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

Research Proposal

Muhindo Mubaraka Master of Computer Science Candidate
Makerere University

Student Number: 2400725633
Registration Number: 2024/HD05/25633U

muhindo.mubaraka@students.mak.ac.ug

Supervisor: Dr. Chongomweru Halimu Department of Computer Science
Makerere University

***Abstract*—Training effective machine learning models today requires bringing all data together in one place. This approach works well when you have fast internet and powerful computers, but it creates serious problems. Farmers and hospitals cannot share their data because of privacy concerns. In rural areas with limited connectivity, uploading large datasets is simply not practical. Even when possible, centralizing sensitive information creates security risks that many organizations cannot accept.**

This research explores a different approach. Instead of moving data to where the model is being trained, we send the model to learn from data that stays in its original location. This is called federated learning. However, when you train models across many separate locations, a new challenge emerges: how do you trust that everyone is participating honestly? How do you verify that the system has not been compromised?

I propose using blockchain technology to solve this trust problem. The blockchain acts as an unchangeable record keeper, documenting every step of the training process and ensuring that no one can cheat the system. To demonstrate that this works in practice, I will apply it to predicting Foot-and-Mouth Disease outbreaks in Ugandan cattle farms, where data privacy and limited infrastructure make traditional approaches impossible.

The outcome will be a working system that shows how organizations can collaborate on building intelligent models without sacrificing data ownership or privacy. This has implications far beyond disease prediction, offering a blueprint for privacy-preserving AI in healthcare, agriculture, and other sensitive domains where data sharing remains a major barrier.

***Index Terms*—Federated Learning, Blockchain, Privacy-Preserving Machine Learning, Disease Prediction, Distributed Systems, Differential Privacy, Byzantine Fault Tolerance**

I. Introduction

Currently, building a good machine learning model involves gathering all data in one central location and using that data to train the model. It is like making it mandatory for any student to be able to learn; they must first physically go to the library. This causes major issues for privacy, it is costly, and in areas with slow internet, it's often simply not feasible.

My research explores a different path. Instead of gathering the data, what if we could send the AI model out to learn from where the data already lives? This is what Federated Learning promises. But this new approach brings its own big question: if no one is sending their data to a central authority, how can we trust that the system is working correctly and hasn't been tampered with?

This is where I propose bringing in blockchain, not for cryptocurrencies, but as a trust mechanism. My research is to design, build, and test a system that combines these two technologies to create a smarter, more private, and trustworthy way of training machine learning models.

To prove this works, I will apply it to a real-world problem: predicting Foot-and-Mouth Disease in Uganda's cattle farms. The heart of my work is a contribution to computer science, creating a practical blueprint for decentralized AI that respects privacy and builds trust directly into the system.

II. Background

The classical way of applying machine learning relies on taking data from different sources, unifying them in one place, and then using that combined data for the training of the model. This method encounters serious problems like the risk of exposing personal information, the need for very powerful and expensive computers, and the inconveniences in places with poor internet connectivity.

Federated learning, on the other hand, suggests an amazing new way of working where the models can be trained without transferring data and just using the local data kept in different places. This procedure allows a simultaneous learning process. But, this method of decentralization has also caused concerns about the integrity of the system and the possibility of verifying the contributions of the different participants.

In the absence of a central authority that is responsible for the inspection of the data, what can we do to ensure that the whole learning process is protected and that no one is sending ill-intentioned updates to harm the model?

This is the point where blockchain technology comes in with a very strong solution. Besides its application in the field of cryptocurrencies, it has been proven that the core of blockchain technology is the creation of trust in an environment where there is no trust. It can be represented as a decentralized ledger that can document the entire learning process step by step.

Improper claims and fraud will be practically impossible in the case of blockchain. By employing blockchain technology, a new system can be introduced in which every model update goes through a process of verification, the whole thing is open, and the outcomes—like an early disease warning—are credible and cannot be altered after they have been announced.

My research is situated at the intersection of these two technologies. I will investigate the way these technologies can be combined into a single, integrated system that is not only private and efficient but also capable of gaining users' trust.

III. Literature Review

A. *Kapalaga et al. (2024) – A Unified FMD Dataset*

The dataset presents a national FMD dataset that is not only a collection of various data sourced from different regions and sources but also a major step towards overcoming the problem of fragmentation.

Moreover, they indicate that the performance of standard ML models suffers heavily when the distribution of important input variables (like rainfall, temperature) changes over time, thus arguing the necessity of models that are capable of dealing with non-stationary data.

Gap: The research is solely centralized, thus all data is obtained and managed in one location. The study did not consider federated learning (FL) and decentralized data processing which could help mitigate concerns regarding data privacy and network inefficiencies, particularly in low-resource areas with inadequate connectivity.

B. *Zhang et al. (2023) – Blockchain-Based FL*

The authors propose a blockchain-based system for federated learning, which aims at the preservation of privacy as well as the fairness of all parties involved. The setup combines blockchain with crypto-based methods such as verifiable random functions (VRFs) and zero-knowledge proofs (ZKPs) to provide fairness and confidentiality during the training of the model.

Gap: Despite the fact that privacy and fairness are mentioned as main advantages, the paper still does not consider the application domains of early warning systems, which are real-world situations, nor does it offer an end-to-end architecture for implementation in actual, low-resource areas such as agriculture or disease prediction.

C. *Teo et al. (2024) – FL in Healthcare*

The paper presents a meticulous examination of the role of federated learning in healthcare through the lens of 612 studies. It points out the main issues and challenges encountered by FL in healthcare, including but not limited to interoperability, legislative barriers, and lack of actual application.

Gap: Despite being very informative about FL in healthcare, the paper does not elaborate on an integrated system that uses both blockchain and FL for decentralized early warning systems.

D. *Chen et al. (2024) – FLock System*

The system demonstrates the creation of a strong privacy-preserving scenario in federated learning setup using blockchain state channels, resistant to model poisoning.

Gap: The emphasis is on secure aggregation and robustness, rather than on creating a complete early warning system with real-time estimations and domain-specific functionality.

IV. Research Gaps

I spotted three major areas where current research is lacking through my literature review, which my research proposes to cover:

1) **The Application Gap:** The frameworks are showcasing the implementation of blockchain in federated learning for general purposes; however, there is no dedicated system for livestock disease prediction. The present technologies are either theoretical or healthcare-centric, thereby creating an urgent demand for a specialized system suited to the agriculture sector for early warning.
2) **The Trust-Verification Gap:** The systematic reviews keep mentioning the privacy advantages of current federated learning research, but they do not come along with the built-in means for the verification of model integrity and the participants' compliance with the protocols.
3) **The Practical Efficiency Gap:** The majority of the proposed frameworks are computationally intensive and not designed to meet the real-world conditions in developing areas. There is no design for a system that is lightweight and efficient at the same time that could operate despite network limitations and resource constraints.

V. Problem Statement

The merging of these gaps discloses the principal problem that my study is going to tackle: Federated learning allows for the development of a system that collaborates and preserves privacy at the same time, but it still relies on complete trust in the participants and the central aggregators.

The inclusion of blockchain as a trust layer usually results in impractical, slow systems that are counterproductive to efficient decentralized learning.

As a result, the main research question is to come up with and test a new system architecture that will allow for the smooth integration of federated learning with blockchain, thus establishing a private, trustworthy, and efficient framework for collaborative prediction—particularly tailored for the real-world constraints and applied to crucial areas like disease early warning.

VI. Research Objectives

A. Main Objective

To develop a private and trustworthy collaborative prediction system, that is blockchain-based and fed with a federated learning system, which will be suitable for resource-constrained environments after thorough evaluation.

B. Specific Objectives

- To create a system architecture that optimally combines federated learning with blockchain technology for safe and auditable model merging.
- To put the whole system concept into a working prototype using lightweight frameworks appropriate for settings with limited computing resources.
- To carry out a performance assessment of the system regarding prediction accuracy, computational efficiency, and resilience to malevolent attacks.
- To further prove the system's applicability through a case study on Foot-and-Mouth Disease prediction using Ugandan cattle farm data.

VII. Research Methodology

A. System Architecture Design

We will design a layered architecture that strategically separates concerns while maintaining integration between federated learning and blockchain components:

- **Data Layer:** Local data remains on farm systems with standardized interfaces for secure access
- **Federated Learning Layer:** Model training occurs locally with secure aggregation protocols
- **Blockchain Layer:** Hyperledger Fabric will provide the trust foundation for recording model hashes and verification
- **Application Layer:** Early warning dashboard and alert system

B. Prototype Implementation

The theoretical design will be made real through working prototype development using technologies that are robust and suitable for the practical world:

- **PySyft** for federated learning module as it is designed for privacy-preserving artificial intelligence
- **Hyperledger Fabric** as blockchain platform for quick, private transactions without excessive computing power
- **Django** for backend management and **React.js** for web-based dashboard

C. Performance Evaluation

We will conduct comprehensive testing across three key dimensions:

1) **Model Accuracy:** Compare prediction performance using F1-score, precision, and recall metrics
2) **System Efficiency:** Measure communication overhead, training convergence time, and computational resource usage
3) **Security Analysis:** Test robustness against model poisoning attacks and privacy preservation

VIII. Timeline and Work Plan

My research will span roughly two years, with each phase building naturally on what came before.

A. Phase 1: Foundation (Months 1-4)

The first few months are about getting my bearings and laying solid groundwork. I will spend time deeply understanding what others have done, making sure I am not reinventing wheels, and getting access to the data I need.

By the end of month four, I should have a clean, unified dataset sitting on my computer, ready to be split up and used for training.

B. Phase 2: Building the Core (Months 5-9)

This is where things get real. I will start coding, first getting a basic federated learning system working without any of the fancy blockchain stuff. Once that is solid, I will layer in the blockchain.

It is tempting to try to build everything at once, but I have learned that leads to impossible-to-debug messes. Better to get one piece working perfectly before adding the next layer.

C. Phase 3: Making It Secure (Months 10-13)

Now I will make the system genuinely privacy-preserving and resistant to attacks. I will implement differential privacy, carefully calibrating how much noise to add.

For differential privacy, the key equation is:

$$\Pr[\mathcal{M}(D) \in S] \leq e^{\epsilon} \cdot \Pr[\mathcal{M}(D') \in S] + \delta \tag{1}$$

where $\mathcal{M}$ is our mechanism, D and D' are neighboring datasets differing by one record, ϵ controls privacy loss, and δ is a small failure probability.

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

$$\theta_{global} = \frac{\sum_{i=1}^{N} n_i \theta_i}{\sum_{i=1}^{N} n_i} \tag{2}$$

D. Phase 4: Making It Adaptive (Months 14-16)

For each district d at time t, I will track the validation loss $L_d^{(t)}$ over a sliding window. If the moving average suddenly increases beyond a threshold:

$$\frac{1}{w} \sum_{k=t-w}^{t} L_d^{(k)} > \mu + 2\sigma \tag{3}$$

where μ and σ are historical mean and standard deviation, that signals drift.

E. Phase 5: Making It Real (Months 17-20)

I will test whether the system actually works under realistic conditions—slow internet, machines that crash, participants who drop offline mid-training. I will also build the dashboard where veterinary officers can see predictions and explanations.

F. Phase 6: Testing Everything (Months 21-23)

Now I run every experiment I can think of. How accurate is the system compared to centralized training? What happens if 30% of participants try to sabotage the model?

G. Phase 7: Writing and Wrapping Up (Month 24)

The final stretch is writing the thesis, preparing papers for publication, and releasing the code and datasets publicly.

IX. Expected Contributions

A. Scientific Contributions

I will be creating the first complete blueprint for combining federated learning with blockchain verification in a way that actually handles real-world messiness. The architecture I design, the lessons I learn, and the benchmark performance numbers I establish—all of that becomes a foundation others can build on.

B. Technical Contributions

The open-source code I release will be production-ready. A researcher in Kenya can download my code, plug in their own data, and have a working system in days instead of months.

C. Practical Impact for Uganda

There is a direct outcome: a working early warning system that could help prevent livestock disease outbreaks. But the bigger impact is demonstrating that Uganda does not have to ship its data abroad to benefit from AI.

D. Global Relevance

The challenges I am tackling—intermittent internet, limited resources, privacy regulations, the need for trust—these are not unique to Uganda. If I can prove that sophisticated AI works in these conditions, that removes a massive barrier for most of the world.

X. Evaluation Metrics

A. Model Performance Metrics

Precision and Recall: For binary classification:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

F1-Score: The harmonic mean:

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

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

B. 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.

C. 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.

D. Efficiency Metrics

Communication Overhead Ratio:

$$\text{Overhead Ratio} = \frac{\text{Bytes in FL}}{\text{Bytes in Centralized}} \tag{6}$$

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

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

E. Robustness Metrics

Accuracy Under Attack: For p fraction of malicious clients:

$$\text{Robustness Score} = \frac{\text{Accuracy with attack}}{\text{Accuracy without attack}} \tag{7}$$

Aim to maintain above 0.80 even with 40% attackers.

XI. Ethical Considerations

A. Protecting Privacy

I will be transparent about limitations. All personally identifiable information will be stripped from datasets before analysis.

B. Informed Consent

Every participant will receive clear explanations of what the system does, what risks exist, and their right to withdraw. University ethics board approval will be obtained before human-subjects research.

C. Avoiding Harm

The system will generate predictions as decision-support tools, not autonomous decisions. Veterinary officers retain final judgment. Built-in explanations help users evaluate predictions.

D. Fair Access to Benefits

All software and research outputs will be released under open-source licenses. The system must work even with intermittent connectivity to serve all districts equitably.

E. Environmental Responsibility

Hyperledger Fabric uses practical Byzantine fault tolerance, not energy-intensive proof-of-work. Federated learning reduces data center energy consumption by keeping data local.

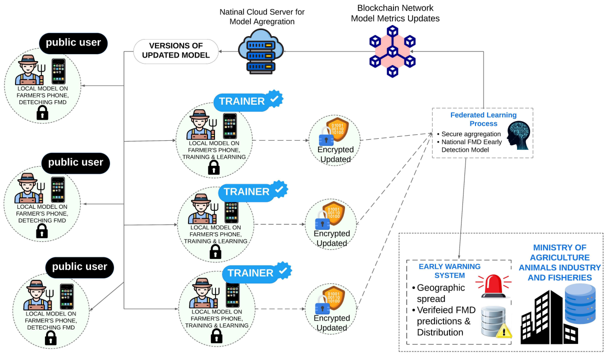

Fig. 1. System workflow showing local data preparation, federated model training, blockchain verification, secure aggregation, and early warning generation. Data never leaves local farms while maintaining verifiable trust through blockchain recording.

XII. Scope and Limitations

A. Research Scope

Geographic: Uganda, specifically districts with ULITS infrastructure.

Disease: Foot-and-mouth disease as primary use case.

Temporal: Historical data 2011-2024; system designed for ongoing operation.

Technical: Integration of federated learning and blockchain with differential privacy, robust aggregation, and adaptation mechanisms.

B. Limitations

Full Production Deployment: Beyond scope; requires additional resources and partnerships.

Attack Coverage: Testing common Byzantine attacks, but cannot cover all possible threats.

Assumption of Honest Majority: System requires fewer than 50% of participants to be malicious—a fundamental limitation of distributed consensus.

Scalability: Testing up to 50 client nodes; larger scale may reveal additional bottlenecks.

C. Future Directions

This research enables extensions to other diseases, multi-country monitoring, real-time sensor integration, advanced personalization, and economic sustainability mechanisms.

XIII. Process Workflow

The workflow in Fig. 1 ensures data privacy while maintaining accountability through blockchain-recorded model updates and aggregation steps.

XIV. Conclusion

This research addresses a fundamental challenge in modern AI: how to build intelligent systems that learn from distributed data without compromising privacy, security, or fairness. By combining federated learning with blockchain verification, we create a trustworthy foundation for collaborative machine learning in high-stakes domains.

The livestock disease use case provides a rigorous testbed with real challenges. Success here demonstrates principles applicable far beyond agriculture. We are pioneering a new paradigm for decentralized AI where data ownership remains with those who generated it, contributions are verifiably fair, systems adapt to changing conditions, and trust is earned through transparency.

The outcome will be open-source tools, empirical evidence, and practical insights that advance privacy-preserving machine learning while delivering tangible benefits to Ugandan livestock keepers.

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

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

[7] 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.

[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

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 developer and his research interests include federated learning and blockchain technology.