IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 6, May 2025
Impact Factor: 7.67

Healthcare Data Security using Federated Learning and Blockchain Technology

Pallavi R A[1], Varshini D[2], V Monisha[3], Y P Lavanya[4], Dr. Nandini S[5]
U.G. Students, Department of Information Science and Engineering[1-4]
Associate Professor, Department of Information Science and Engineering[5]
pallavigowda1555@gmail.com, varshinigowda072@gmail.com
monisha.v953@gmail.com, lavanyayp2@gmail.com, nandinis@sjcit.ac.in
S J C Institute of Technology, Chickaballapura, Karnataka, India

Abstract: *This paper presents a secure, scalable, and privacy-preserving framework for automated skin cancer detection by integrating Federated Learning (FL) and Blockchain technology. A pre-trained ResNet50 convolutional neural network is fine-tuned using the HAM10000 dataset to classify dermoscopic images into seven distinct skin lesion categories. FL enables decentralized model training by allowing multiple healthcare institutions to collaboratively improve model performance without transferring sensitive patient data to a central server. Instead, only model parameters are exchanged, preserving data locality and confidentiality.To ensure transparency and trust among participants, a Blockchain-based infrastructure is employed to immutably log model updates. Smart contracts, deployed on the Ethereum blockchain using Ganache and MetaMask, manage authentication, access control, and validation of contributions from participating nodes. A decentralized ledger stores hashed model updates, ensuring data integrity and resistance to tampering. The system also incorporates cryptographic mechanisms to secure communication channels and maintain model authenticity during transmission.A user-friendly web interface, developed with Flask and HTML, facilitates interaction with the system, while MetaMask integration enables secure, blockchain-backed user verification. The proposed architecture supports interoperability and scalability, making it well-suited for deployment in smart healthcare ecosystems. Experimental results, evaluated using metrics such as classification accuracy, communication latency, and privacy protection, demonstrate the effectiveness of the proposed approach. This work contributes a robust and transparent AI-driven solution tailored for privacy-aware, decentralized digital healthcare environments.*

Keywords: Federated Learning, Blockchain Technology, Privacy-Preserving Healthcare, Skin Cancer Detection,ResNet50,Smart Contracts, Secure Model Aggregation

I. INTRODUCTION

As healthcare increasingly adopts data-driven technologies, Artificial Intelligence (AI) and Machine Learning (ML) are playing transformative roles in improving diagnostic accuracy and patient care. However, these advancements raise critical concerns around data privacy, security, and regulatory compliance, especially with sensitive medical information. Traditional centralized ML systems aggregate data in cloud environments, exposing it to risks such as breaches, unauthorized access, and loss of control. Additionally, they often struggle to manage heterogeneous data across institutions.To overcome these challenges, this work proposes a secure framework combining Federated Learning (FL) and Blockchain technology. FL allows healthcare institutions to collaboratively train models without sharing raw patient data, preserving privacy by transmitting only model updates. While this approach enhances confidentiality, FL alone cannot prevent malicious updates or unauthorized participation.To reinforce security and trust, the framework integrates a Blockchain layer using Ethereum, Ganache, and MetaMask. Smart contracts are employed for access control, contribution validation, and immutable logging of model updates. This decentralized ledger ensures tamper-proof collaboration and transparency among participants.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 6, May 2025
Impact Factor: 7.67

II. PROBLEM STATEMENT

The rise of AI in healthcare has heightened concerns about patient data privacy and security. Traditional machine learning requires centralized data storage, exposing sensitive medical records to breaches, unauthorized access, and regulatory challenges. This is especially critical in applications like skin cancer detection, where data sensitivity and diagnostic accuracy are both essential.Centralized systems also reduce transparency and data ownership, and they present single points of failure. Existing methods often fail to balance privacy with performance or support secure collaboration among untrusted parties.To address these issues, this project proposes a unified framework that combines Federated Learning and Blockchain. The goal is to enable secure, transparent, and privacy-preserving collaborative training of skin cancer detection models—without sharing or centralizing sensitive healthcare data.

III. LITERATURE REVIEW

• Kairouz, P., McMahan, H. B., et al. (2021)

This paper titled "Advances and Open Problems in Federated Learning" provides a comprehensive survey of the FL landscape. It discusses the principles, architecture, and algorithms involved in FL, and identifies open challenges that are yet to be addressed. It serves as a key theoretical base for this project by outlining the scope and future direction of FL.

• Li, T., Sahu, A. K., et al. (2020)

The paper "Federated Optimization in Heterogeneous Networks" focuses on optimizing FL systems under conditions of data and system heterogeneity. It introduces adaptive federated optimization methods to address issues like non-IID data and variable computation capabilities across devices, which are significant concerns in real-world FL deployments.

• Woisetschlager, H., Erben, A., Wang, S., Mayer, R., & Jacobsen, H.-A. (2024)

The paper "A Survey on Efficient Federated Learning Methods for Foundation Model Training" discusses the challenges and methodologies for applying FL to large-scale foundation models. It emphasizes the need for efficient training techniques and addresses issues related to model scalability and resource constraints. This reference is pertinent for understanding the integration of FL with advanced machine learning models.

• Ye, M., Fang, X., Du, B., Yuen, P. C., & Tao, D. (2023)

In "Heterogeneous Federated Learning: State-of-the-art and Research Challenges", the authors delve into the complexities arising from heterogeneity in FL systems, including statistical, model, and system heterogeneity. They propose a taxonomy for existing methods and discuss future research directions. This work is crucial for projects dealing with diverse data distributions and system capabilities.

• Liu, Y., Yuan, X., Xiong, Z., Kang, J., Wang, X., & Niyato, D. (2020)

The article "Federated Learning for 6G Communications: Challenges, Methods, and Future Directions" explores the role of FL in the context of emerging 6G networks. It discusses the potential applications, challenges, and methodologies for integrating FL into next-generation communication systems. This reference is valuable for understanding the intersection of FL and advanced communication technologies.

• Liu, Y., Kang, Y., Zou, T., Pu, Y., He, Y., Ye, X., Ouyang, Y., Zhang, Y.-Q., & Yang, Q. (2022)

In "Vertical Federated Learning: Concepts, Advances and Challenges", the authors focus on Vertical Federated Learning (VFL), where different parties hold different features of the same data instances. The paper discusses the unique challenges of VFL, including privacy preservation and effective model training. This reference is essential for projects involving collaborative learning across organizations with complementary datasets.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 6, May 2025
Impact Factor: 7.67

IV. DESIGN AND IMPLEMENTATION

Design Details

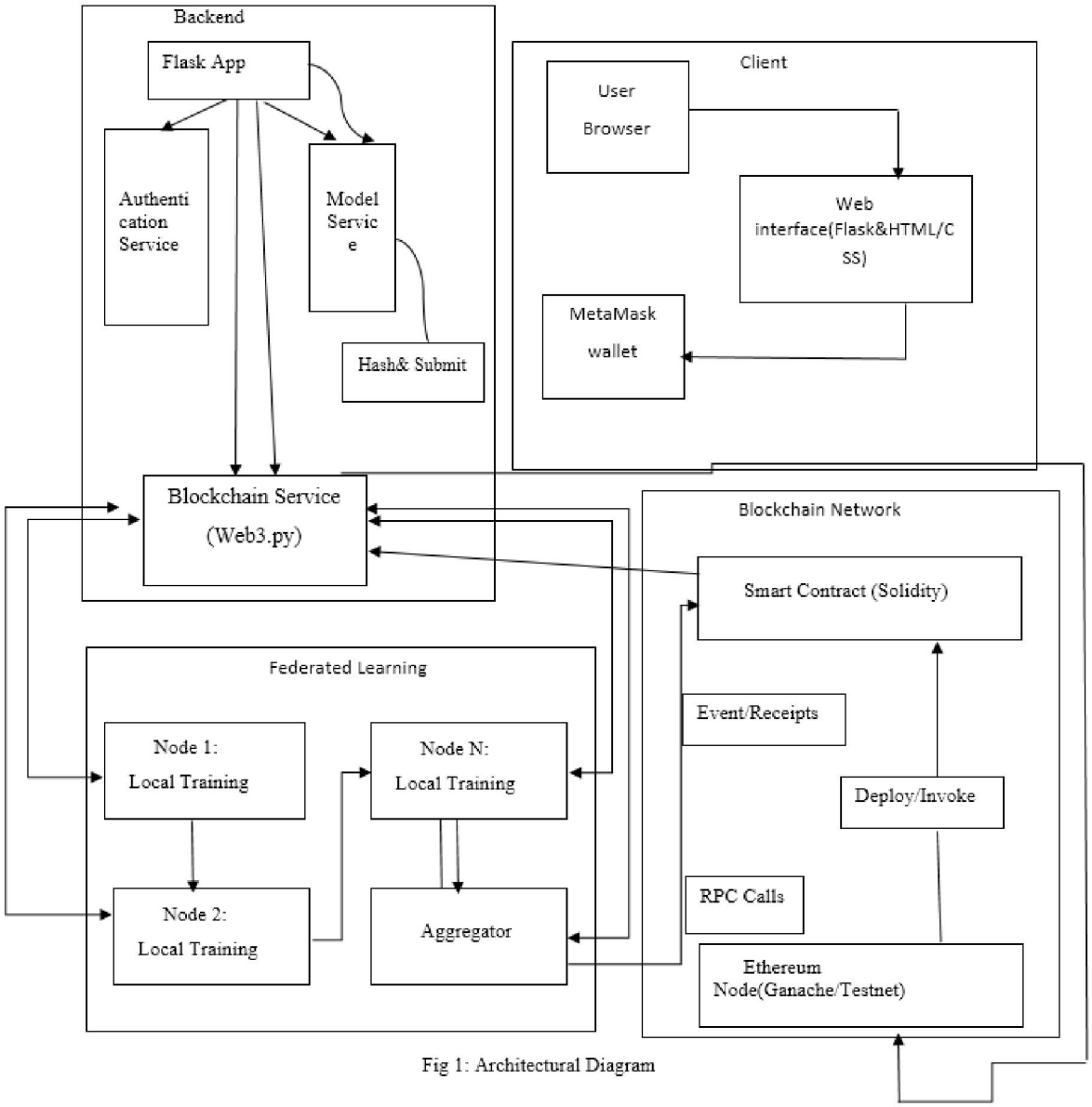

Fig 1: Architectural Diagram

1. Client Layer

• User Browser (A): The practitioner's browser.
• Web Interface (B): Flask-served UI where users log in, view dashboards, and trigger training rounds.
• MetaMask Wallet (C): Provides cryptographic login and signing for blockchain interactions.

2. Backend Layer

• Flask App (D): Central server orchestrating authentication, model logic, and blockchain calls.
• Authentication Service (E): Verifies MetaMask signatures and manages user sessions.
• Model Service (F): Handles global model initialization, distribution, and aggregation.
• Blockchain Service (G): Uses Web3.py to send transactions and listen for events.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 6, May 2025
Impact Factor: 7.67

3. Federated Learning Layer

• Nodes (N1, N2, Nx): Simulated client institutions training locally on private data.
• Aggregator (H): Collects local updates and performs Federated Averaging to update the global model.

4. Blockchain Network Layer

• Ethereum Node (BC): A local Ganache or testnet node that processes transactions.
• Smart Contract (SC): Solidity contract that validates and immutably records model-update hashes.
• Flows:
o RPC Calls from the Backend's Blockchain Service to the Ethereum node.
o Deploy/Invoke actions on the smart contract.
o Hash & Submit by the Model Service for each local update.
o Events/Receipts returned to the Blockchain Service to confirm successful logging.

This architecture ensures that users authenticate first, then interact with the federated - learning pipeline and blockchain in a modular, secure manner.

Implementation Details

This algorithm outlines the interaction between federated learning clients, a centralized Flask-based aggregator, and an Ethereum blockchain to ensure secure, transparent, and privacy-preserving model training.
Input:
• MMM: Pre-trained global model (e.g., ResNet50)
• DiD_iDi: Local dataset at node iii
• NNN: Total number of participating nodes
• Wi\mathcal{W}_iWi: Local model weights
• H()H()H(): Hashing function (e.g., SHA-256)
• SCSCSC: Smart contract for logging updates
• MetaMask\text{MetaMask}MetaMask: Web3 wallet for cryptographic login
• EnodeE_{node}Enode: Ethereum node (Ganache/Testnet)
Output:
• M∗M^*M∗: Converged global model after secure federated learning rounds

Step-by-Step Workflow:

Step 1: User Authentication via MetaMask

1. User accesses the web interface through a browser.
2. Flask frontend prompts MetaMask for a cryptographic signature.
3. Flask backend verifies the signature using Web3.py.

Step 2: Model Initialization

1. The global model MMM is initialized in the ModelService.
2. Model weights are distributed to all NNN client nodes.

Step 3: Local Training (Client-Side)

For each node i∈{1,2,...,N}i \in \{1, 2, ..., N\}i∈{1,2,...,N}:
1. Load model MMM.
2. Train on local dataset DiD_iDi to obtain updated weights Wi\mathcal{W}_iWi.
3. Generate hash: hi=H(Wi)h_i = H(\mathcal{W}_i)hi=H(Wi).

Step 4: Blockchain Logging

1. Submit hih_ihi to the blockchain via the BlockchainService.
2. Interact with SmartContract SCSCSC using RPC to:
o Verify identity (based on MetaMask address).
o Record the hash hih_ihi with timestamp and contributor ID.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 6, May 2025
Impact Factor: 7.67

3. Receive event confirmation or transaction receipt.

Step 5: Model Aggregation

1. After all hih_ihi are verified, aggregator collects corresponding Wi\mathcal{W}_iWi.
2. Apply Federated Averaging (FedAvg):

M←∑i=1N∣Di∣∑j∣Dj∣⋅WiM \leftarrow \sum_{i=1}^N \frac{|D_i|}{\sum_j |D_j|} \cdot \mathcal{W}_iM←i=1∑N∑j∣Dj∣∣Di∣⋅Wi

3. Update global model MMM and redistribute to clients.

Step 6: Iterative Convergence

1. Repeat Steps 3–5 for multiple rounds until convergence criteria are met (e.g., accuracy threshold or max epochs).

Step 7: Final Model Deployment

1. Final model M∗M^*M∗ is served to authorized users through the web interface.
2. Optionally, store the final model hash on-chain for immutability.

Security Guarantees

- Authentication via MetaMask ensures verified user identities.
- Integrity is preserved by hashing and recording updates immutably.
- Privacy is maintained since raw patient data never leaves local nodes.

V. CONCLUSION

This Blockchain-Enhanced Federated Learning framework demonstrates a viable approach for collaborative, decentralized AI in healthcare. By combining local ResNet50 training on private datasets with immutable, smart-contract-backed logging of model updates, the system achieves stronger generalization in skin-lesion classification while preserving data locality. MetaMask-based authentication and on-chain permissions ensure that only verified institutions contribute to—and benefit from—the shared global model. The modular architecture, comprising Flask services, an aggregator, and a blockchain node, supports seamless scaling as new participants join. Real-time dashboards provide transparency into training performance and transaction latency, and built-in safeguards protect against tampering or replay attacks. Together, these features offer a robust, extensible foundation for privacy-aware, multi-party AI collaboration. Future work may include integration with real-world clinical workflows, support for additional disease modalities, and exploration of advanced aggregation algorithms to further enhance model resilience and fairness.

REFERENCES

[1] P. Kairouz, H. B. McMahan, et al., "Advances and Open Problems in Federated Learning," Foundations and Trends® in Machine Learning, vol. 14, no. 1–2, pp. 1–210, 2021.
[2] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated Optimization in Heterogeneous Networks," in Proceedings of Machine Learning and Systems (MLSys), vol. 2, pp. 429–450, 2020.
[3] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated Machine Learning: Concept and Applications," ACM Transactions on Intelligent Systems and Technology (TIST), vol. 10, no. 2, pp. 1–19, 2019.
[4] K. Bonawitz, H. Eichner, et al., "Towards Federated Learning at Scale: System Design," in Proc. 2nd SysML Conf., Stanford, CA, USA, 2019.
[5] A. Hard, K. Rao, R. Mathews, S. Ramaswamy, F. Beaufays, S. Augenstein, H. Eichner, C. Kiddon, and D. Ramage, "Federated Learning for Mobile Keyboard Prediction," arXiv preprint arXiv:1811.03604, 2018.
[6] H. Woisetschlager, A. Erben, S. Wang, R. Mayer, and H.-A. Jacobsen, "A Survey on Efficient Federated Learning Methods for Foundation Model Training," arXiv preprint arXiv:2401.11012, 2024.
[7] M. Ye, X. Fang, B. Du, P. C. Yuen, and D. Tao, "Heterogeneous Federated Learning: State-of-the-art and Research Challenges," Information Fusion, vol. 98, pp. 1–20, 2023.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 6, May 2025
Impact Factor: 7.67

[8] Y. Liu, X. Yuan, Z. Xiong, J. Kang, X. Wang, and D. Niyato, "Federated Learning for 6G Communications: Challenges, Methods, and Future Directions," IEEE Open Journal of the Communications Society, vol. 1, pp. 324–340, 2020.

[9] Y. Liu, Y. Kang, T. Zou, Y. Pu, Y. He, X. Ye, Y. Ouyang, Y.-Q. Zhang, and Q. Yang, "Vertical Federated Learning: Concepts, Advances and Challenges," IEEE Internet of Things Journal, vol. 9, no. 22, pp. 22193–22213, 2022.