TECHNO INTERNATIONAL NEW TOWN

Project topic: **Privacy Preserving Using Federated Learning for Collaborative Healthcare Data Analysis.**

Mentored By- **DR. TAPASI BHATTACHARJEE**

Subject: **Project 1- IT792**

Group name: **TECHMIND INNOVATORS**

Group members

ANWARUL HAQUE (18700221042)
FAISAL SHAMIM (18700221057)
TAHSEEN ATIQUE ALI (18700221045)
MD MUJTABA (18700221047)

Abstract

- Protecting privacy in healthcare is crucial for collaborative data analysis.
- Centralized data sharing risks breaches, necessitating decentralized solutions.
- This work combines **Federated Learning (FL)** and **Homomorphic Encryption (HE)**:
  - **FL**: Enables model training without sharing patient data.
  - **HE**: Ensures secure computations on encrypted data.
- Allows secure aggregation of encrypted model updates across hospitals.
- Develops accurate prediction models while preserving confidentiality.
- Advances decentralized healthcare analytics with a privacy-first framework.

Problem Definition

Barriers to Collaborative Data Sharing - Healthcare data is siloed across organizations and it also limits innovation and results in isolated, less effective machine learning models.

Privacy Threats in Federated Learning (FL) - Parameter Leakage: Attackers infer sensitive data from shared gradients or model updates (gradient inversion) and **Malicious Participants**: Untrusted entities may cause data poisoning or backdoor attacks.

Lack of Robust Privacy Mechanisms - Homomorphic Encryption (HE): Stronger privacy guarantees but computationally expensive for resource-constrained settings.

Scalability and Reliability Issues - Heterogeneous Data: Institutional differences create biased models that fail to generalize and verifying participant reliability is critical to prevent malicious contributions.

Objective and Motivation

- **Maintaining privacy of Client**

Traditional systems , tend to have an interface pipeline where the data from the client is shared from a **third party** (usually cloud service).

- **Lack of Centralized Data**

Traditionally, developing and training a model on extensive, diverse data is resource-intensive, requiring significant space and time.

- **Reduced Communication Cost**

Transmitting only model updates rather than large datasets reduces network bandwidth usage.

Federated Learning

Federated Learning is a decentralized approach to machine learning where multiple devices or servers collaboratively train a model while keeping their data local. Instead of sending raw data to a central server, each participant trains the model on their local data and only shares model updates with a central server.

Homomorphic Encryption

Homomorphic Encryption is a form of encryption that allows computations to be performed on ciphertexts without needing to decrypt them first. The result of the computation, when decrypted, matches the result of operations performed on the plaintext.

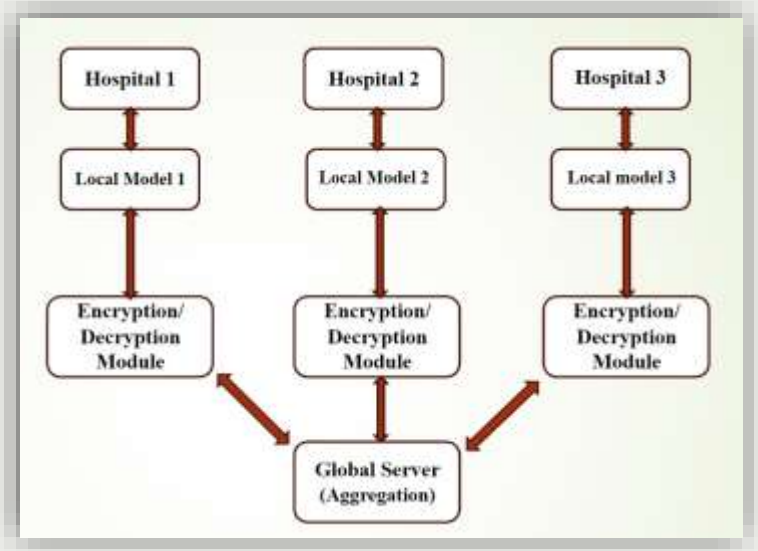
Proposed Methodology

Federated Learning approach:

A) Training occurs locally on user devices or nodes, using their private data.

B) Model updates (not raw data) are shared with a central server.

C) Global aggregation on the server refines the model without exposing sensitive data.



Homomorphic Encryption:

Ensures computations on encrypted data are possible without decrypting it. Adds a robust layer of privacy and prevents data breaches.

Implementation Tools/Technologies Used:

Development Environment:

- **Google Colab -** Cloud-based platform to simulate federated learning across multiple nodes.
- **Google Drive -** Acts as a centralized location for storing global models and aggregated updates.

Programming and Libraries:

- **Python -** Core programming language for implementing the federated learning framework.
- **TensorFlow Federated (TFF) -** A specialized library for simulating federated learning workflows.
- **Pandas and NumPy -** Used for efficient data preprocessing and manipulation.

System Architecture

Federated Learning:

- Hospitals and healthcare facilities act as independent clients.
- Each institution trains a machine learning model locally using sensitive patient data.

Privacy with Homomorphic Encryption:

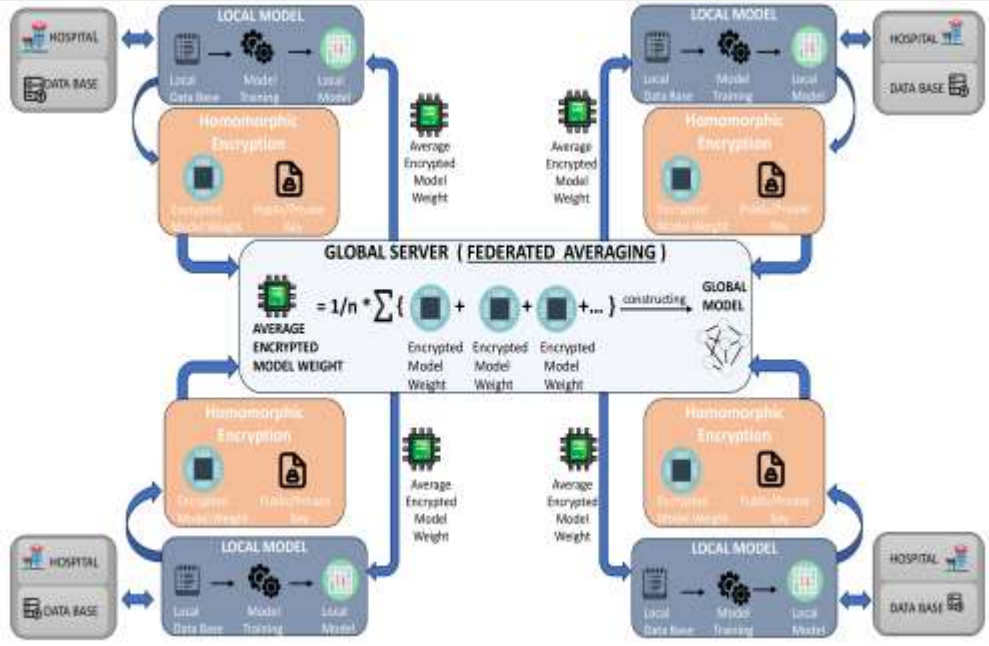
- Enables computations on encrypted data without revealing its content.
- Protects the privacy of model updates during training and transmission.

Central Server:

- Receives encrypted model updates, not raw data, from participating institutions.
- Aggregates updates to create a global model.

Collaboration Without Data Sharing:

- Hospitals collaborate to enhance the global model without sharing raw patient data.
- Patient privacy remains intact throughout the process.



Features of the Proposed Model

Homomorphic Encryption

- Ensures computations on encrypted data are possible without decrypting it.
- Adds a robust layer of privacy and prevents data breaches.

Real-Time Applications

- Specifically developed for hospital management systems.
- Can be integrated into existing systems to enhance privacy and data security.

Scalability

Designed to work seamlessly across multiple distributed nodes, ensuring no dependency on a centralized data pool.

Collaboration

Multi-user support to ensure secure data sharing among hospitals, researchers, or other stakeholders and maintains compliance with data protection regulations.

Workflow/Implementation

Data Preparation/ Initialization

- Each participating hospital or node maintains its own dataset locally.
- Data preprocessing is performed at the local level to ensure readiness.

Federated Learning Model Training.

- Each node trains the model on its local dataset.
- Local models are encrypted before sharing updates with the global server.

Global Aggregation

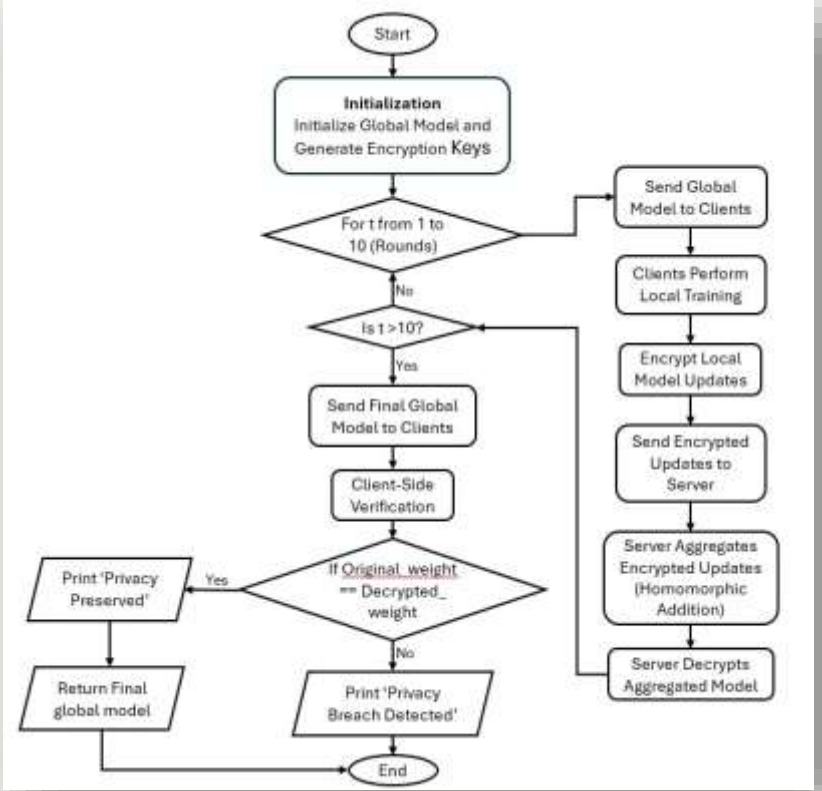
- The global server aggregates encrypted updates from all nodes.
- Homomorphic encryption ensures computations are performed without exposing raw updates.

Model Distribution

- The updated global model is shared back with all nodes.
- Participating nodes benefit from collective learning while maintaining data privacy.

Iteration

- Steps 2 to 4 are repeated for multiple rounds to improve model accuracy.

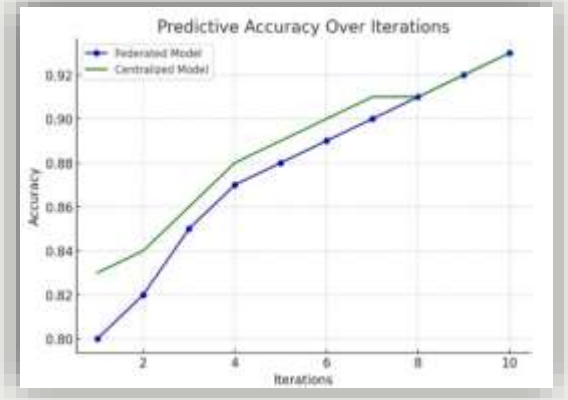


Outcomes

The effectiveness of Federated Learning (FL) and Homomorphic Encryption (HE) in securely analyzing healthcare data, focusing on predictive accuracy, model robustness, privacy security, and computational efficiency across different healthcare facilities.

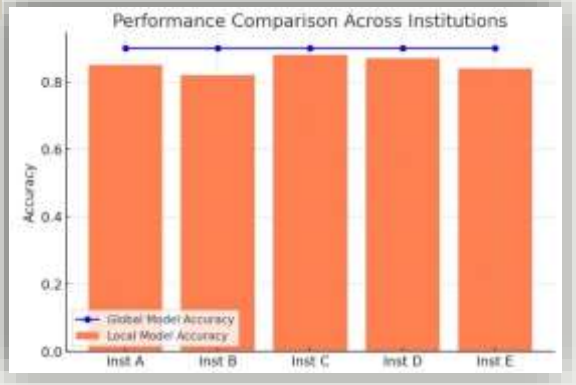
Predictive Accuracy -

- The Federated Learning model aggregated encrypted updates from healthcare facilities.
- Achieved predictive accuracy comparable to centralized models.
- Ensured high-quality results while safeguarding data confidentiality.



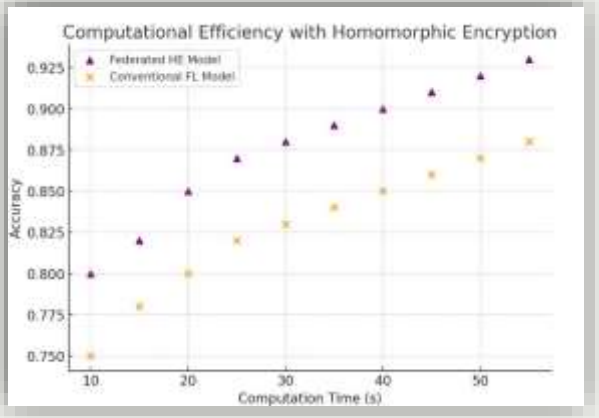
Model Robustness -

- The model demonstrated robustness across training cycles with minimal differences between local and global models.
- Encrypted updates improved accuracy with each iteration.



Computational Efficiency -

- Homomorphic Encryption with Federated Learning improved computational efficiency.
- Ensured controllable processing speeds and maintained prediction accuracy.



FEDERATED LEARNING

Data Localization: Data remains decentralized at individual institutions, avoiding the need to move sensitive data to a central repository.

Privacy Preservation: Protects privacy by sharing encrypted model updates instead of raw data.

Lower Risk of Data Breaches: Since raw data isn't transmitted, the risk of breaches is significantly reduced.

CENTRALIZED LEARNING

Data Aggregation: Requires collecting and storing all data in a central repository for training, increasing the risk of breaches and unauthorized access.

Less Emphasis on Privacy: Privacy depends on securing the centralized server, making it vulnerable to attacks.

Dependence on Data Centralization: Institutions must trust the central server to securely handle and analyze sensitive data.

Challenges In federated learning

Data imbalance

(Data distribution across devices may vary, leading to biased model training and reduced accuracy.)

Statistical heterogeneity

(Variations in data generated by different users or devices make it difficult to develop a universal model)

System heterogeneity

(Devices involved in federated learning have different computational capabilities, network speeds, and storage, causing inefficiencies)

Resources allocation

(Efficient allocation of computational and communication resources is challenging due to device and network constraints)

Expensive communication

(Transmitting model updates frequently between devices and the central server is bandwidth-intensive and costly)

REAL LIFE APPLICATIONS

Secure Wearable Device Data Integration

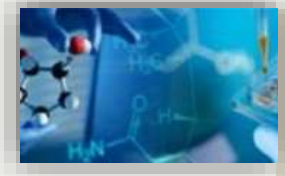
helps in aggregating data from multiple devices, the model enables accurate analysis for early chronic disease detection.

Enhanced Remote Patient Monitoring

Decentralized data exchange between healthcare providers is made possible by federated learning, which protects patient privacy while fostering collaborative insights.

Collaborative Clinical Research

Through the safe sharing of encrypted datasets, the suggested model facilitates collaboration between institutions.



Precision Medicine and Drug Development

To develop customized drugs, pharmaceutical corporations use real-world health data.

KEY TAKEAWAYS

01

Privacy-Preserving Data Analysis:
Federated Learning (FL) combined with Homomorphic Encryption (HE) ensures secure and private machine learning. No raw data is shared, safeguarding sensitive patient information.

02

Collaborative Healthcare Innovation:
Enables healthcare institutions to collaboratively train predictive models without violating privacy regulations like GDPR and HIPAA.

03

High Accuracy and Robustness:
Achieves near-centralized model performance, demonstrating scalability and reliability across diverse datasets.

04

Real-Time Applications:
Supports IoT and EHR integration, paving the way for real-time analytics in healthcare systems.

05

Future-Ready Framework:
Optimized for scalability, efficiency, and compliance, this model sets a benchmark for privacy-conscious data analytics.

Future Scope

Enhanced Privacy Mechanisms:

- Integrate advanced cryptographic techniques like Differential Privacy to complement Homomorphic Encryption.
- Research on mitigating emerging threats, such as adversarial attacks on federated models.

Scalability Across Sectors:

- Expand beyond healthcare to domains like finance, education, and smart cities.
- Adapt the framework for large-scale deployments involving diverse institutions.

Real-Time Federated Analytics:

- Incorporate edge computing to enable faster training and analysis for real-time applications.
- Develop adaptive models that can respond dynamically to changes in decentralized environments.

Integration with IoT Devices:

- Utilize federated learning in wearable healthcare devices to enable personalized diagnostics and monitoring.
- Aggregate data from IoT ecosystems while ensuring privacy compliance.

THANK YOU,

wishing you all a great day ahead…