such as rainfall and temperature.

3.2.2 Data Cleaning Process

Data cleaning will involve several steps:

- **Missing Value Treatment:** Records with missing critical fields (location, date, diagnosis) will be flagged. Numerical missing values will be imputed using median imputation for continuous variables and mode imputation for categorical variables where appropriate.
- **Duplicate Removal:** Duplicate records arising from multiple reporting channels will be identified using fuzzy matching on location, date, and herd identifiers.
- **Outlier Detection:** Statistical methods will be used to identify implausible values (e.g., negative animal counts, future dates) which will be corrected or removed.

- **Standardization:** Date formats, location names, and categorical variables will be standardized across all district records.

3.2.3 Data Preprocessing

The cleaned data will be preprocessed for machine learning:

- **Feature Engineering:** Temporal features (month, season, year), spatial features (district, region), and derived features (outbreak frequency, proximity to previous outbreaks) will be created.
- **Normalization:** Numerical features will be normalized using min-max scaling to ensure comparable ranges.
- **Encoding:** Categorical variables will be encoded using one-hot encoding or label encoding as appropriate for the prediction model.
- **Train-Test Split:** The dataset will be split temporally, with earlier years used for training and recent years reserved for testing to simulate real-world prediction scenarios.

3.3 System Architecture Design

The system architecture integrates federated learning with blockchain for secure and verifiable model aggregation.

3.3.1 Layered Architecture

A layered architecture will be designed that strategically separates concerns:

- **Data Layer:** Local FMD data remains on district systems with standardized interfaces for secure access. Each participating district maintains control over its raw data.
- **Federated Learning Layer:** Model training occurs locally at each district node. Only model updates (gradients or weights) are shared, not raw data. PySyft framework will be used for implementing federated averaging with secure aggregation.
- **Blockchain Layer:** Hyperledger Fabric will provide the trust foundation. Each model update submission and aggregation step will be recorded as transactions on the blockchain, creating an immutable audit trail.
- **Application Layer:** An early warning dashboard built with Django backend and React.js frontend will present predictions to veterinary officers with explanations.

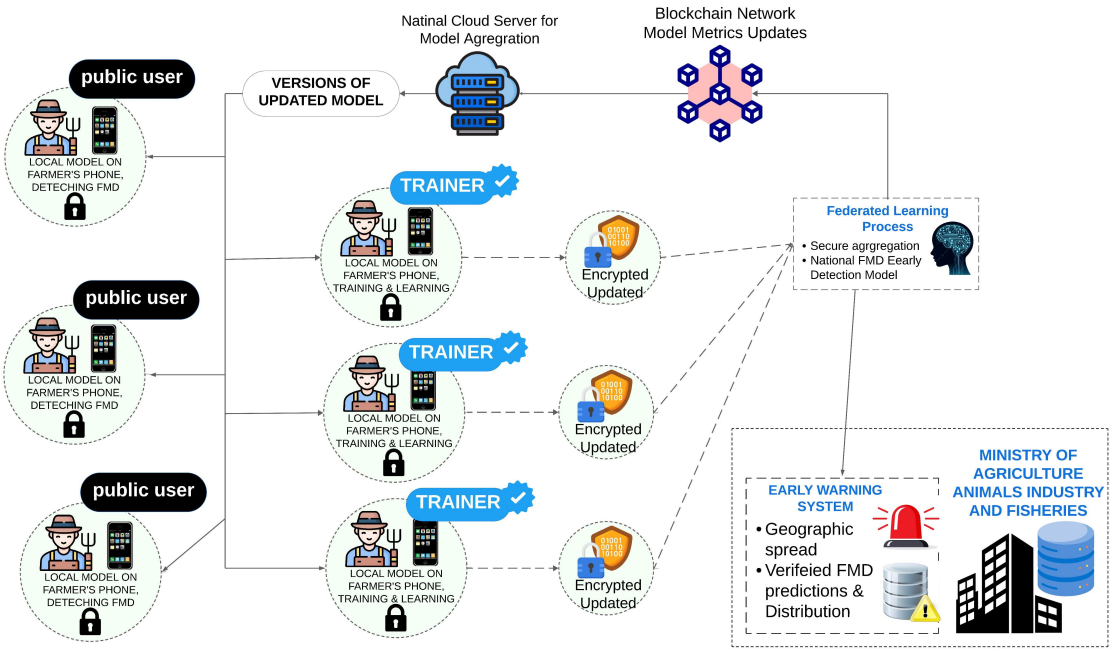


Figure 1: System Architecture Design showing the layered approach with Data Layer (district databases), Federated Learning Layer (local training and secure aggregation), Blockchain Layer (Hyperledger Fabric with PBFT consensus), and Application Layer (Django backend and React.js dashboard).

3.3.2 Blockchain Integration Design

The blockchain integration will record:

- Hash of each model update submitted by participating nodes
- Aggregation operations performed by the central server
- Verification proofs that allow participants to confirm their contributions were included

Hyperledger Fabric is selected because it uses practical Byzantine fault tolerance (PBFT) consensus, which is more energy-efficient than proof-of-work and suitable for permissioned networks where participants are known.

3.4 Prototype Implementation

The implementation focuses on a lightweight prototype that works efficiently in resource-limited environments.

3.4.1 Technology Stack

- **Federated Learning Framework:** PySyft will be used for implementing federated learning. PySyft supports privacy-preserving operations including secure aggregation and differential privacy, and integrates well with PyTorch for deep learning models.
- **Blockchain Platform:** Hyperledger Fabric will provide the blockchain layer. A minimal network configuration will be designed with peers representing participating districts and an ordering service for transaction sequencing.
- **Machine Learning Model:** A lightweight neural network will be trained for FMD outbreak prediction. The model architecture will be chosen to balance accuracy with computational requirements suitable for resource-limited devices.
- **Backend Services:** Django will handle API endpoints, user authentication, and coordination between federated learning and blockchain components.
- **Frontend Dashboard:** React.js will be used for the early warning interface, displaying predictions, confidence levels, and explanations for veterinary officers.

3.4.2 Federated Training Process

The federated training process follows these steps:

1. The central server initializes a global model and distributes it to all participating district nodes.
2. Each district node trains the model locally on its FMD data for several epochs.
3. Local model updates are computed and hashed before transmission.
4. Updates are sent to the server along with blockchain transactions recording the submission.
5. The server aggregates updates using federated averaging and records the aggregation on the blockchain.
6. The updated global model is distributed to participants, and the process repeats.

For aggregation, if there are N clients and client i sends update θ_i with local dataset size n_i, standard federated averaging computes:

$$\theta_{global} = \frac{\sum_{i=1}^{N} n_i \theta_i}{\sum_{i=1}^{N} n_i} \tag{1}$$

3.4.3 Privacy Mechanisms

Differential privacy will be implemented to protect individual farm records. The key equation is:

$$\Pr[\mathcal{M}(D) \in S] \leq e^{\epsilon} \cdot \Pr[\mathcal{M}(D') \in S] + \delta \tag{2}$$

where $\mathcal{M}$ is the mechanism, D and D' are neighboring datasets differing by one record, ϵ controls privacy loss, and δ is a small failure probability. The privacy budget will be calibrated to balance protection with model utility.

3.5 System Evaluation

The system will be evaluated across prediction accuracy, computational efficiency, and security.

3.5.1 Accuracy Evaluation

Model performance will be measured using standard classification metrics:

Precision and Recall:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

F1-Score:

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

The federated model's performance will be compared against a centralized baseline trained on the same data collected in one location. Target: F1-score above 0.80, AUC-ROC above 0.85.

3.5.2 Efficiency Evaluation

Communication Overhead: The ratio of data transmitted in federated learning compared to centralized training:

$$\text{Overhead Ratio} = \frac{\text{Bytes in FL}}{\text{Bytes in Centralized}} \tag{5}$$

Target: Ratio less than 0.1 (90% bandwidth reduction).

Training Time: Time per training round will be measured. Target: Under 30 minutes per round on typical hardware available in Ugandan districts.

Resource Usage: Memory consumption and CPU utilization will be monitored to ensure the system runs on resource-limited devices.

3.5.3 Security Evaluation

Model Poisoning Resistance: The system will be tested with varying fractions of malicious participants (10%, 20%, 30%, 40%) sending corrupted updates. Robustness score:

$$\text{Robustness Score} = \frac{\text{Accuracy with attack}}{\text{Accuracy without attack}} \tag{6}$$

Target: Maintain score above 0.80 even with 40% attackers.

Privacy Testing: Membership inference attacks will be conducted to test whether attackers can determine if specific farm data was included in training.

Blockchain Integrity: Verification that all model updates are correctly recorded and that tampering attempts are detected.

3.6 Practical Effectiveness Assessment

The system's practical effectiveness will be assessed through testing with real FMD data.

3.6.1 Simulation of District Nodes

Up to 50 simulated nodes will be created, each representing a Ugandan district. FMD data will be distributed across these nodes according to actual geographic distribution of outbreaks. This setup mimics real-world deployment where each district veterinary office would operate a node.

3.6.2 Realistic Conditions Testing

Testing will be conducted under conditions that reflect the actual operating environment:

- **Network Constraints:** Simulated slow and intermittent internet connectivity typical of rural Uganda
- **Node Failures:** Random node dropouts during training rounds to test system resilience
- **Heterogeneous Data:** Uneven distribution of FMD cases across districts reflecting real patterns

3.6.3 Distribution Shift Detection

The system will implement mechanisms for detecting when data distributions change over time. For each district d at time t, validation loss $L_d^{(t)}$ is tracked over a sliding window.

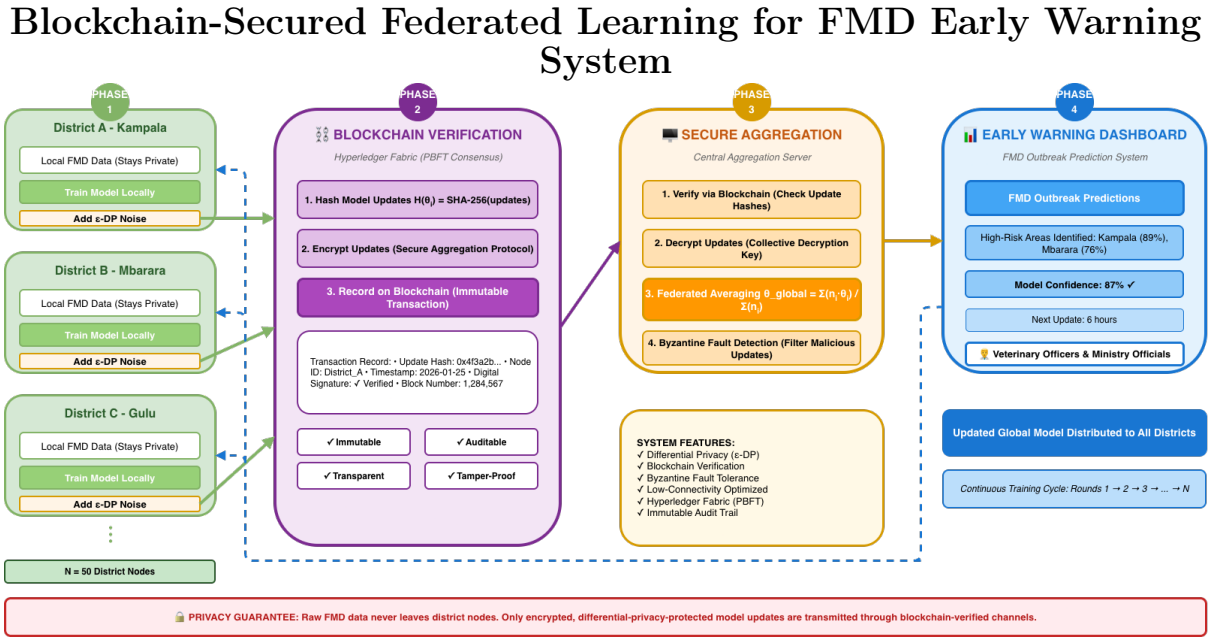


Figure 2: The diagram illustrates the four-phase workflow: (1) Local Training at district nodes with differential privacy, (2) Blockchain Verification using Hyperledger Fabric PBFT consensus, (3) Secure Aggregation with Byzantine fault detection, and (4) Early Warning Dashboard for veterinary officers. Raw FMD data never leaves district nodes, ensuring privacy while blockchain verification maintains trust and auditability.

If the moving average exceeds a threshold:

$$\frac{1}{w}\sum_{k=t-w}^{t} L_d^{(k)} > \mu + 2\sigma \tag{7}$$

where μ and σ are historical mean and standard deviation, this signals distribution drift requiring model retraining.

3.6.4 Early Warning Dashboard Testing

The dashboard interface will be evaluated for usability by presenting predictions to potential end users (veterinary officers or agricultural extension workers) and collecting feedback on clarity, usefulness, and actionability of the information presented.

3.7 System Workflow

The workflow in Fig. 2 illustrates the complete blockchain-secured federated learning pipeline. In Phase 1, each of the 50 district nodes trains ML models locally on their FMD data, applying differential privacy (ε-DP) noise to model updates before transmission. Phase 2 utilizes Hyperledger Fabric's PBFT consensus mechanism to hash, encrypt, and record all model updates on an immutable blockchain ledger with full auditability.

Phase 3 performs secure aggregation using federated averaging ($\theta_{global} = \sum(n_i \cdot \theta_i) / \sum(n_i)$) with Byzantine fault detection to filter malicious updates. Finally, Phase 4 delivers high-confidence (87%) FMD outbreak predictions to veterinary officers and ministry officials through a continuously updated early warning dashboard. This privacy-preserving architecture ensures that sensitive livestock data remains at source while maintaining verifiable trust throughout the entire training cycle.

3.8 Blockchain Transaction Flow

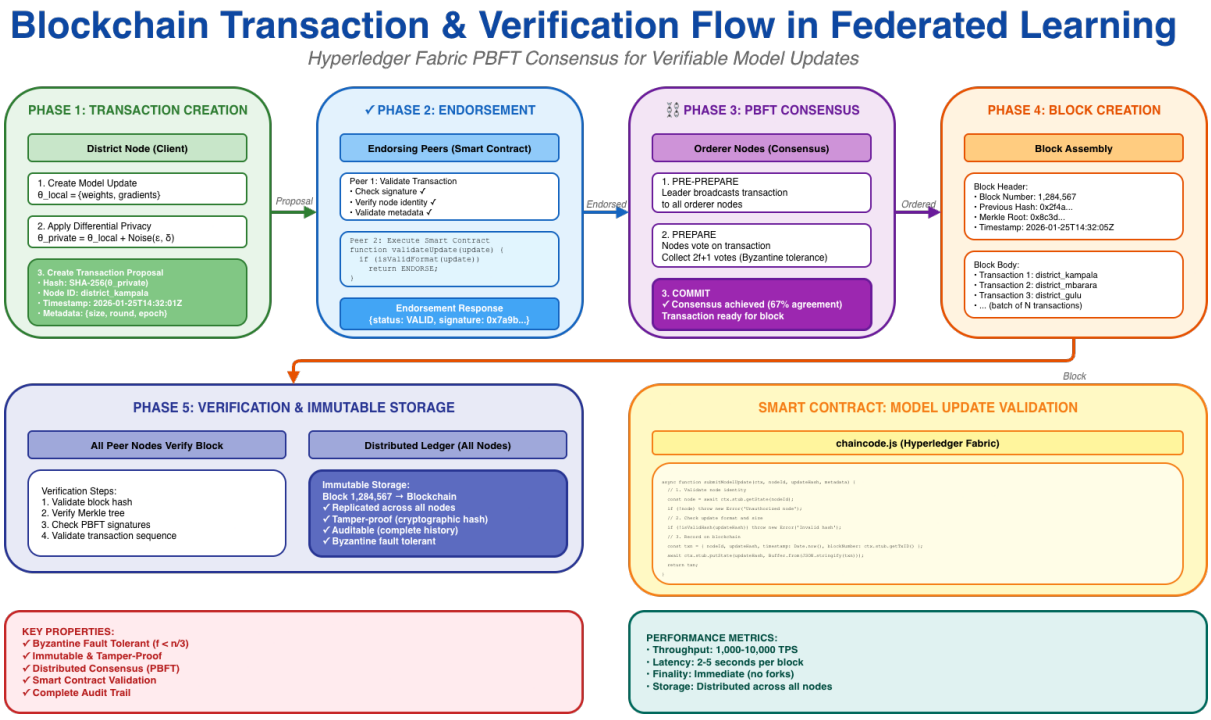


Figure 3: The five-phase blockchain transaction flow: (1) Transaction Creation with differential privacy at district nodes, (2) Endorsement via smart contract execution by endorsing peers, (3) PBFT Consensus with Pre-Prepare, Prepare, and Commit steps achieving 67% agreement, (4) Block Creation with header and body assembly including Merkle root, and (5) Verification & Immutable Storage across distributed ledger. The system achieves 1,000-10,000 transactions per second with 2-5 second latency while maintaining Byzantine fault tolerance where $f < n/3$.

Fig. 3 details the blockchain transaction and verification mechanism that ensures trust and auditability in the federated learning process. When a district node completes local training, it creates a transaction proposal containing the SHA-256 hash of the differentially-private model update along with node identity, timestamp, and metadata. This proposal is submitted to endorsing peers who execute the smart contract to validate the transaction format, verify the node's signature, and check metadata integrity. Upon successful endorsement, the transaction enters the PBFT consensus phase where orderer nodes execute a three-step protocol: Pre-Prepare (leader broadcasts), Pre-

pare (nodes vote with 2f+1 votes required for Byzantine tolerance), and Commit (67% agreement threshold). Once consensus is achieved, transactions are batched into blocks containing a header (block number, previous hash, Merkle root, timestamp) and body (N transactions). All peer nodes verify the block by validating the hash, Merkle tree, PBFT signatures, and transaction sequence before committing to their local copy of the distributed ledger. This architecture provides complete auditability with an immutable record of all model updates, cryptographic proof of integrity, and Byzantine fault tolerance supporting up to f malicious nodes where $f < n/3$. The system maintains high performance with throughput of 1,000-10,000 transactions per second and finality within 2-5 seconds, making it suitable for real-time federated learning deployments in resource-constrained agricultural settings.

3.9 Research Timeline

The research will span 12 months. Table 2 summarizes the key activities for each phase.

Table 2: Research Timeline Summary

Month	**Phase**	**Key Activities**
1-2	Literature Review & Data Preparation	Complete literature review; obtain FMD data from ULITS; clean and preprocess data
3-5	Core System Implementation	Implement federated learning infrastructure; integrate blockchain verification layer
6-7	Security & Privacy	Implement differential privacy; add attack resistance mechanisms
8	Adaptive Mechanisms	Implement distribution shift detection; add model retraining triggers
9	Testing & Dashboard	Test under realistic conditions; develop early warning dashboard
10-11	Comprehensive Evaluation	Conduct experiments on accuracy, efficiency, and security; analyze results
12	Documentation	Write thesis; prepare publications; release open-source code

3.10 Expected Contributions and Deliverables

3.10.1 Scientific Contributions

This research provides a practical solution to the trust problem in federated learning by integrating blockchain verification in resource-limited settings. The system architecture,

implementation strategies, and benchmark performance data will serve as a foundation for future work in privacy-preserving distributed machine learning.

The research demonstrates how to balance security, privacy, and computational efficiency in environments with limited resources and unreliable connectivity.

3.10.2 Technical Contributions

The open-source implementation will include production-ready code for blockchain-verified federated learning, complete documentation, and deployment guidelines. Researchers and practitioners can adapt this system to their own applications without building from scratch.

All components will be modular to allow customization for different use cases, datasets, and infrastructure constraints.

3.10.3 Practical Impact for Uganda

The system provides a working early warning mechanism for Foot-and-Mouth Disease that could help veterinary officers respond quickly to outbreaks. Beyond the specific application, it demonstrates that Uganda can benefit from advanced AI without centralizing sensitive data or relying on external infrastructure.

This establishes a model for how developing countries can maintain data sovereignty while leveraging collaborative machine learning.

3.10.4 Global Relevance

The challenges addressed—intermittent connectivity, limited computational resources, privacy requirements, and trust verification—are common across developing regions. Success in the Ugandan context provides evidence that sophisticated machine learning systems can work in similar conditions worldwide.

The approach applies beyond agriculture to healthcare, disaster response, and other domains where distributed data and limited resources create barriers to AI adoption.

3.11 Evaluation Metrics

3.11.1 Model Performance Metrics

Precision and Recall: For binary classification:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{8}$$

F1-Score: The harmonic mean:

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{9}$$

Target: F1-score above 0.80, AUC-ROC above 0.85.

3.11.2 Privacy Metrics

Differential Privacy Budget: Track cumulative ϵ over training rounds, targeting $\epsilon < 10$.

Membership Inference Attack Success Rate: Measure if attackers can identify whether specific farm data was in training set.

3.11.3 Fairness Metrics

Gini Coefficient: Measures inequality in district selection frequency. Target below 0.3.

Performance Equity: Variance in F1-scores across districts. Lower variance means more equitable performance.

3.11.4 Efficiency Metrics

Communication Overhead Ratio:

$$\text{Overhead Ratio} = \frac{\text{Bytes in FL}}{\text{Bytes in Centralized}} \tag{10}$$

Expect ratio less than 0.1 (90% bandwidth reduction).

Training Time: Target under 30 minutes per round on realistic hardware.

3.11.5 Robustness Metrics

Accuracy Under Attack: For p fraction of malicious clients:

$$\text{Robustness Score} = \frac{\text{Accuracy with attack}}{\text{Accuracy without attack}} \tag{11}$$

Aim to maintain above 0.80 even with 40% attackers.

4 Ethical Considerations

4.1 Data Privacy and Protection

All personally identifiable information will be removed from datasets before use. The federated learning approach ensures that raw farm-level data never leaves its source lo-

cation. Limitations of privacy guarantees will be clearly documented and communicated to participants.

4.2 Informed Consent and Participation

If human subjects are involved, university ethics board approval will be obtained before data collection. Participants will receive clear explanations of how the system works, what data is used, potential risks, and their right to withdraw at any time without penalty.

4.3 Responsible Use of Predictions

The system provides decision-support tools, not autonomous decisions. Veterinary officers maintain authority and responsibility for final judgments. Predictions include explanations to help users evaluate reliability and make informed decisions.

4.4 Equitable Access

All software and research outputs will be released under open-source licenses to ensure broad access. The system is designed to work with intermittent connectivity so that all districts, regardless of infrastructure quality, can benefit equally.

4.5 Environmental Impact

Hyperledger Fabric uses practical Byzantine fault tolerance rather than energy-intensive proof-of-work consensus. Federated learning reduces network traffic and data center energy consumption by processing data locally.

5 Research Scope

This research focuses on designing and evaluating a blockchain-verified federated learning system for FMD prediction in Uganda. The scope includes:

- **Geographic Coverage:** Ugandan districts participating in the Uganda Livestock Information Tracking System (ULITS)
- **Disease Application:** Foot-and-Mouth Disease early warning as the primary test case
- **Data Timeline:** Historical FMD outbreak data from 2011-2024 for training and evaluation
- **Technical Components:** Federated learning with blockchain verification, differential privacy, secure aggregation, and adaptation to distribution shifts

5.1 Acknowledged Limitations

Production Deployment: Full operational deployment across Uganda requires additional resources, partnerships, and infrastructure beyond this research scope.

Security Testing: Evaluation covers common Byzantine attack scenarios but cannot exhaustively test all possible threat vectors.

Trust Model: The system assumes an honest majority (less than 50% malicious participants), which is a fundamental requirement for Byzantine fault-tolerant consensus.

Scale: Testing involves up to 50 simulated nodes representing districts. Performance at larger scales may differ.

5.2 Future Research Directions

This foundation enables future work on multi-disease monitoring, cross-country collaboration, real-time sensor integration, personalized district-level models, and economic incentive mechanisms for sustained participation.

6 Conclusion

This research addresses the trust problem in federated learning by integrating blockchain verification, specifically designed for resource-limited environments. The system enables collaborative machine learning without centralizing data, making it practical for settings where privacy concerns, limited connectivity, and infrastructure constraints prevent traditional approaches.

Using Foot-and-Mouth Disease prediction in Ugandan cattle farms as a test case provides real-world validation under challenging conditions. The FMD application addresses a genuine need for early warning in the livestock sector while demonstrating that blockchain-verified federated learning can work despite intermittent internet, limited computing resources, and distributed data sources.

The research contributes a working system architecture, implementation strategies, and empirical evidence for privacy-preserving collaborative machine learning in developing regions. Beyond the specific application to FMD in Uganda, the approach applies to healthcare, agriculture, and other domains where data remains distributed due to privacy, sovereignty, or infrastructure limitations.

The outcome includes open-source tools, performance benchmarks, and practical lessons that advance the field of privacy-preserving machine learning while addressing a real problem in Uganda's livestock sector.

References

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in *Proc. 20th Int. Conf. Artificial Intelligence and Statistics (AISTATS)*, 2017.

[2] P. Kairouz *et al.*, "Advances and Open Problems in Federated Learning," *Foundations and Trends in Machine Learning*, vol. 14, no. 1–2, pp. 1–210, 2021.

[3] K. Bonawitz *et al.*, "Practical Secure Aggregation for Privacy-Preserving Machine Learning," in *Proc. ACM SIGSAC Conf. Computer and Communications Security*, 2017.

[4] Y. Liu, X. Qu, and G. Chen, "Blockchain-Based Practical and Privacy-Preserving Federated Learning with Verifiable Fairness," *Mathematics*, vol. 11, no. 5, p. 1091, 2023.

[5] R. Chen, Y. Li, and M. Zhang, "FLock: Robust and Privacy-Preserving Federated Learning based on Practical Blockchain State Channels," *Cryptology ePrint Archive*, Paper 2024/1797, 2024.

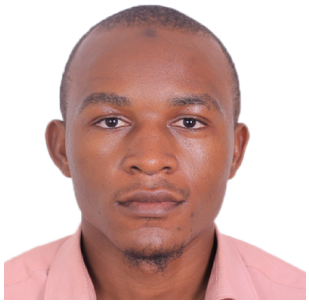
[6] R. S. Antunes, C. A. da Costa, and A. Küdde, "Federated Machine Learning in Healthcare: A Systematic Review," *ACM Computing Surveys*, vol. 55, no. 5, pp. 1–35, 2022.

[7] T. Kapalaga, M. Mubangizi, and P. Kisaakye, "A Unified Foot and Mouth Disease Dataset for Uganda: Evaluating Machine Learning Predictive Performance Degradation Under Varying Distributions," *Frontiers in Artificial Intelligence*, vol. 7, 2024.

[8] C. Dwork and A. Roth, "The Algorithmic Foundations of Differential Privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014.

[9] Hyperledger Foundation, "Hyperledger Fabric," [Online]. Available: https://www.hyperledger.org/use/fabric

About the Author



Muhindo Mubaraka received his B.Sc. in Computer Science from the Islamic University of Technology, Bangladesh, in 2022. He is currently pursuing a Master of Science in Computer Science at Makerere University. He works as a software engineer and his research interests include federated learning and blockchain technology.