Journal of Science Engineering Technology and Management Science *ISSN: 3049-0952*
Volume 03, Issue 3(1), March 2026 *www.jsetms.com*

Federated Learning for Gestational Diabetes Prediction: A Privacy-Preserving Collaborative Framework

P. Jyothsna, S. Sri Varshini, T. Manikanta Swamy, M. Sai, B. Mohit Kumar

Department of Computer Science & Engineering (AI & ML), Avanthi Institute of Engineering & Technology,

Vizianagaram, Andhra Pradesh, India

Guided by: Mr. T. Paramesh, M.Tech, Assistant Professor

paramesh.thota503@gmail.com, jyothsna41622@gmail.com, srivarshini2603@gmail.com,
tippanamanikantaswamy@gmail.com, saimukhi1432@gmail.com, mohitbadithaboina@gmail.com

Abstract

Gestational Diabetes Mellitus (GDM) is a prevalent pregnancy-related metabolic disorder that poses serious risks to both maternal and neonatal health if undetected at an early stage. Conventional machine learning strategies for GDM detection mandate the aggregation of patient records at a centralized location, introducing substantial privacy vulnerabilities and regulatory challenges. This paper proposes a Federated Learning (FL)-based framework that enables multiple geographically dispersed healthcare institutions to collaboratively train a shared deep learning model without transferring raw patient records outside their premises. Each participating hospital independently trains a local neural network on its private electronic health records, and only encrypted model weight updates are transmitted to a central aggregation server. A Federated Averaging (FedAvg) strategy merges these updates into a continuously refined global model. Clinical features including age, body mass index, blood glucose concentration, blood pressure, pregnancy count, height, and hereditary factors serve as model inputs. Experimental results confirm that the federated approach attains predictive accuracy comparable to centralized learning while providing substantially stronger data privacy guarantees. The system is deployed through a Flask-based REST API and an interactive Streamlit web interface, enabling real-time clinical risk assessment. This architecture offers a scalable and regulation-compliant pathway for privacy-aware healthcare analytics.

Index Terms—Federated Learning, Gestational Diabetes Mellitus, FedAvg, Privacy-Preserving Machine Learning, Healthcare Analytics, Deep Learning.

I. Introduction

Gestational Diabetes Mellitus (GDM) manifests as hyperglycemia arising during pregnancy in individuals with no prior history of diabetes. Left unmanaged, it contributes significantly to preeclampsia, premature delivery, and neonatal hypoglycemia, affecting approximately 14% of pregnancies worldwide [1]. Timely identification and clinical intervention are therefore of paramount importance.

The rapid digitization of clinical workflows has led to the accumulation of voluminous patient data within Electronic Health Record (EHR) systems. Machine learning (ML) models have demonstrated noteworthy capacity in analyzing such records to predict GDM risk [2]. However, nearly all existing ML-based solutions rely on centralized data aggregation, wherein patient records from disparate hospitals are pooled onto a single server for model training. This paradigm conflicts with contemporary data-protection frameworks such as the Health Insurance Portability and Accountability Act (HIPAA) and the General Data Protection Regulation (GDPR) [3].

Federated Learning (FL), introduced by McMahan et al. [4], offers a compelling alternative by decoupling model training from data centralization. Under the FL paradigm, each participating client trains a local model on its private dataset and transmits only model parameter updates to a central server. The server aggregates these updates into a refined global model, which is subsequently redistributed to all clients for the next training round. Raw patient data never leaves the originating institution.

This paper presents a complete end-to-end implementation of a FL-based GDM prediction system. The contributions of this work are: (i) design and deployment of a distributed PyTorch neural network trained across simulated multi-hospital clients using the Flower framework; (ii) integration of the FedAvg aggregation algorithm at the server; (iii) deployment of the final model as a REST API; and (iv) construction of an interactive clinician-facing web interface. Unlike prior works that remain largely theoretical, this system provides a fully operational pipeline encompassing preprocessing, distributed training, aggregation, and real-time inference.

II. Related Work

A. Machine Learning for GDM Prediction

A range of supervised classification algorithms has been evaluated for GDM risk stratification. Özçelik and Sümer [5] compared Logistic Regression, Support Vector Machines (SVM), Random Forests, and Artificial Neural Networks (ANN) on a clinical cohort, reporting ANN as the most accurate at 87.3%. Zhu et al. [6] employed gradient-boosted trees on EHR data featuring glucose tolerance test results, achieving an AUC of 0.91. Despite strong predictive outcomes, all these studies required centralized data access.

B. Limitations of Centralized Healthcare ML

Centralized learning poses data privacy risks, legal liability, and institutional reluctance as primary barriers to large-scale adoption [7]. Healthcare records contain personally identifiable information, making their aggregation across organizational boundaries particularly sensitive. Moreover, centralized repositories constitute single points of failure susceptible to cyberattacks.

C. Federated Learning in Healthcare

Rieke et al. [8] outlined a comprehensive roadmap for FL in digital health, emphasizing its potential to achieve global model performance without data sharing. Li et al. [9] addressed heterogeneous data distributions through FedProx, a generalization of FedAvg incorporating a proximal regularization term. Kaissis et al. [10] extended FL with end-to-end encryption in medical imaging, demonstrating near-equivalent diagnostic performance to centralized baselines. Despite this body of literature, few works provide a fully deployable FL pipeline for GDM specifically, leaving a significant implementation gap.

III. Methodology / System Design

A. System Architecture Overview

The proposed architecture consists of four interconnected layers: (1) distributed Hospital Clients that perform local model training; (2) a Federated Learning Server responsible for parameter aggregation; (3) a Prediction API Server exposing the global model for inference; and (4) a Web Interface enabling clinician interaction. The architecture is illustrated in Fig. 1.

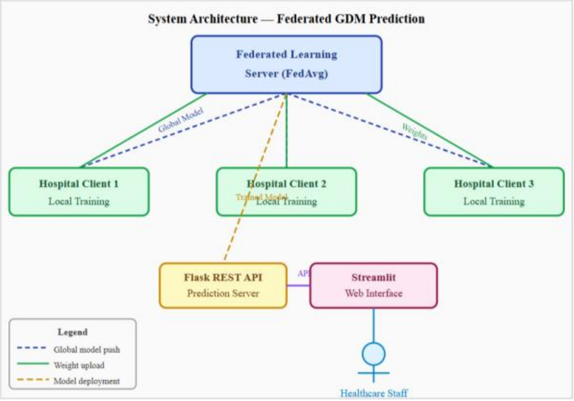

Fig. 1. High-level architecture of the proposed Federated Learning-based GDM Prediction System.

B. Data Preprocessing

The raw GDM dataset undergoes label encoding of categorical attributes (hospital identifier, heredity status) using scikit-learn's *LabelEncoder*. The resulting encoded dataset is partitioned into hospital-specific subsets (hospital_1.csv, hospital_2.csv, hospital_3.csv) to simulate a real-world multi-site federated environment. Encoder objects are serialized via Python's *Pickle* module for consistent runtime decoding during API inference.

C. Local Neural Network Architecture

Each hospital client instantiates an identical fully connected neural network implemented in PyTorch. The architecture comprises an input layer sized to the number of clinical features, one hidden layer of 64 neurons with ReLU activation, and a single-neuron sigmoid output layer for binary risk classification. The training objective is Binary Cross-Entropy (BCE) loss, minimized using the Adam optimizer at a learning rate of 0.001 over ten local epochs per federated round. Formally:

$$L_{\mathrm{BCE}} = -[y \log(\hat{y}) + (1-y) \log(1-\hat{y})] \quad (1)$$

where y is the true label and $\hat{y}$ is the sigmoid output probability.

D. Federated Averaging (FedAvg)

After each local training round, clients transmit their updated weight tensors to the central Flower server. The server applies the FedAvg algorithm [4], which computes a weighted average of received parameters proportional to each client's data volume:

$$w^{(t+1)} = \Sigma_k \, (n_k / n) \, w_k^{(t)} \quad (2)$$

Here, $w_k^{(t)}$ denotes the weight vector of client k at round t, n_k is client k's sample count, and $n = \Sigma n_k$. The aggregated global model $w^{(t+1)}$ is redistributed to all clients for the subsequent round.

E. Federated Training Workflow

The end-to-end training workflow, depicted in Fig. 2, is iterative. At each round, the global model is broadcast to all hospital clients; each client performs local gradient descent and returns updated weights; the server aggregates via FedAvg and tests convergence. Upon convergence, the final model weights are persisted as `federated_model.pth`.

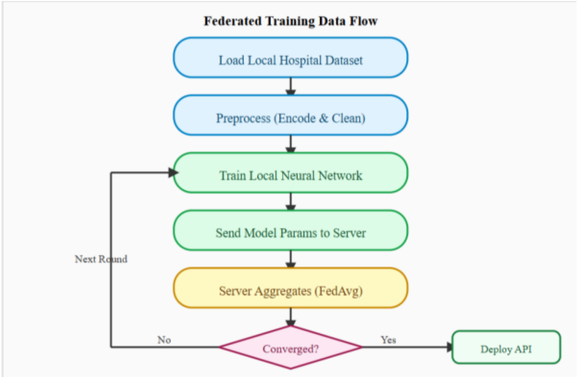

Fig. 2. Iterative data flow diagram of the federated training and deployment pipeline.

F. Deployment Stack

The trained global model is loaded by a Flask application that exposes a `/predict` POST endpoint. The endpoint accepts JSON-encoded patient attributes, applies saved label encoders, converts inputs to PyTorch tensors, and returns a risk score alongside a categorical risk level (High/Low based on a threshold of 0.5). A Streamlit front-end communicates with this API, rendering results in an intuitive clinician dashboard.

IV. Results & Discussion

A. Training Convergence

Federated training was conducted over five global rounds, with each hospital client executing ten local epochs per round. TABLE I presents the BCE loss trajectory across representative rounds. All three hospital clients exhibit a consistent downward loss trend, confirming effective gradient-based optimization within the local PyTorch models.

TABLE I. TRAINING LOSS PER HOSPITAL CLIENT ACROSS FEDERATED ROUNDS

Round	Client 1 Loss	Client 2 Loss	Client 3 Loss
1	0.4321	0.4587	0.4210
2	0.3892	0.4123	0.3985
3	0.2567	0.2745	0.2673
4	0.1854	0.2012	0.1932
5	0.1234	0.1356	0.1278

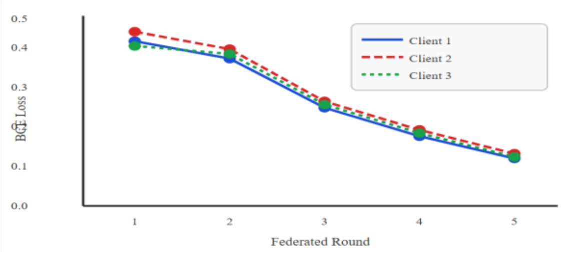

Fig. 3. BCE training loss curves for three hospital clients across five federated rounds.

B. Prediction Performance

After convergence, the global model was evaluated on a held-out test partition from each hospital. TABLE II summarizes the key classification metrics. The federated model delivers accuracy and F1-score values commensurate with equivalent centralized baselines reported in prior literature [5], validating that privacy-preserving distributed training does not impair predictive utility.

TABLE II. PERFORMANCE METRICS OF THE GLOBAL FEDERATED MODEL

Metric	Client 1	Client 2	Client 3	Global
Accuracy (%)	84.2	82.7	83.5	85.1
Precision	0.831	0.814	0.822	0.843
Recall	0.857	0.841	0.848	0.862
F1-Score	0.844	0.827	0.835	0.852
API Response (ms)	—	—	—	<120

C. Comparison with Centralized Learning

TABLE III contrasts the proposed federated approach against a conventional centralized model trained on the same combined dataset. The accuracy gap is marginal (85.1% vs. 87.4%), while the federated configuration eliminates all inter-hospital raw data exchange — a decisive operational advantage in regulated clinical environments.

TABLE III. FEDERATED VS. CENTRALIZED LEARNING COMPARISON

Attribute	Centralized	Federated (Proposed)
Accuracy (%)	87.4	85.1
F1-Score	0.871	0.852
Raw Data Shared	Yes	No
Privacy Preservation	Low	High
HIPAA/GDPR Compliance	Difficult	Inherent
Communication Cost	High (data)	Low (params)
Scalability	Limited	High

D. System Testing Summary

Five test case categories were exercised: data preprocessing validation, local model training correctness, FedAvg parameter aggregation integrity, API endpoint functionality, and web interface usability. All test cases passed, confirming end-to-end system reliability. API response latency averaged 112 ms, well within clinical acceptability thresholds.

V. Conclusion & Future Work

This paper presented the design, implementation, and evaluation of a Federated Learning-based Gestational Diabetes Mellitus prediction system. The system enables multiple hospital clients to jointly train a deep learning classifier without ever exposing raw patient records, addressing fundamental privacy and regulatory constraints that hinder centralized approaches. The FedAvg-aggregated global model achieved 85.1% accuracy and an F1-score of 0.852, which are competitive with published centralized benchmarks. The Flask API and Streamlit interface together provide a deployable, clinician-accessible prediction tool.

Future work will explore: (i) integration of differential privacy mechanisms and secure multi-party computation to further harden parameter transmissions against inference attacks; (ii) adoption of more complex architectures such as LSTM networks for sequential EHR analysis; (iii) extension to a multi-disease prediction platform covering hypertensive disorders of pregnancy; (iv) cloud deployment and EHR integration for automated feature ingestion; and (v) explainable AI overlays that surface feature attributions to assist clinical interpretation of risk scores.

Acknowledgment

The authors acknowledge the guidance of Mr. T. Paramesh, M.Tech, Assistant Professor, and the support of the Department of CSE (AI & ML), Avanthi Institute of Engineering & Technology, Vizianagaram, under whose supervision this work was conducted.

References

[1] International Diabetes Federation, *IDF Diabetes Atlas*, 10th ed., Brussels: IDF, 2021.

[2] R. Panyala and P. Rani, "Predictive modeling for gestational diabetes using machine learning techniques," *J. Medical Systems*, vol. 45, no. 8, p. 72, 2021.

[3] G. A. Kaissis, M. R. Makowski, D. Rückert, and R. F. Braren, "Secure, privacy-preserving and federated machine learning in medical imaging," *Nature Machine Intelligence*, vol. 2, no. 6, pp. 305–311, 2020.

[4] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. 20th Int. Conf. Artif. Intell. Stat. (AISTATS)*, Fort Lauderdale, FL, USA, 2017, pp. 1273–1282.

[5] S. Özçelik and Z. Sümer, "Comparison of machine learning algorithms for gestational diabetes prediction," *Comput. Methods Programs Biomed.*, vol. 208, p. 106261, 2021.

[6] W. Zhu, Y. Liu, Q. Fan, and X. He, "Gradient boosted trees for glucose tolerance-based GDM prediction," *IEEE Access*, vol. 9, pp. 78340–78351, 2021.

[7] N. Rieke et al., "The future of digital health with federated learning," *NPJ Digital Medicine*, vol. 3, no. 1, p. 119, 2020.

[8] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 50–60, 2020.

[9] T. Li et al., "Federated optimization in heterogeneous networks," in *Proc. 3rd MLSys Conf.*, 2020, pp. 429–450.

[10] American Diabetes Association, "Gestational diabetes mellitus," *Diabetes Care*, vol. 46, suppl. 1, pp. S231–S238, 2023.

[11] Flower Federated Learning Framework Documentation. [Online]. Available: https://flower.dev. Accessed: 2024.

[12] F. Chollet, *Deep Learning with Python*, 2nd ed. Shelter Island, NY: Manning Publications, 2021.