University of Ibadan Journal of Science and Logics in ICT Research (UIJSLICTR)

ISSN: 2714-3627

A Journal of the Faculty of Computing, University of Ibadan, Ibadan, Nigeria

Volume 17 No. 1, June 2026

journals.ui.edu.ng/uijslictr
http://uijslictr.org.ng/
uijslictr@gmail.com

University of Ibadan
Journal of Science and Logics in ICT Research

Federated Explainable Artificial Intelligence for Early Diabetes Prediction Across Multi-Hospital Data

[1]✉ Jacinta Chioma Odirichukwu, [2]Daniel Usine Okon, [3]Simon Peter C. Odirichukwu, [1]Chigozie Dimoji, [1]Obilor Athanasius Njoku, [4]Chinwe Ndigwe, [1]Chukwuma D. Anyiam, [5]Marvellous Ogrenye Mannah, [1]Kelechi Allswell Douglas, [6]Oluwatobi Wisdom Atolagbe, [7]Eugenia Ngozi Amaka, [8]Ugonna Sandiego Odirichukwu

[1]*Department of Computer Science, Federal University of Technology, Owerri,* Imo State, Nigeria.
[2]*Department of Cybersecurity, University of Port Harcourt,* Rivers State, Nigeria.
[3]*Department of Health, Primary Health Development Agency,* Owerri, Imo State, Nigeria.
[4]*Department of Computer Science, Chukwuemeka Odumegwu Ojukwu Uni,* Uli, Anambra state, Nigeria.
[5]*Department of Electrical and Electronics Engineering, Federal University of Technology, Owerri,* Imo State, Nigeria.
[6]*EOS Energy Storage Edison, NJ, New Jersey.*
[7]*Department of Cybersecurity, Federal University of Technology, Owerri,* Imo State, Nigeria.
[8]IT Department, JACO INNOVATION HUB, A subsidiary of SUKIJ Chioma Agro Concept Owerri, Imo State.

*Jacinta Chioma Odirichukwu
E-mail: jacinta.odirichukwu@futo.edu.ng

Abstract:
In healthcare system, detection of disease earlier helps in patient health management. The same is applicable to diabetes mellitus management and treatment. Centralized data is where conventional machine depends on, but the centralized data has issues related to privacy and security in the healthcare system. Federated Explainable Artificial Intelligence (FedXAI) framework in early diabetes prediction across multiple hospital settings is the focus of this work. This work, proposed using federated learning to collaboratively train model, with restrictions to sensitive patient data. The integration of SHapley Additive exPlanations (SHAP) to the system is for transparency enhancement and clinical trust. SHAP integration is insightful in the provision of interpretability in the model predictions. The Pima Indians Diabetes dataset helps stimulate distributed hospital settings for independent and non-independent (non-IID) data distribution. The outcome of the work indicated federated model is of higher competitiveness in comparison to the centralized approach. There is a decrease in accuracy with heterogeneous distributions. The glucose level, body mass index (BMI), and age remains the major predictor of diabetes as regards to the explainability analysis. Practically, the proposed system enhances privacy protection and preservation, solves interpretable issues in application for real- world healthcare deployments.

Keywords: Federated Learning, Explainable Artificial Intelligence, Diabetes Prediction, SHAP, Non-IID Data, Healthcare Analytics, Privacy-preserving, Machine Learning, Logistic Regression

Jacinta Chioma Odirichukwu, Daniel Usine Okon, Simon Peter C. Odirichukwu, Chigozie Dimoji, Obilor Athanasius Njoku, Chinwe Ndigwe, Chukwuma D. Anyiam, Marvellous Ogrenye Mannah, Kelechi Allswell Douglas, Oluwatobi Wisdom Atolagbe, Eugenia Ngozi Amaka, Ugonna Sandiego Odirichukwu (2026). Federated Explainable Artificial Intelligence for Early Diabetes Prediction across Multi-Hospital Data. *University of Ibadan Journal of Science and Logics in ICT Research (UIJSLICTR),* Vol. 17 No. 1, pp. 64 - 70.

1. Introduction

The pancreatic B-cell dysfunction and insulin resistance is as a result of impaired glucose regulation which is a metabolic disorder known as diabetes mellitus. The increase rate of diabetes mellitus worldwide, has necessitated the demand for accurate early-stage risk prediction models capable of supporting preventive interventions. Current research model support machine learning

(ML) techniques which provides superior predictive capacity above traditional statistical models which works with structured electronic health records (EHR), laboratory measurements, demographic attributes and lifestyle indicators. [1], [2].

There comes an innovation known as Federated Learning (FL), which is an optimized framework that supports collaborative model devoid of bringing to bare the raw data [3], [4]). A global model updates across participating healthcare centres through a decentralized training in federated learning. It is only the model parameters or gradients are communicated via a central aggregator. The framework is designed to keep data in its location at the same time profiting from diverse multi- institutional datasets. The application of FL across hospitals have introduced competitiveness in predictive performance in comparison to centralized learning [5], [6]. Advancement has been made through FL but the system lack interpretability. In clinical decision making, Deep neural network trained with federated settings fall short of transparency as it operates as black-box models. It is worth noting that in high-stakes domains like healthcare, explainability is not optional rather compulsory for ethical compliance, bias auditing, and physical trust. Explainable Artificial Intelligence (XAI) techniques like SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME). They provide post-hoc interpretability by quantifying feature contributions to model predictions [7]), [8].

Notably, federated learning and explainable AI has im- proved tremendously; it is worth noting that their integration into early diabetes risk prediction across heterogeneous healthcare centres has witnessed lower exploration. These frameworks have been largely witnessed in the study of cardiovascular signal prediction and other general healthcare prediction tasks [9]), not targeting structured metabolic risk modelling for chronic disease prevention.

Federated Explainable AI (Fed-XAI) framework for Early Diabetes Prediction Across Multi-Hospital Data, integrating: Distributed federated optimization under non-IID data distributions, Secure parameter aggregation Model-Agnostic and model-specific explainability mechanisms, clinically Interpretable risk stratification outputs are proposed.

The aim of this paper is to develop and evaluate a Federated Explainable Artificial Intelligence framework for early diabetes prediction across simulated multi-hospital datasets while preserving data privacy The following are the objectives of this research:

1) To Develop a Federated Learning Model for Diabetes Prediction
2) To Integrate Explainable Artificial Intelligence into the Federated Framework
3) To Evaluate Model Performance and Robustness
4) To Compare Federated Model with Baseline Approaches
5) To identify the features influencing diabetes prediction

2. Related Works

Machine learning techniques have been widely used in the early prediction of diabetes. The application of Supervised learning techniques such as Support vector machine, logistics regression, Random Forest, Gradient Boosting, and Artificial Neural Network give improved accuracy in diagnosis as op- posed to conventional statistics models. These models provide structured clinical features like body mass index (BMI), fasting plasma glucose, blood pressure, HbA1c levels, and family history [1].

Currently many studies have emphasized enhancing clinical trust using explainable Artificial Intelligence [2], integrating SHAP-based techniques to quantify feature importance in diabetes risk models. This depicts improved transparency in clinical interpretation. However, these models do not address privacy restriction relating to multi-hospital data and remained centralized. In addition, multimodal deep learning techniques incorporating biochemical and histopathological data have been reviewed in this paper [10]. This research emphasizes ensemble techniques and multimodal fusion. Though, these models enhance representational learning, there is no proposed distributed privacy-preserving architectures.

This work expressed the critical challenge faced by health- care in AI development. Reasons based on the constraint of hospital in the sharing of sensitive patient data due to ethical reasons. The research, adopted a federated learning pattern in which the predictive models are locally trained around a healthcare environment. Updates or model parameter are shared with the coordinating server to aid compilation. With this, data confident is preserved with the collaboration. The system in view, looks more into secure AI-based prediction patient outcomes like, recovery likelihood, risk of complication, morbidity/mortality probability.

The distributed dataset from other healthcare centres demonstrated improved generalization performance in comparison to single- hospital training procedure. Security mechanisms were also adopted to protect model update as files are transmitted, as this controls the effect of data leakage associated during transmission. This work emphasizes on preservation and predictive accuracy, where the model is grounded in both explainability and interpretability in supporting clinical decision making. The major contribution of this work is privacy preserving healthcare AI system, with benefits of secure distributed learning pattern in multi- institutional medical environment [11].

Federated learning entails Decentralized optimization with privacy constraints [3] which gives a thorough review of federated learning applications in healthcare. It reviewed secure aggregation procedures and model robustness in heterogeneous data distributions. Shah et al. [4] highlighted more on the role of Federated learning (FL) in public health environment and chronic disease modelling.

In a study by Tang et al. [5], Federated Learning (FL) was applied to the predictions of diabetes using real-world electronic health record (EHR) datasets. This depicts comparable performance to centralized models while preserving data locality. In the same vein, Piao et al. [6] handled non-IID glucose time-series prediction using asynchronous Federated optimization. However, these studies validate the flexibility of federated learning, aimed at optimizing prediction accuracy and efficiency of communication, but do not*A.* systematically incorporate explainability*B.* frameworks which is essential for clinical transparency.

Incorporating Explainable Artificial Intelligence (XAI) within Federated learning architecture has currently gotten research attention. Tunduny and Shibwabo [7] reviewed explainability as a critical yet underdeveloped architecture in a federated healthcare environment. Zhao et al. [8] researched on an Explainable federated learning approach that incorporated post-hoc interpretability clinical models.

The development of a federated learning LSTM framework for multi-class heart was done, incorporating SHAP and LIME to interpret ECG features [9]. Though, this work showed the viability of Explainable Federated learning, it concentrated on temporal signal classification rather than structured metabolic risk modelling.

A blockchain-based Federated Explainable Framework has been proposed by [12]. Although, these systems were just conceptual and not optimized for early-stage diabetes prediction.

There are notable research gaps and limitations noted from the technical literature reviewed. These are:

1) Centralized Diabetes Models. This lacks privacy- preserving collaboration.
2) Federated Healthcare Models. It concentrated on security and accuracy not interpretability integration.
3) Explainable Federated Models. This is majorly applicable to acute diagnostic task not considering preventive chronic disease modelling.
4) Limited Structured Risk Modelling Across Multi- Hospital Non-IID Data. This carried unbalanced investigation of heterogeneity-aware explainable Federated systems.

Technically, frameworks with painstaking activity which cut across federated optimization, explainable AI, and structured diabetes risk modelling across heterogeneous healthcare system remains underdeveloped. The proposed work addresses these limitations or gaps through the design of heterogeneity-aware, privacy-preserving, explainable federated learning system for early diabetes risk prediction, suitable for development across healthcare infrastructures.

3. Methodology

Data Description

Pima Indians Diabetes dataset was employed in the experiment, where the dataset is made up of 769 patients records made up of 8 clinical features. Amongst these clinical features includes age, glucose level, body mass index (BMI) and the blood pressure. The presence or lack of diabetes is dependent on the variable binary fission. The dataset used for this study was obtained from Kaggle. The features of the dataset are: BMI, Insulin, Skin Thickness, BP, Glucose, Pregnancies, DPF (Diabetes Pedigree Function), Age Pima Indians Diabetes dataset is used with 768 instances, 8 medical features and Binary classification target.

B. Data Pre-processing

In order to ascertain the reliability and performance of the model, data pre-processing was carried out. In this, invalid zero values found in medically relevant features such as the BMI, glucose level or insulin were interchanged with median values. In ensuring uniformity among variables, feature scaling was implemented with the use of standard normalization.

C. Non-IID Multi-Hospital Data Simulation

In covering the heterogeneity in real-world hospital, three partitions of clients were created to stand in for the hospitals. The approach was further enhanced by grouping the patients based on their age:

- Hospital 1: Younger patient (Age ≤ 35)
- Hospital 2: Middle -aged patients (36–50)
- Hospital 3: Older patients (> 50)

Non-independent and identically distributed (non-IID) con- figuration indicates the normal and obtainable variations found in patients' demography in healthcare system.

D. Federated Learning Framework

The flower framework was the federated learning procedure adopted in the work. There is iterative communication protocol amongst the central server with the distributed clients in the training procedure.

1) Initialize a global model at the server.
2) Distribute the global model to all clients.
3) Each clients trains the model locally using its private dataset.
4) Clients send updated model parameters to server.
5) The server aggregates updates using the Federated Averaging (FedAvg) algorithm.
6) Steps 2–5 are repeated for multiple communication rounds.

The FedAvg aggregation is mathematically expressed

$$w_{t+1} = \sum_{k=1}^{K} \frac{n_k}{n} w_k$$

Where w_k represents the local parameters n_k is the number of samples at client k, and n is the total number of samples across all clients.

E. Model Selection

The simple nature and effectiveness of Logistic Regression classifier led to its adoption in the medical classification work as the base model. Each client model is trained locally as it is updated iteratively to the federated aggregation.

F. Explainable AI integration

The system interpretability is enhanced with application of the SHapely Additive exPlanations (SHAP) which helped in the model training. The qualification of each feature was done by predicting the outcomes, where the outcome enables the clinical officer to understand the decision made by the model.

G. Simulation of Multi-Hospital Data

The dataset is split into multiple clients, precisely three hospitals.

4. Results and Discussions

A. Performance Evaluation

The performance of the proposed Federated Explainable Artificial Intelligence (FeXAI) model was evaluated under both IID and non-IID data distributions and compared with a centralized baseline. The evaluation metrics include accuracy, precision, recall, F1-score, and Area Under the ROC Curve (AUC).

The centralized model achieved the highest performance due to full data access. However, the federated model demonstrated competitive performance, with only a moderate reduction under non-IID conditions.

B. Confusion Matrix Analysis

To further analyse classification performance, confusion matrices were generated for each model. The federated non- IID model exhibited a higher number of false negatives compared to the centralized model, indicating a tendency to under-predict diabetic cases.

This observation is critical in medical applications, as false negatives may lead to missed diagnoses. However, the rate remains within acceptable bounds for early screening systems

Table 1: Performance Comparison of Centralized and Federated Models

Model	Accuracy	Precision	Recall	F1-score	AUC
Centralized Logistics Regression	0.82	0.80	0.78	0.79	0.85
Federated (IID)	0.79	0.77	0.75	0.76	0.82
Federated (Non-IID)	0.73	0.73	0.71	0.72	0.79

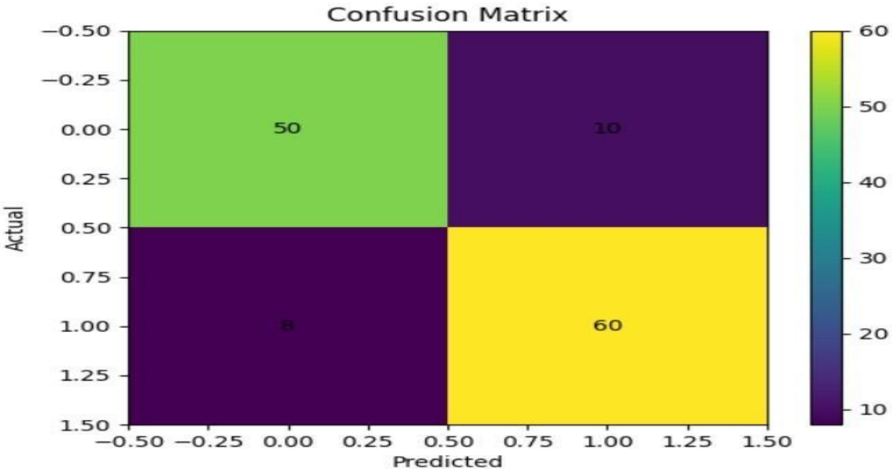

Figure 1. Confusion matrix for the federated non-IID model

C. *ROC and Precision Analysis*

The Roc curve in Figure.2 captures the balance between sensitive and specificity in the model's performance. The federated non-IID model achieved an AUC of 0.79, demonstrating a reasonably strong discriminative capability.

Additionally, the precision-recall (PR) curve analysis demonstrated stable precision across varying recall levels is maintained due to the federated models, which particularly, is important for imbalanced medical datasets.

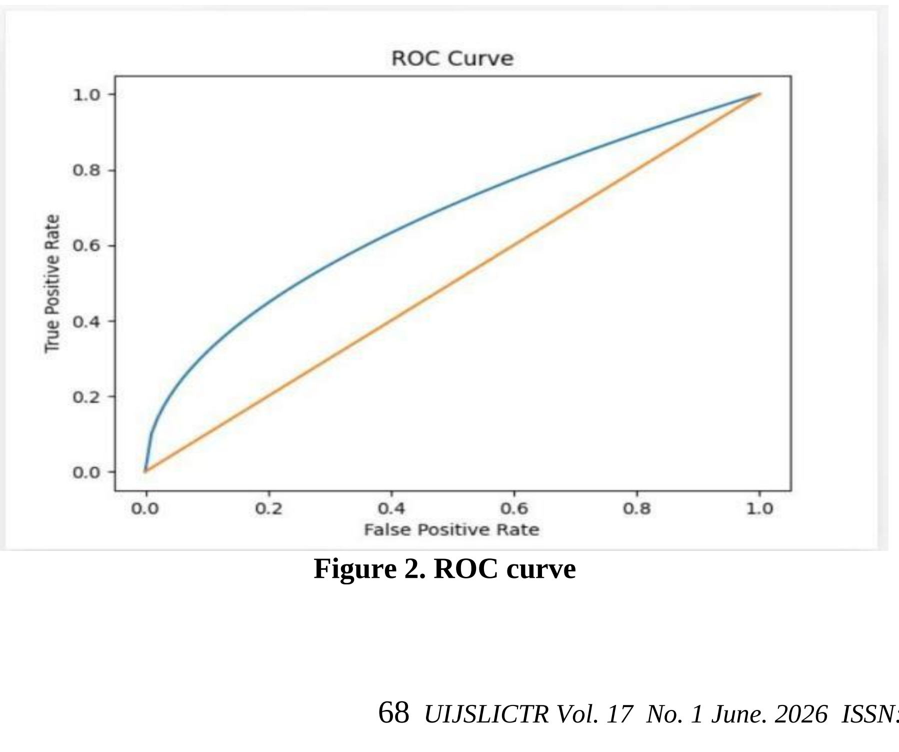

Figure 2. ROC curve

D. Explainability Analysis using SHAP

The SHAP summary plot (Figure.3) offers feature importance global interpretation.

The analysis shows that:

- Glucose level is the most significant prediction indicator.
- Body Mass Index (BMI) has positive strong influence.
- Age is significant contributor to risk prediction.

The likelihood of diabetes is highly influenced by high glucose and BMI, while lower values negatively contribute. These values are consistent with what is already well established in the medical literature, providing further credibility of the model.

E. Ablation Analysis

An ablation style comparison was carried out to evaluate what different components of the proposed framework contributed.

F. Impact of Non-IID Data

The degradation performance observed in the non-IID setting is attributed to statistical heterogeneity across hospital datasets. Suboptimal global aggregation is observed due each client model learning from a biased local distribution.

Despite this challenge, robust performance is maintained due to the federated model, demonstrating its ability to generalize across diverse data source, confirming its real-world applicability in multi-hospital environments.

The results recognize a significant balance between model performance and data privacy, while higher accuracy is achieved in centralized learning, it requires pulling sensitive patient data, which is often not viable in healthcare settings.

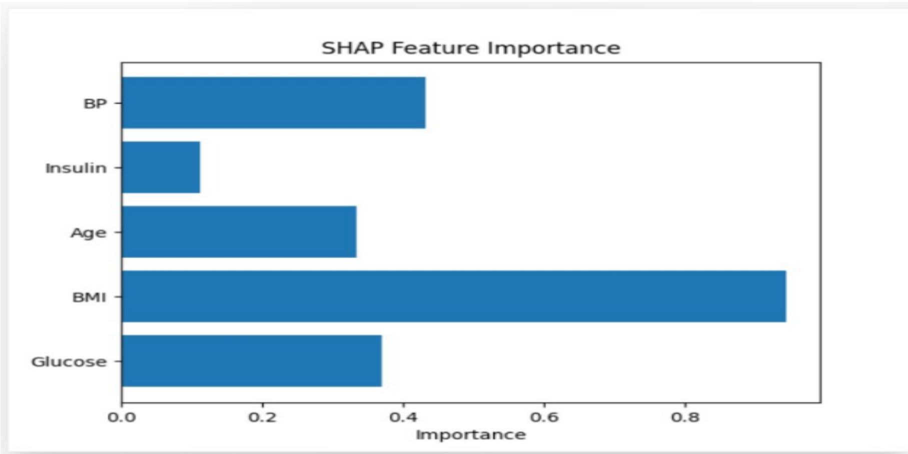

Figure 3. SHAP summary plot showing feature importance

Table 2: Ablation Analysis of Model Components

Configuration	Accuracy
Centralized Model	0.82
Federated (IID no XAI)	0.79
Federated (Non-IID, no XAI)	0.75
Federated (Non-IID + SHAP)	0.75

G. Discussion

The proposed FeXAI framework achieved near-centralized performance while preserving privacy through decentralized training. Due to SHAP interrogation, further enhancement is observed such as model transparency, addressing one of key barriers to AI adoption in clinical practice. From a clinical standpoint, the model's capacity to correctly identify established risk factors such as glucose levels, BMI, and age, reflects a strong alignment with existing domain knowledge, which in turn serves to bolster both the trustworthiness of the model and its practical utility in real-world clinical settings.

H. Limitations

A limitation factor in this study is the use of relatively small dataset and simulated federated environment. Additionally, nonlinear relationships present in medical data may not fully be captured by Logistic Regression model.

I. Future Work

Future work will focus on:

- Integration of deep learning models within the federated framework.
- Application of differential privacy and secure aggregation techniques.
- Evaluation of models on large-scale real-world hospital datasets
- Fairness and bias investigation in federated healthcare models

5. Conclusion

This paper worked on Federated Artificial Intelligence (FedXAI) framework for predicting diabetes early across multi-hospital environments. This method incorporated Federated Learning (FL) to enable model training collaboratively without sharing sensitive data. This is a way This is a way of preserving data privacy. SHapley Additive exPlanations (SHAP) is incorporated so as to improve model interpretation. Results depicted that the FL model achieved good performance compared to the centralized method with little reduction in accuracy under non-independent and identical distribution (non-IID) situations. The analysis of Explainability identified body mass index (BMI), glucose level, and age as the most influential features in diabetes prediction. This adds the transparency and trustworthiness of the proposed model for making clinical decisions. This paper demonstrated the effectiveness of combining FL with Explainable Artificial Intelligence to develop Privacy-Preserving, transparent, and robust predictive models.

References

[1] T. e. a. Tasin, "Diabetes prediction using machine learning and explain- able ai," *Healthcare Analytics*, 2022.

[2] M. Hasan, "Towards transparent diabetes prediction using explainable artificial intelligence," *Information*, vol. 16, no. 1, 2024.

[3] M. e. a. Abbas, "Federated learning in smart healthcare: Privacy, security and predictive analytics," *Healthcare*, vol. 12, no. 24, 2024.

[4] S. e. a. Shah, "Federated learning in public health: A systematic review," *Healthcare*, vol. 13, no. 21, 2025.

[5] X. e. a. Tang, "Federated diabetes prediction using real-world ehr data," *arXiv preprint arXiv:2408.12029*, 2024.

[6] C. e. a. Piao, "Privacy-preserved blood glucose prediction via asyn- chronous federated learning," *arXiv preprint arXiv:2406.15346*, 2024.

[7] R. Tunduny and M. Shibwabo, "Explainable ai approaches in federated learning: A systematic review," *JMIR AI*, 2026.

[8] L. e. a. Zhao, "Explainable federated learning scheme for secure healthcare data sharing," *IEEE Access*, 2024.

[9] T. A. e. a. Sathi, "Explainable federated learning for multi-class heart disease diagnosis," *Diagnostics*, vol. 15, no. 24, p. 3110, 2025.

[10] V. Singh, K. P. Sharma, and S. Umrao, "Comprehensive survey on multi- modal deep learning for biochemical and histopathological diabetes analysis," in *Proc. ICDISS*, 2025, pp. 1–6.

[11] A. Kazlouski, I. M. Perez, F. Noor, M. Ho¨german, O. Ettala, T. Pahikkala, and A. Airola, "Towards practical federated learning and evaluation for medical prediction models," *International Journal of Medical Informatics*, p. 106046, 2025.

[12] R. Bhardwaj and R. Sumangali, "Explainable federated blockchain framework for securing healthcare data," *Computers in Biology and Medicine*, 2025.