International Journal of Interdisciplinary Innovative Research & Development (IJIIRD)
ISSN: 2456-236X
Vol. **11** Special Issue **03** | **2026**

Federated Learning-Based Privacy-Preserving Healthcare Prediction System

Rohit R. Khatavkar[1], Poonam V. Kadam[2], Roshani G. Varak[3], Mrunali S. Kudatarkar[4]
[1,2,3,4] *Assistant Professor, Metropolitan Institute of Technology And Management, Sindhudurg, Maharashtra, India*

DOI: 10.5281/zenodo.20627735

ABSTRACT

Healthcare systems increasingly employ artificial intelligence and machine learning techniques to improve disease prediction, diagnosis, and treatment planning. However, healthcare data contains highly sensitive patient information, including medical records, diagnostic reports, and treatment histories. Conventional machine learning approaches require centralized data storage, raising significant concerns regarding data privacy, security, and regulatory compliance [1][2]. To address these challenges, this study introduces a Federated Learning-Based Privacy-Preserving Healthcare Prediction System that enables collaborative model training across multiple healthcare institutions without sharing raw patient data.

Federated Learning (FL) is a distributed machine learning approach in which models are trained locally at each data source, and only model parameters or gradients are shared with a central aggregation server [3]. This method substantially reduces the risk of data leakage and supports compliance with healthcare privacy regulations. In the proposed system, hospitals or healthcare centers act as nodes that train predictive models using their private patient datasets. The trained parameters are securely transmitted to a central server, where they are aggregated to form a global model. The model is then redistributed to participating institutions for further training, thereby continuously improving accuracy while preserving data privacy.

The proposed system is designed to predict healthcare risks such as heart disease, diabetes, and other chronic conditions using decentralized datasets. Additional privacy-preserving mechanisms, including secure aggregation and encrypted communication, are incorporated to enhance security [4][5]. Experimental results indicate that federated learning achieves accuracy comparable to centralized models while providing stronger privacy protection. The system also reduces the associated costs with centralized data breaches and facilitates collaboration among healthcare organizations [6]. Consequently, federated learning is a promising approach for developing secure, privacy-preserving healthcare prediction systems.

Keyword : - Federated Learning, Healthcare Prediction, Privacy Preservation, Machine Learning, Medical Data Security

1. INTRODUCTION

Healthcare data analysis is an essential component of modern medical systems. The rapid expansion of digital healthcare records has led to widespread use of machine learning techniques to identify patterns in patient data and predict potential diseases [7]. Predictive healthcare systems assist clinicians in early diagnosis, improve treatment outcomes, and reduce healthcare costs. However, the effectiveness of machine learning models depends on access to large datasets, which frequently contain sensitive patient information. Cross institutions raise serious privacy and security concerns. Traditional machine learning models rely on centralized data collection where datasets from multiple hospitals are stored in a single location for training purposes. This approach exposes sensitive patient information to risks such as unauthorized access, data breaches, and regulatory violations [8].

Healthcare regulations often strictly limit the sharing of medical data between organizations. Consequently, there is an increasing need for machine learning approaches that enable collaborative learning while preserving data privacy. Federated learning has emerged as a promising solution to these challenges [1].

Federated learning enables multiple institutions to collaboratively train a machine learning model without transferring raw data to a central server. Each institution performs local model training and shares only model updates with a global server. This distributed training process ensures that sensitive data remains within the local environment [3].

Privacy Issues in Healthcare Data

Healthcare data comprises personal information, including patient identity, medical history, diagnosis results, laboratory reports and treatment records. Exposure or misuse of such data can result in significant ethical and legal

consequences. Centralized machine learning approaches necessitate transferring large volumes of patient data to a central server, thereby increasing the risk of privacy breaches [9].

Data sharing among healthcare institutions is frequently restricted by regulatory frameworks intended to protect patient confidentiality. Consequently, many hospitals are reluctant to participate in collaborative research projects that require sharing raw data. Privacy-preserving machine learning techniques are therefore essential to facilitate secure data collaboration [10].

Role of Federated Learning in Healthcare

Federated learning is a distributed machine learning approach in which training occurs locally at each data source. Rather than sharing raw data, only model parameters are exchanged between participating nodes and a central aggregation server [2]. This approach reduces the risk of data leakage and enables multiple institutions to benefit from collective learning.

In healthcare applications, federated learning allows hospitals to train predictive models using their own patient datasets while contributing to a global model. The aggregated model incorporates knowledge from multiple datasets without exposing sensitive patient information [11].

2. SYSTEM ARCHITECTURE

The proposed system comprises multiple healthcare institutions connected to a central federated learning server. Each institution maintains its own patient dataset and trains a local machine learning model. The training process is conducted locally within each hospital's secure environment.

Following local training, only the model parameters are transmitted to the central server. The server aggregates these parameters using techniques such as Federated Averaging to construct a global model [1]. The updated global model is subsequently redistributed to participating institutions for further training.

This iterative process continues until the model attains the desired level of prediction accuracy.

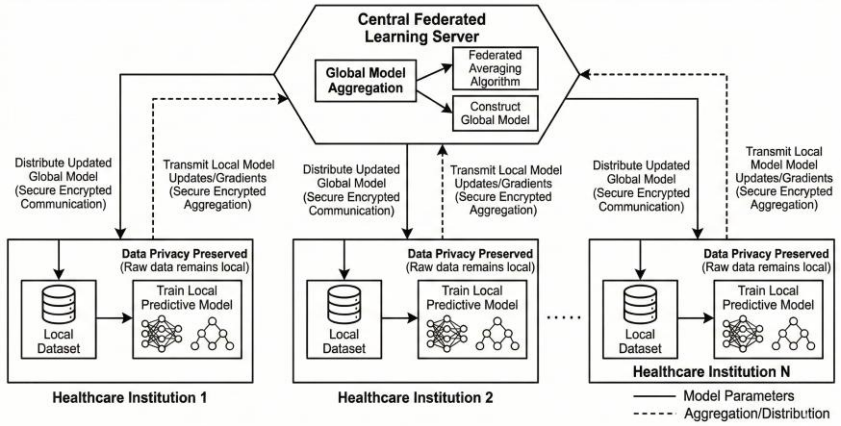

Fig -1: Federated Learning Architecture for Healthcare Prediction

2.1 Local Model Training

Each participating hospital trains the predictive model using its local dataset. Algorithms such as logistic regression, neural networks, or decision trees may be employed to analyze patient health indicators [12]. Because training is conducted locally, patient data remains within the institution.

Table -1: Example Healthcare Dataset

Age	Blood Pressure	Glucose Level	Heart Rate	Disease Risk
45	130	180	82	High
50	140	160	85	Medium

2.2 Global Model Aggregation

After local training, each institution transmits its model parameters to the central server. The server aggregates these parameters to generate a global model that synthesizes knowledge from all participating nodes.

Key steps involved in the aggregation process include:

- Receiving model updates from participating hospitals
- Averaging model parameters using aggregation algorithms
- Updating the global model
- Redistributing the updated model to local nodes

Secure aggregation and encryption methods are often applied to ensure that model updates cannot reveal sensitive information [13].

3. IMPLEMENTATION AND RESULTS

The proposed federated learning system was implemented using distributed machine learning frameworks. Multiple simulated healthcare datasets were used to evaluate the performance of the system. The model was trained across multiple nodes representing different hospitals.

The results demonstrated that the federated learning model achieved prediction accuracy comparable to centralized machine learning approaches while substantially enhancing privacy protection [14].

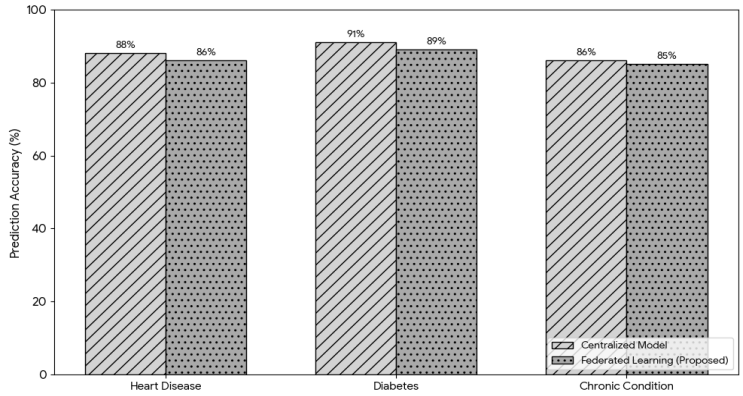

Chart -1: Prediction Accuracy Comparison

3.1 Performance Evaluation

The system was evaluated based on several performance metrics including prediction accuracy, data privacy level, and communication overhead. The federated learning model demonstrated strong performance across all metrics.

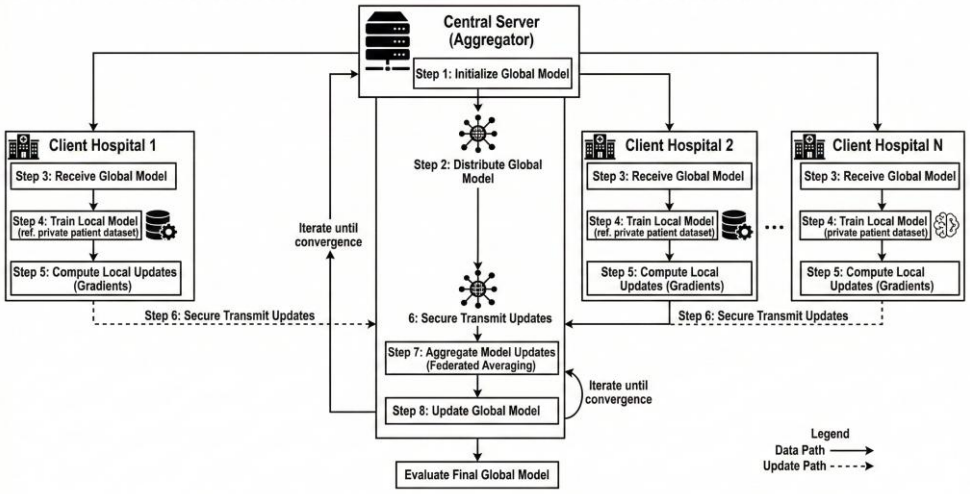

Fig -2: Model Training Process in Federated Learning

The results indicate that federated learning effectively enables collaborative healthcare prediction without compromising patient privacy.

4. CONCLUSIONS

This study presented a federated learning-based healthcare prediction system designed to protect patient privacy while enabling collaborative model training across multiple healthcare institutions. By allowing hospitals to train models locally and share only model parameters, the system eliminates the need for centralized data storage and significantly reduces privacy risks.

Experimental results demonstrate that federated learning achieves high prediction accuracy while maintaining strong privacy protection. Future research can focus on integrating advanced privacy mechanisms such as differential privacy, homomorphic encryption, and secure multiparty computation to further enhance system security [15].

5. REFERENCES

[1]. McMahan, B., et al., "Communication Efficient Learning of Deep Networks from Decentralized Data."
[2]. Kairouz, P., et al., "Advances and Open Problems in Federated Learning."
[3]. Yang, Q., Liu, Y., Chen, T., Tong, Y., "Federated Machine Learning: Concept and Applications."
[4]. Bonawitz, K., et al., "Practical Secure Aggregation for Privacy-Preserving Machine Learning."
[5]. Sheller, M. J., et al., "Federated Learning in Medicine: Facilitating Multi-Institutional Collaborations Without Sharing Patient Data."
[6]. Rieke, N., et al., "The Future of Digital Health with Federated Learning."
[7]. Esteva, A., et al., "A Guide to Deep Learning in Healthcare."
[8]. Price, W., Cohen, I., "Privacy in the Age of Medical Big Data."
[9]. Shickel, B., et al., "Deep Learning in Electronic Health Records."
[10]. Li, T., et al., "Federated Learning: Challenges, Methods, and Future Directions."
[11]. Kaissis, G., et al., "Secure, Privacy-Preserving and Federated Machine Learning in Medical Imaging."
[12]. Topol, E., "High-performance medicine: the convergence of human and artificial intelligence."
[13]. Geyer, R., Klein, T., Nabi, M., "Differentially Private Federated Learning."
[14]. Xu, J., et al., "Federated Learning for Healthcare Informatics."
[15]. Truex, S., et al., "Hybrid Federated Learning Approaches for Privacy-Preserving Data Analytics."