Department of Information Technology

Vidyalankar Institute of Technology
Accredited A+ by NAAC

BE Major Project

Semester: VIII

(ITB/10: Federated Learning for Smart Healthcare using Blockchain)

GROUP MEMBERS:

(ANISH KAMBLE- 21101B0017)

(IRFAN ANSARI- 21101B0038)

(DEVANK SHINDE- 21101B0047)

UNDER THE GUIDANCE OF

(PROF. NEHA KUDU)

AY 2024-25

Outline

Introduction

Motivation

Objectives

Problem Statement

Literature Survey (Papers and/or existing applications- prepare In tabular form)

Technology Stack (Hardware/Software)

Implementation

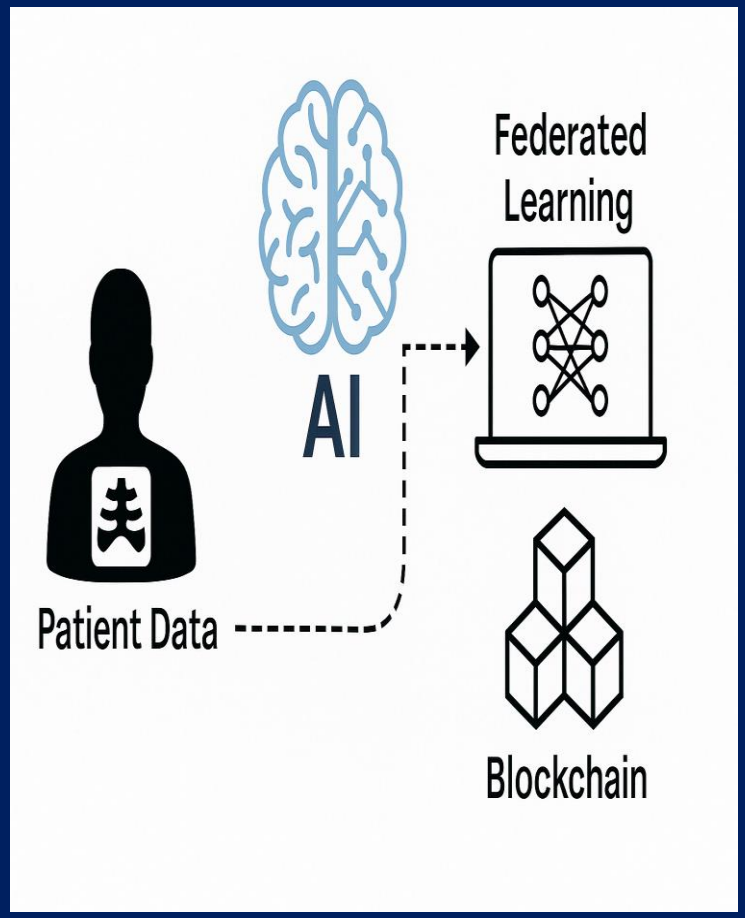
Result

References

Abstract

- Federated Learning (FL) enables secure, decentralized model training across institutions without sharing patient data.
- This framework combines FL with blockchain to enhance privacy and data integrity.
- Used for breast cancer detection using histopathology, mammography, and ultrasound datasets.
- Blockchain ensures immutable model logging and regulatory compliance.

Introduction



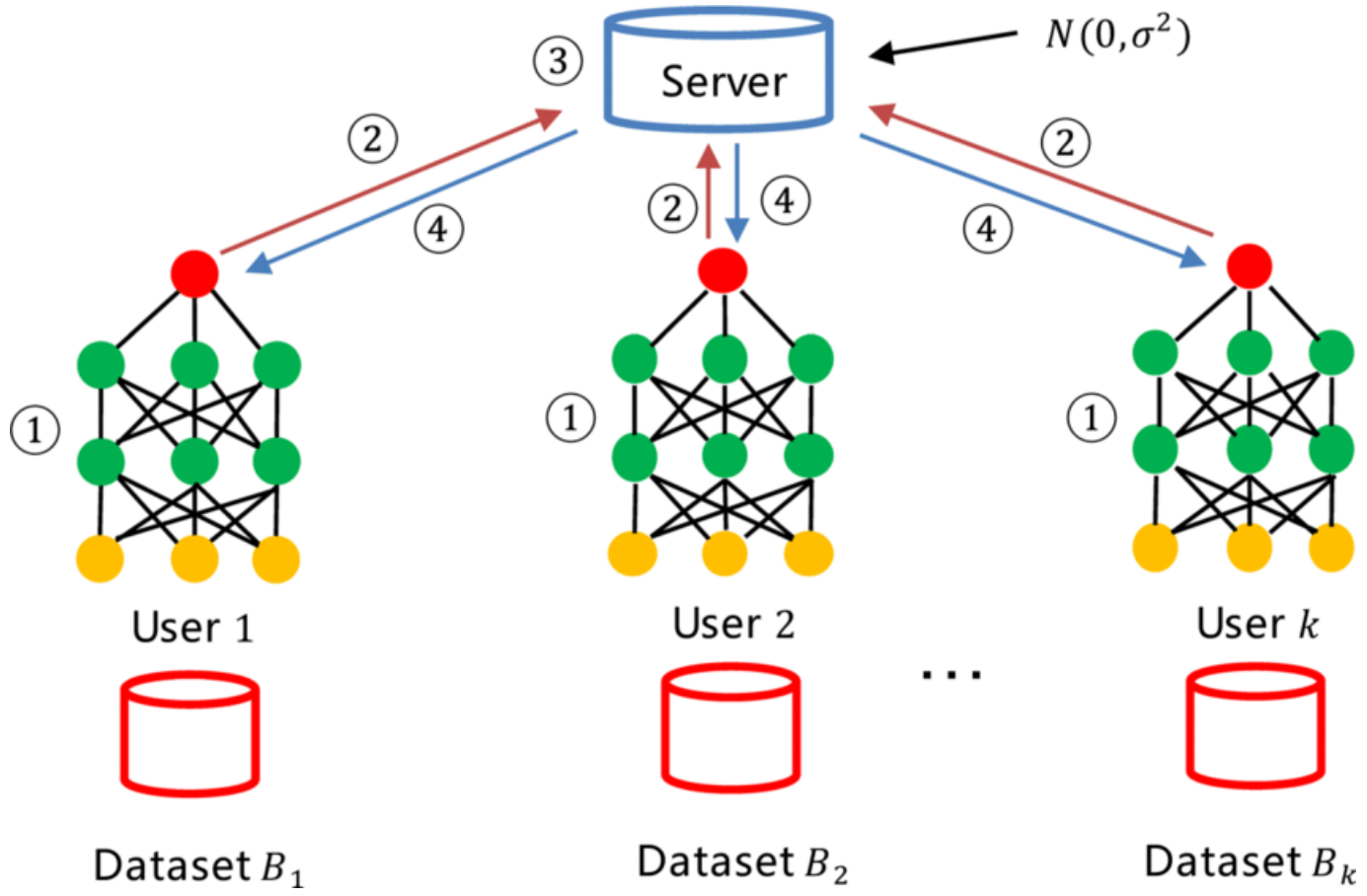
Artificial Intelligence (AI) has transformed healthcare by enabling faster and more accurate disease detection. However, training effective AI models typically requires large-scale patient data, raising concerns about privacy, data security, and regulatory compliance.

Federated Learning (FL) addresses this by allowing multiple institutions to train a shared model without sharing raw data. Each client trains locally and sends only model updates to a central server, preserving data confidentiality.

While FL protects privacy, it lacks mechanisms for verifying and tracing model updates. To enhance trust and security, we integrate blockchain technology to log cryptographic hashes of updates on an immutable ledger.

Our approach combines FL and blockchain to enable secure, transparent, and privacy-preserving AI for sensitive healthcare tasks like breast cancer detection.

- Each hospital (client) trains a local model on its own medical images (e.g., histopathology, mammography, ultrasound).
- Only model updates (not data) are sent to a central server.
- The server aggregates updates to create a global model.
- The improved model is shared back with all hospitals.
- Blockchain logs each update for transparency and trust.



Motivation

• **Privacy Concerns**: Centralized AI training exposes sensitive patient data, risking breaches and non-compliance with HIPAA/GDPR.

• **Federated Learning (FL)**: Enables model training across institutions without sharing raw data, preserving privacy.

• **Trust Gap in FL**: Lacks mechanisms for verifying and auditing model updates, making the system vulnerable to tampering.

• **Blockchain Integration**: Adds transparency, traceability, and security by immutably logging model updates on a decentralized ledger.

• **Healthcare Need**: Ensures privacy-preserving, secure, and collaborative AI development across hospitals and research centers.

Objectives

- **Enable Privacy-Preserving AI**
  Train models without sharing raw patient data.
- **Integrate Blockchain for Trust**
  Log model updates securely to ensure transparency and auditability.
- **Improve Diagnostic Accuracy**
  Use federated learning across diverse medical imaging datasets.
- **Ensure Regulatory Compliance**
  Align with HIPAA, GDPR, and healthcare data privacy standards.
- **Promote Secure Collaboration**
  Facilitate multi-institutional AI training with verifiable contributions.
- **Maintain System Scalability**
  Design a framework that works across various data sources and clients.

Problem Statement

- **Privacy Risks in Centralized Learning**
  Transferring medical data to central servers compromises patient confidentiality.
- **Lack of Trust in FL Alone**
  FL doesn’t provide mechanisms to verify or audit model updates.
- **Regulatory Challenges**
  Centralized data handling often violates HIPAA and GDPR standards.
- **Limited Collaboration**
  Institutions hesitate to participate due to security and ownership concerns.
- **No Built-in Audit Trail**
  Difficulty in tracing model contributions and ensuring update integrity.

Future Scope

- **Integrate Differential Privacy**
Add formal privacy guarantees to further protect sensitive data.
- **Adaptive Model Weighting**
Balance contributions from clients with varying data sizes and quality.
- **Scalability to More Institutions**
Extend the framework to support larger, real-world healthcare networks.
- **Optimize Blockchain Efficiency**
Use lightweight consensus mechanisms to reduce latency.
- **Cloud & Mobile Deployment**
Enable secure access and real-time updates across devices and platforms.
- **Explainable AI (XAI)**
Improve transparency in predictions for better clinical trust and usability.

Literature Survey & Existing Papers

Paper	Technology Used	Application	Key Findings
Kaissis et al., Nature Mach. Intell., 2020	Federated Learning	Medical Imaging	FL preserves privacy without sacrificing performance
Sheller et al., Brainlesion, 2019	FL (Multi-institutional setup)	Brain Tumor Segmentation	Feasible collaborative training without data sharing
Kuo et al., JAMA, 2017	Blockchain	Biomedical Data Management	Ensures trust, integrity, and access control
Nguyen et al., IEEE Trans. Intell. Transp. Syst., 2021	FL + Blockchain	Autonomous Systems	Blockchain adds verifiability and prevents update tampering
Dayan et al., Nature Medicine, 2021	Federated Learning	COVID-19 Outcome Prediction	FL is effective across hospitals for real-time model sharing
Rieke et al., npj Digital Medicine, 2020	Federated Learning	Digital Health	FL enhances generalizability in multi-center studies

Technology Stack

Component	Technology	Purpose
Model Training	TensorFlow, Keras	Build and train CNN models on medical image data
Federated Learning	Flower Framework	Coordinate local training and global aggregation
Blockchain	Hyperledger Fabric	Securely log model updates, ensure auditability
Programming Language	Python	Core development for ML, FL, and blockchain layers
Data Communication	REST APIs	Enable client-server interaction and update flow
Image Processing	OpenCV, NumPy	Preprocess histopathology, mammography, ultrasound
Evaluation & Logging	Matplotlib, TensorBoard	Visualize accuracy, loss, and training metrics

Implementation

1: Dataset Preparation
Assign and preprocess imaging data for each client (histopathology, mammography, ultrasound).

2 : Local Model Training
Clients train CNN models on local data using TensorFlow without sharing raw data.

3: Federated Learning Setup
Use the Flower framework to coordinate communication between server and clients.

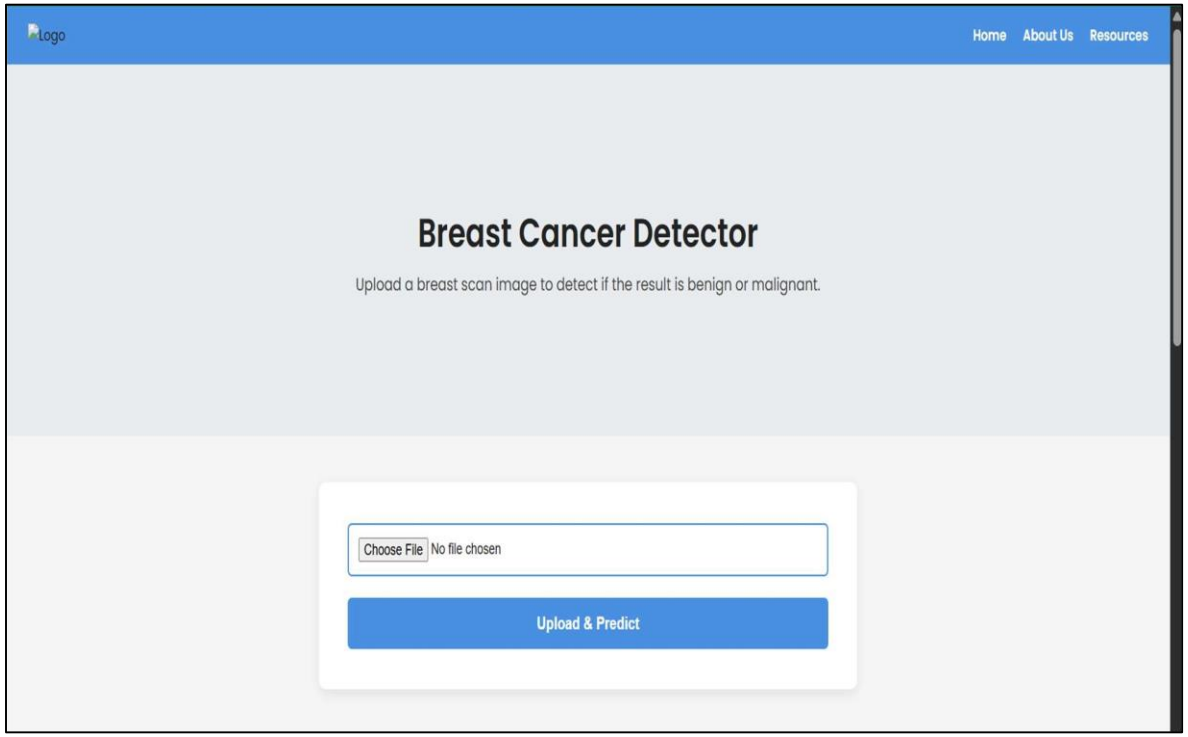
4 : Apply Blockchain Layer
Hash and log model updates to Hyperledger Fabric for integrity and traceability.

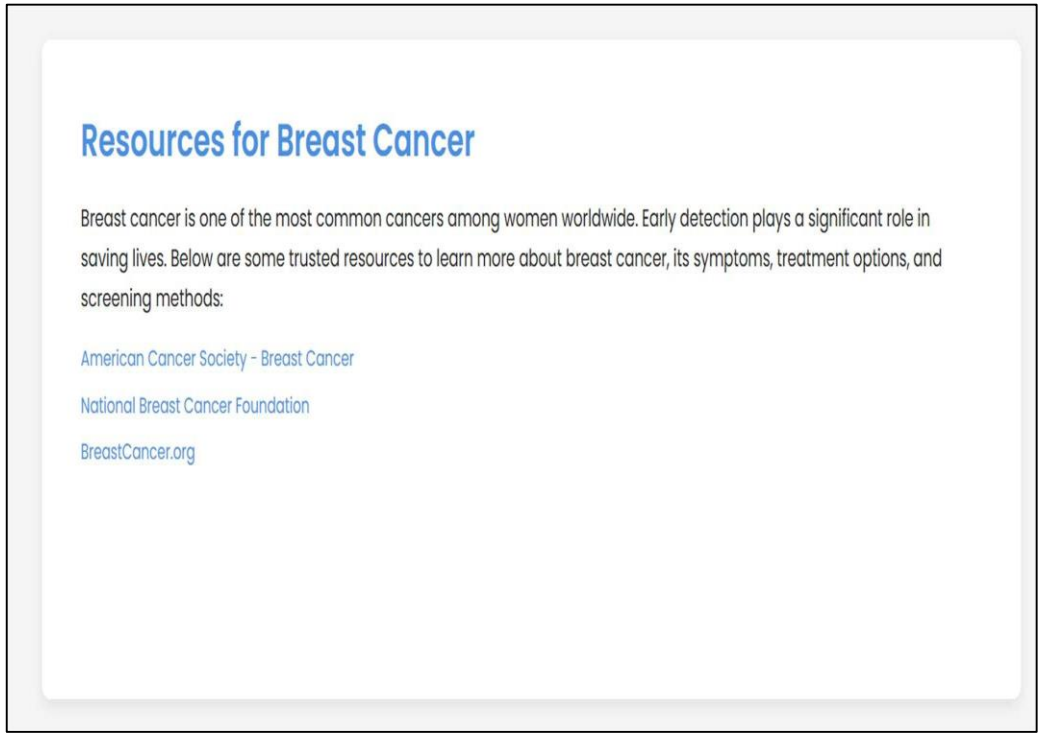
5: Model Aggregation
The server validates hashes and aggregates weights using Federated Averaging (FedAvg).

6: Model Evaluation
Track performance metrics (Accuracy, AUC, Loss) over communication rounds.

7: Compliance Check
Ensure the system meets HIPAA and GDPR requirements by keeping data local.

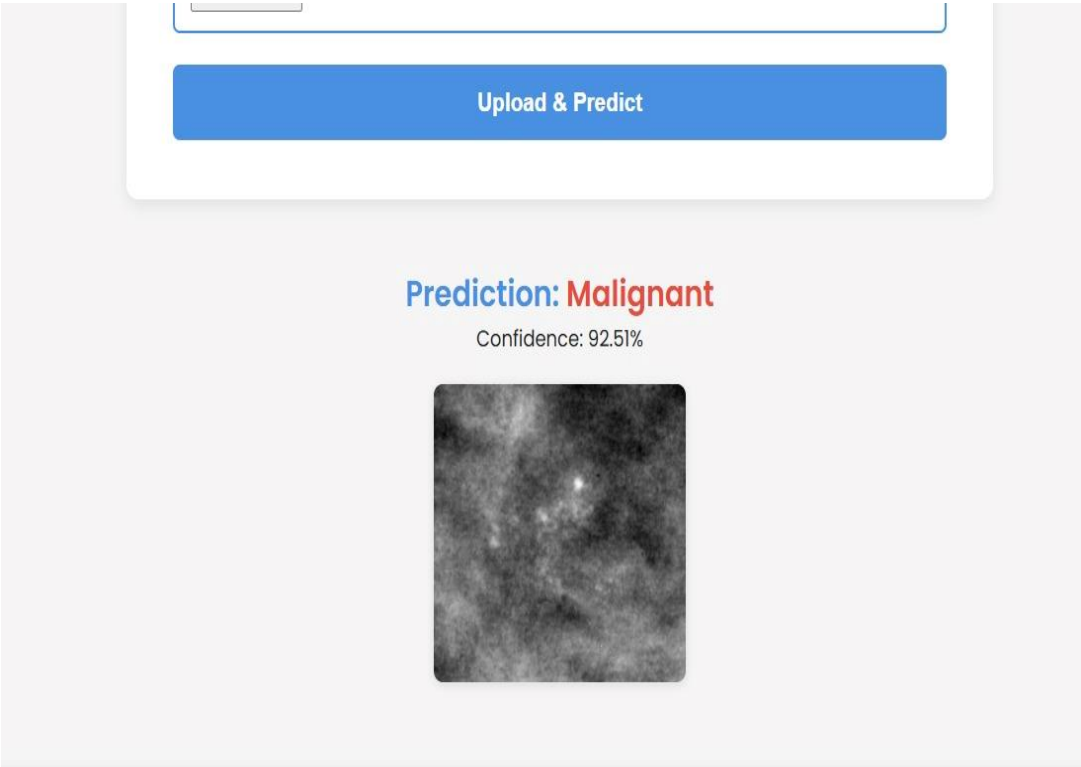
Result





LANDING PAGE

Result



Client	Final Accuracy	Final AUC	Observations
Histopathology	81.29%	0.8788	Best learning curve with balanced performance
Ultrasound	73.08%	0.5962	Strong baseline despite limited data
Mammography	86.02%	0.7542	High accuracy, AUC improved despite higher loss

References

No.	Title & Author(s)	Link
1	Secure, privacy-preserving and federated machine learning in medical imaging Kaissis et al., 2020	Link
2	Multi-institutional deep learning modeling without sharing patient data Sheller et al., 2019	Link
3	Blockchain distributed ledger technologies for biomedical and health care applications Kuo et al., 2017	Link
4	Federated learning with blockchain for autonomous vehicles Nguyen et al., 2021	Link
5	The future of digital health with federated learning Rieke et al., 2020	Link
6	Federated learning for predicting clinical outcomes in COVID-19 patients Dayan et al., 2021	Link