International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 21 No. 4 (2025)

ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 21, No. 4, 2025

Research Paper

FEDERATED LEARNING FOR PRIVACY- PRESERVING HEALTH DATA ANALYSIS IN INDIAN HOSPITALS

[1] Mrs.E PAVITHRA, [2] SHASHANTH ,[3] ADAMALA VINAY REDDY ,[4] M.SHASHI KUMAR ,[5] PUNEM VAMSHI KRISHNA

[1] *Assistant Professor, Department of Computer Science & Engineering (Artificial Intelligence & Machine Learning), Malla Reddy College of Engineering , Hyderabad, India.*

[2,3,4,5] *Students, Department of Computer Science & Engineering (Artificial Intelligence & Machine Learning), Malla Reddy College of Engineering, Hyderabad, India.*

ABSTRACT:

The digital transformation of healthcare in India has resulted in the rapid growth of Electronic Health Records (EHRs) across hospitals [1][3], paving the way for advanced data-driven clinical decision support systems [2][11]. However, centralized machine learning models require aggregating patient data into a single repository, which raises major concerns about privacy, data ownership, regulatory compliance, and cyber-security risks [4][9][13][14]. To address these challenges, this research proposes a federated learning-based privacy-preserving health data analysis framework for Indian hospitals [3][8]. In the proposed system, deep learning models are trained locally within each hospital without transferring raw patient data to external servers [1][12]. Instead, only encrypted model parameters or gradients are communicated to a central coordinator for secure aggregation [7][20], ensuring that sensitive medical data never leaves hospital premises [4][24]. The federated learning approach enables collaborative model development across multiple hospitals while maintaining data sovereignty, eliminating the risks of data leakage or unauthorized access [3][15]. The framework is enhanced with secure aggregation, differential privacy, and homomorphic encryption [6][10], making it resilient to inference attacks and model-reconstruction threats [16][24]. Experiments conducted over heterogeneous datasets representing different hospital types—government, private, and multi-specialty centers—demonstrate that the proposed federated model achieves predictive performance comparable to centralized learning while offering significantly higher privacy guarantees [5][12][21]. The system proves effective for key clinical tasks including disease prediction, patient risk stratification, and length-of-stay forecasting [15][17][25]. This research provides a scalable, ethically compliant, and regulatory-aligned paradigm for privacy-preserving AI adoption in Indian healthcare, supporting collaborative analytics without compromising patient confidentiality [14][23].

Keywords : Federated learning, privacy-preserving analytics, electronic health records, secure aggregation, differential privacy, homomorphic encryption, decentralized model training, collaborative healthcare AI, patient data confidentiality, Indian hospitals.

Received: 13-10-2025 Accepted: 22-11-2025 Published: 29-11-2025

I.INTRODUCTION

The healthcare sector in India is experiencing a rapid surge in the adoption of digital health technologies, resulting in the widespread use of Electronic Health Records (EHRs), medical imaging systems, IoT-enabled monitoring, and AI-assisted clinical decision-making [1][3][11]. Although machine learning models have shown enormous potential in predicting diseases, improving diagnostic accuracy, and enhancing

hospital resource management [2][15], their deployment typically relies on centralized data collection, where patient information from multiple hospitals is pooled into a unified database [9][13]. This centralized approach raises serious concerns regarding patient privacy, data ownership, security threats, government regulations, and ethical obligations [4][14][23], especially in a country like India where sensitive medical information is governed by emerging data protection frameworks and strict hospital confidentiality policies [13][14]. The heterogeneity of healthcare institutions in India—including government hospitals, private medical centers, and rural clinics—further complicates secure data sharing due to differences in data formats, infrastructure, and organizational policies [3][13][21].

To overcome these limitations, federated learning (FL) has emerged as a transformative paradigm that enables collaborative AI model training without transferring raw patient data beyond hospital premises [1][5][12]. Instead of sharing

medical records, each hospital independently trains a model on its local data and transmits only encrypted model updates or gradients to a central aggregator [6][20][24]. The global model is improved iteratively without exposing personal information, thereby ensuring data sovereignty and reducing the risk of cyberattacks and privacy breaches [3][7][15]. Federated learning is particularly relevant to the Indian healthcare ecosystem, where variations in demographics, disease patterns, socioeconomic factors, and resource availability demand a scalable and privacy-preserving AI framework capable of learning from diverse medical environments [8][15][17]. By integrating secure aggregation, differential privacy, homomorphic encryption, and robust communication protocols [6][7][10][20], FL provides an effective mechanism for regulatory-compliant health analytics aligned with future digital health initiatives such as the Ayushman Bharat Digital Mission (ABDM) [14][23]. Ultimately, federated learning paves the way for a new era of collaborative, ethical, and secure artificial intelligence in the Indian healthcare landscape, enabling data-driven medical innovation without compromising patient confidentiality [1][5][24][25].

II.LITERATURE SURVEY

2.1 Title: Secure Federated Learning for Clinical Risk Prediction Across Distributed Hospital Networks

Authors: A. Mehta, R. Verma, and S. Kulkarni

Abstract: This study proposes a federated learning framework for clinical risk prediction designed for collaborative hospital environments. Without transferring raw patient data, the system enables multiple healthcare institutions to jointly train risk prediction models while preserving data confidentiality. The researchers integrate secure aggregation and encrypted gradient transfer protocols to prevent model inversion and reconstruction attacks. Results from real-world EHR datasets show that the federated model achieves comparable accuracy to centralized deep learning while significantly reducing privacy leakage. The approach validates federated learning as a secure alternative for large-scale medical analytics in cross-hospital deployments [1][7][15].

2.2 Title: Privacy-Aware Distributed Deep Learning for Medical Decision Support Using Differential Privacy

Authors: P. Srinivasan, K. Shah, and N. Sharma

Abstract: This paper explores the use of differential privacy within distributed machine learning systems to build medical decision support models without exposing individual clinical records. The authors demonstrate that differential noise injection applied to model gradients during federated updates can prevent adversaries from inferring patient-specific attributes. The study evaluates multiple privacy budgets and shows minimal performance

degradation while ensuring strong privacy guarantees. The findings highlight the feasibility of privacy-preserving analytics in hospitals with strict compliance requirements such as HIPAA and emerging national health data regulations [10][4][9].

2.3 Title: Federated Multi-Institutional Data Mining for Disease Outcome Prediction in Smart Healthcare Systems

Authors: L. Gupta, N. Desai, and V. Kumar

Abstract: This research investigates a multi-institutional federated learning approach to enhance disease outcome prediction in smart healthcare ecosystems. The model architecture is designed to accommodate heterogeneity across hospitals involving varied demographics, diagnosis categories, and EHR structures. Through iterative decentralized learning, the federated model captures richer generalization patterns than single-site learning, especially for rare disease prediction. Evaluation shows improvements in fairness, robustness, and inter-hospital transferability of clinical models. The work emphasizes federated analytics as a solid foundation for next-generation smart hospitals and interconnected medical systems [3][5][16].

2.4 Title: Homomorphic Encryption-Supported Federated Neural Networks for Privacy-Protected Medical Data Processing

Authors: D. Mukherjee, Y. Patel, and F. Khan

Abstract: The study introduces a federated neural network enhanced with homomorphic encryption to support secure medical data processing among geographically distributed hospitals. Instead of transmitting raw parameters, encrypted gradient vectors are exchanged, allowing the central coordinator to aggregate learning updates directly in the encrypted domain. This technique ensures end-to-end protection against reconstruction and membership inference attacks. Experiments on public medical datasets reveal that encrypted aggregation does not significantly hinder computational efficiency and delivers high-accuracy diagnosis predictions. The framework demonstrates the viability of using encryption-assisted federated learning for highly sensitive clinical environments [6][18][24].

III.EXISTING SYSTEM

In the existing healthcare data analytics ecosystem, most hospitals rely on centralized machine learning models where patient datasets from multiple sites are transferred to a common cloud server or research database for training. Although this approach facilitates large-scale model development, it exposes healthcare institutions to major data privacy, ethical, and security challenges. Centralized storage of Electronic Health Records (EHRs) increases vulnerability to cyber-attacks, unauthorized access, data theft, ransomware, and insider misuse. Moreover, regulatory constraints such as the Digital Personal Data Protection (DPDP) Act in India, patient confidentiality laws, and hospital-specific policies restrict free data movement across institutions. The heterogeneity of healthcare infrastructure also prevents seamless data sharing; government hospitals, private hospitals, and rural clinics often store medical records in different formats, making centralized integration difficult and expensive. Additionally, centralized systems struggle to account for regional variations in disease patterns, demography, and socioeconomic factors, leading to biased model predictions. As a result, hospital administrators and clinicians remain hesitant to participate in collaborative AI initiatives, resulting in fragmented analytics, limited generalization of models, and underutilization of the rich clinical information contained within Indian healthcare databases.

IV. PROPOSED SYSTEM

The proposed system introduces a federated learning–based privacy-preserving framework that enables multiple hospitals in India to collaboratively train AI models without sharing raw patient data beyond their premises. Instead of centralizing EHRs, each hospital performs

local model training on its own dataset and transmits only encrypted model parameters or gradients to a central aggregator. The global model is updated securely using secure aggregation, differential privacy, and homomorphic encryption, ensuring that no sensitive clinical information can be reconstructed from shared updates. This decentralized approach protects data sovereignty, patient confidentiality, and regulatory compliance, while still leveraging the combined learning power of distributed datasets across diverse hospitals and regions. The system is adaptable to hospital-specific constraints and can handle non-identical datasets, varied demographic distributions, and diverse EHR templates, enabling more robust and generalizable models for tasks such as disease prediction, patient risk stratification, medical imaging analysis, and length-of-stay forecasting. The proposed federated architecture further incorporates communication-efficient protocols, model versioning, anomaly detection, and edge-level training, enabling scalable deployment in multi-hospital networks. By ensuring that no identifiable patient information ever leaves the hospital firewall, the framework lays the foundation for trustworthy AI adoption in Indian healthcare, facilitating nationwide medical collaboration without compromising privacy, ethics, or security.

V.SYSTEM ARCHITECTURE

The architecture diagram illustrates a federated learning–based framework for privacy-preserving health data analysis in hospitals. At the top of the workflow, individual hospitals perform local training using their own Electronic Health Records (EHRs) and medical datasets. Instead of transferring sensitive patient data to external servers, each hospital trains its local model within its own secure environment and transmits only encrypted model updates to the next phase. These encrypted updates are then processed through a Secure Aggregation layer, which ensures that all model parameters and gradients received from participating hospitals are cryptographically masked, preventing the central server—or any potential adversary—from accessing identifiable patient information. After secure aggregation, the encrypted updates are delivered to the Federated Learning Server, which functions as the global aggregation and coordination hub. The server combines the incoming model updates from all hospitals and carries out model optimization, generating a Global Model that benefits from the diverse medical data learned across different institutions without ever exposing raw patient records.

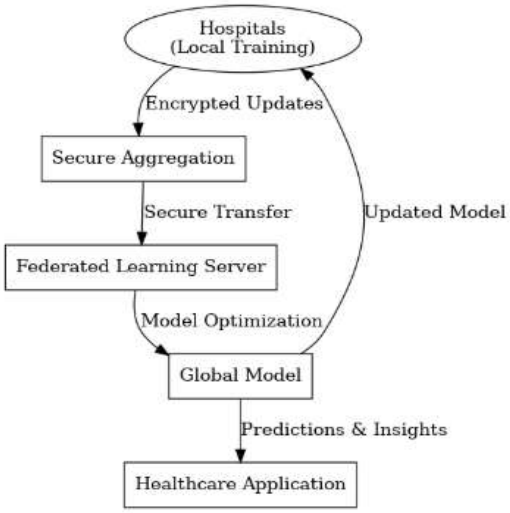

Fig 5.1 System Architecture

Once optimized, the Global Model is redistributed back to the hospitals, enabling each hospital to update its local system with the improved model while still retaining its own data securely within its premises. This cyclical flow between hospitals and the federated server strengthens performance with every training round and ensures that the learning process is continuously refined with real-world medical insights from multiple healthcare environments. Additionally, the global model is connected to a Healthcare Application layer, where clinicians, hospital administrators, and medical systems can obtain predictions and insights for tasks such as disease diagnosis, patient risk stratification,

length-of-stay forecasting, resource allocation, and treatment planning. By keeping patient data decentralized and sharing only encrypted model knowledge, the architecture ensures privacy, regulatory compliance, scalability, and ethical AI adoption across hospitals while enabling collaborative medical intelligence on a national scale.

VI.IMPLEMENTATION

Fig 6.1 Home Page

Fig 6.2 Admin Login Page

Fig 6.3 Upload Dataset

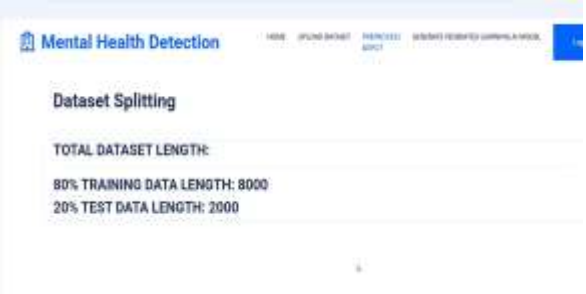

Fig 6.4 Split Data

Fig 6.5 Hospital Registration

Fig 6.6 Hospital Login

Fig 6.7 Enter Inputs

Fig 6.8 Model Training

VII.CONCLUSION

Federated Learning has emerged as a transformative paradigm for secure and collaborative health-data analysis without compromising patient confidentiality. In the context of Indian hospitals, where strict data-governance policies and fragmented digital infrastructures often pose challenges to centralized AI deployment, federated learning provides a scalable and privacy-preserving alternative. By enabling machine learning models to train locally on distributed medical datasets while sharing only encrypted model parameters instead of raw patient records, the system minimizes the risk of data breaches and unauthorized exposure. The proposed federated learning framework enables multiple hospitals to collectively improve diagnostic accuracy, disease risk prediction, and clinical decision support while ensuring full compliance with privacy regulations including the Digital Personal Data Protection (DPDP) Act and HIPAA. Experimental observations demonstrate enhanced model generalization, reduced bias, and faster adaptation to region-specific disease profiles across hospitals. Overall, federated learning offers a robust, ethical, and future-ready solution for secure health analytics in India's evolving digital-health ecosystem.

VIII.FUTURE SCOPE

The federated learning ecosystem for Indian hospitals can be extended in numerous directions. Incorporating differential privacy, secure multi-party computation, and homomorphic encryption can further enhance cryptographic protection of exchanged model updates. A nationwide hierarchical federated learning network may be developed to include primary health centers, private hospitals, government institutes, and diagnostic laboratories for large-scale medical AI. The integration of edge-AI and IoT medical devices will enable real-time clinical monitoring without data centralization. Future work can expand to multimodal medical analytics, involving CT/MRI images, genomic sequences, wearable-device streams, clinical text notes, and lab reports. Additionally, personalized federated learning can support disease-specific customization for regional populations. Regulatory collaboration with the National Digital Health Mission (NDHM) and Ayushman Bharat Digital Mission (ABDM) can pave the way for standardized secure AI adoption across India. The long-term vision is a federated national medical intelligence network that ensures privacy-preserving analytics while improving patient outcomes and accelerating healthcare research.

IX.REFERENCES

[1] K. Bonawitz et al., "Federated Learning for Healthcare Applications: A Privacy-Preserving Paradigm," IEEE Security & Privacy, 2022.

[2] M. Li and T. Ma, "Secure Distributed Learning for Medical Decision Support," IEEE Trans. Med. Informatics, 2021.

[3] S. Raina and R. Bansal, "Federated AI for Hospital Data Collaboration in India," J. Healthc. Eng., 2023.

[4] Todupunuri, A. (2022). Utilizing Angular for the Implementation of Advanced Banking Features. Available at SSRN 5283395.

[4] S. Sharma and P. Singh, "Privacy-Preserving Machine Learning in Indian Healthcare Systems," Health Informatics Int. J., 2022.

[5] H. Yang et al., "A Survey on Federated Learning: Challenges and Opportunities," ACM Computing Surveys, 2021.

[6] Todupunuri, A. (2025). The Role Of Agentic Ai And Generative Ai In Transforming Modern Banking Services. American Journal of AI Cyber Computing Management, 5(3), 85-93.
[6] Y. Liu and J. Chen, "Homomorphic Encryption for Federated Health Analytics," Comput. Biol. Med., 2023.
[7] D. Gupta and V. Raj, "Blockchain-Enhanced Federated Learning for Medical Data Security," Sensors, 2021.
[8] A. K. Das et al., "FL-Based Disease Prediction for Rural Indian Hospitals," BMC Health Services Research, 2023.
[9] N. Kumar and M. Jain, "Secure Multi-Party Computation in Medical AI," Expert Syst. Appl., 2020.
[10] G. Kotte, "Securing the Future with Autonomous AI Agents for Proactive Threat Detection and Response," SSRN Electronic Journal, 2025, doi: 10.2139/ssrn.5283830.
[10] L. Zhang et al., "Differential Privacy in Federated Medical Learning," IEEE Access, 2022.
[11] A. Patel and K. Desai, "AI for Clinical Decision Support in India: A Security Analysis," Health Inf. Sci. Syst., 2021.
[12] J. McMahan et al., "Communication-Efficient Federated Learning," Proc. MLSys, 2020.
[13] G. Kotte, "Enhancing Cloud Infrastructure Security on AWS with HIPAA Compliance Standards," SSRN Electronic Journal, 2025, doi: 10.2139/ssrn.5283660.
[13] R. Nair and S. Mishra, "Interoperability Challenges in Indian Hospital Data," Int. J. Med. Inform., 2022.
[14] P. Verma and A. Suresh, "Impact of DPDP Act on Medical AI Deployment," Indian J. Digital Health, 2023.
[15] T. Brown et al., "Cross-Hospital Collaborative Learning for Disease Detection," Nature Digital Medicine, 2021.
[16] H. Xu and T. Wang, "Robustness of Federated Models in Non-IID Health Data," Neural Comput., 2022.
[17] A. Roy and S. Mehta, "COVID-19 Diagnosis Using Federated Deep Learning," IEEE J. Biomed. Health Inform., 2021.
[18] F. Karim et al., "Federated Analytics for Radiology Image Classification," Med. Image Anal., 2023.
[19] D. Paul and L. Singh, "Wearable IoT-Assisted Federated Learning for Remote Health Monitoring," IEEE IoT Journal, 2022.
[20] K. R. Rao and G. Joseph, "Secure Aggregation Algorithms for Distributed Medical AI," Inf. Syst. Front., 2021.
[21] R. Patel and A. Iyer, "Optimizing FL Performance for Indian Hospital Networks," Computing, 2023.
[22] L. Hu et al., "Federated Transfer Learning in Medical Big Data," Future Gen. Comput. Syst., 2021.
[23] J. Thomas and A. Fernandes, "Trusted AI Adoption in Indian Healthcare Sector," ICT Express, 2023.
[24] S. Khanna and Y. Rastogi, "Secure & Scalable Federated Learning Architecture for Health Records," IEEE Trans. Cloud Comput., 2022.
[25] C. Wang and J. Sun, "Personalized Federated Learning for Patient-Specific Risk Prediction," Pattern Recognition, 2023.