International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0158
e ISSN: 2584-2854
Volume: 04 Issue: 04
April 2026
Page No: 1031 - 1037

Privacy Preserving Multi Disease Prediction Framework in Healthcare Using Federated Learning

Sowmiya J[1],
[1]*PG - Data Science, K.S.Rangasamy College of Technology, Tiruchengode - 637215, Tamil Nadu, India*
Email ID: *sow*miyajayakumar18@gmail.com[1]

Abstract
The rapid adoption of machine learning in healthcare has enhanced disease prediction capabilities. However, traditional centralized approaches pose significant risks to patient data privacy and security. To overcome these limitations, this paper presents a privacy-preserving multi-disease prediction framework based on Federated Learning, a decentralized approach that enables collaborative model training without sharing raw data. The proposed system supports the prediction of multiple diseases, including heart disease, lung disease, Parkinson's disease, and diabetes. In this framework, datasets are distributed across multiple simulated hospital clients, where local models are trained independently using Logistic Regression and Neural Networks. The locally trained model parameters are securely aggregated at a central federated server using the Federated Averaging algorithm to generate a global model. Performance evaluation is conducted using metrics such as accuracy, precision, recall, and F1-score, and the model with superior performance is selected for final prediction. The system is integrated into a Streamlit-based web application to provide real-time, user-friendly disease prediction. Experimental results demonstrate that the proposed approach achieves high predictive accuracy while ensuring data privacy, scalability, and efficient model collaboration. This work highlights the potential of Federated Learning as a secure and practical solution for multi-disease prediction in modern healthcare systems.

Keywords: *Disease prediction; Federated learning; Healthcare analytics;Neural networks; Privacy-preserving machine learning.*

1. Introduction
The increasing prevalence of chronic diseases such as heart disease, lung disease, Parkinson's disease, and diabetes has become a major concern in global healthcare. Early and accurate prediction of these diseases is essential for effective treatment and reducing mortality rates. Traditional diagnostic systems often rely on centralized data collection, where patient data is stored and processed in a single location. However, such approaches raise serious concerns related to data privacy, security, and regulatory compliance, especially when dealing with sensitive medical information (Kairouz et al., 2021; Li et al., 2020; Sheller et al., 2020; Alexander, 2025).Machine learning techniques have shown significant potential in improving disease prediction by analyzing large-scale medical datasets. Models such as Logistic Regression and Neural Networks are widely used due to their effectiveness in classification tasks. Logistic Regression provides a simple and interpretable approach, whereas Neural Networks are capable of capturing complex nonlinear relationships in data. However, most existing studies focus on single-disease prediction and depend on centralized architectures, limiting their applicability in real-world healthcare environments.To address these challenges, Federated Learning has emerged as a decentralized approach that enables collaborative model training without sharing raw data. In this approach, data remains locally within institutions, and only model updates are shared, ensuring privacy preservation (McMahan et al., 2017; Yang et al., 2019). Despite its advantages, limited research has explored the integration of federated learning with multi-disease prediction systems.This paper proposes a privacy-preserving multi-disease prediction framework using Federated Learning. The system predicts heart disease, lung disease, Parkinson's disease, and diabetes using Logistic Regression and Neural Network models. The models are evaluated

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0158
e ISSN: 2584-2854
Volume: 04 Issue: 04
April 2026
Page No: 1031 - 1037

based on accuracy, and the best performing model is selected for deployment. The proposed approach ensures data privacy while providing accurate and scalable disease prediction.

1.1.Problem Statement

Existing disease prediction systems primarily focus on single diseases and rely on centralized data storage, which poses significant risks related to data privacy and security. Additionally, there is a lack of unified frameworks that support multiple disease prediction while maintaining high accuracy and privacy.

1.2.Objective of the Study

The objective of this work is to develop a privacy-preserving multi-disease prediction system using Federated Learning. The study aims to compare Logistic Regression and Neural Network models based on accuracy and identify the most effective model for predicting heart disease, lung disease, Parkinson's disease, and diabetes.

2. Literature Review

Recent advancements in machine learning and healthcare analytics have significantly improved disease prediction systems. Traditional centralized machine learning approaches require data from multiple sources to be collected and stored in a central server, which introduces challenges related to data privacy, security, and regulatory compliance.Federated Learning (FL) has emerged as a decentralized learning paradigm that enables collaborative model training without sharing raw data. Kairouz et al. (2021) provided a comprehensive survey on federated learning, highlighting its potential in privacy-sensitive applications such as healthcare. The study discusses challenges such as communication efficiency, data heterogeneity, and system scalability.Li et al. (2020) explored optimization techniques in federated learning and proposed methods to improve convergence in distributed environments. Their work forms the foundation for federated averaging algorithms used in many practical implementations.In the healthcare domain, Sheller et al. (2020) demonstrated the effectiveness of federated learning in medical data analysis, emphasizing its ability to preserve patient privacy while enabling collaborative learning across institutions. Their research highlights the applicability of FL in disease prediction tasks.Several studies have focused on individual disease prediction using machine learning techniques. For instance, logistic regression and neural networks have been widely applied for heart disease, diabetes, and Parkinson's disease prediction due to their ability to model linear and nonlinear relationships effectively. However, most existing systems are limited to single-disease prediction and rely on centralized data processing. Recent work by Alexander (2025) highlights the importance of integrating federated learning with healthcare systems to enable secure and scalable multi-institutional collaboration.Despite these advancements, there is limited research on unified multi-disease prediction frameworks that combine federated learning with real-time deployment. This gap motivates the development of the proposed system, which integrates multiple disease prediction models within a privacy-Preserving Federated Learning Framework.

3. Methodology

3.1.System Overview

The proposed system is a privacy-preserving multi-disease prediction framework based on Federated Learning. It enables collaborative model training across multiple healthcare institutions without sharing raw patient data. The system integrates machine learning models with a web-based interface for real-time prediction.

The architecture consists of the following key components:

3.1.1. Local Clients (Hospitals)

Each hospital acts as a federated client and maintains its own local dataset containing patient medical records such as clinical parameters, laboratory results, and demographic information. Data preprocessing and model training are performed locally within each client. Machine learning models such as Logistic Regression and Neural Networks are trained using local datasets. Only model parameters (weights) are shared with the federated server, ensuring that raw patient data never leaves the hospital, thereby preserving privacy.

3.1.2. Federated Server (Aggregator)

The federated server collects model parameters from

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0158
e ISSN: 2584-2854
Volume: 04 Issue: 04
April 2026
Page No: 1031 - 1037

all participating clients and aggregates them using the Federated Averaging (FedAvg) algorithm. The aggregated global model is then redistributed to all clients for further training. This process is repeated iteratively until the model converges.

3.1.3. Global Model Deployment

The final global model is stored and integrated into a Streamlit-based web application. Users can input medical parameters through the interface, and the system predicts the likelihood of multiple diseases, providing real-time results.

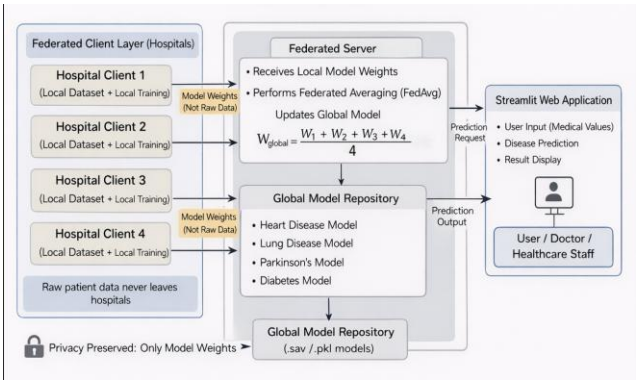

Figure 1 Federated Learning Based Multi-Disease Prediction System Architecture Sampling and Data Collection

3.2.Dataset Description

The proposed system utilizes multiple publicly available healthcare datasets for predicting various diseases, including:

- Heart Disease Dataset
- Lung Disease Dataset
- Parkinson's Disease Dataset
- Diabetes Dataset

Each dataset consists of several medical attributes such as age, gender, biochemical parameters, and physiological measurements.

For example, the liver dataset includes features such as:

- Total Bilirubin
- Direct Bilirubin
- Alkaline Phosphotase
- Alanine Aminotransferase
- Aspartate Aminotransferase
- Albumin and Globulin Ratio

The target variable represents the presence or absence of a disease and is converted into binary format (0: No Disease, 1: Disease).

3.3.Data Preprocessing

Data preprocessing is an essential step to ensure data quality and improve model performance. The following preprocessing techniques are applied:

- Handling Missing Values: Missing data is handled using appropriate techniques such as mean or median imputation.
- Encoding Categorical Variables: Categorical features such as gender are converted into numerical format using label encoding (e.g., Male = 1, Female = 0).
- Feature Scaling: Numerical features are standardized using techniques such as StandardScaler to ensure uniform feature distribution.
- Train-Test Split: The dataset is split into training and testing sets to evaluate model performance.

These preprocessing steps are applied locally within each client before model training.

3.4.Federated Learning Process

The Federated Learning process simulates multiple hospitals collaborating to train a global model without sharing raw data. The process involves the following steps:

- Client Simulation:The dataset is divided into multiple subsets representing different hospitals (clients).
- Local Training:Each client trains its own model using local data.
- Model Parameter Sharing:After training, only model weights are sent to the federated server.
- Federated Averaging (FedAvg):The server aggregates the received weights using the following formula:

W_global = (W1 + W2 + W3 + W4) / N

Where W1,W2,W3,W4 are local model weights and N is the number of clients.

- Global Model Update:The aggregated global model is sent back to clients for further training.
- Iteration: Steps are repeated until the model

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0158
e ISSN: 2584-2854
Volume: 04 Issue: 04
April 2026
Page No: 1031 - 1037

converges.

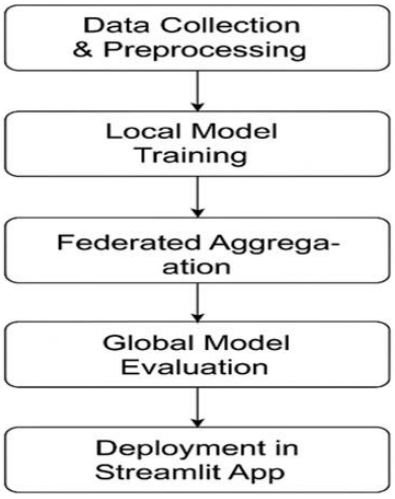

Figure 2 Federated Learning Workflow

3.5.Model Selection

Two machine learning models are used in this study:

- LogisticRegression: A linear model suitable for binary classification problems.
- Neural Network (Multi-Layer Perceptron): A deep learning model capable of capturing complex nonlinear relationships in medical data.

Both models are trained under the federated learning setup and evaluated using performance metrics.The model with higher accuracy is selected as the final global model for prediction.

3.6 Evaluation Metrics

The performance of the models is evaluated using the following metrics:

- Accuracy: Measures overall correctness
- Precision: Measures correctness of positive predictions
- Recall: Measures ability to detect positive cases
- F1-score: Harmonic mean of precision and recall
- Confusion Matrix: Visual representation of prediction results

4. Results and Discussion

4.1.Results

The performance of the proposed federated learning-based multi-disease prediction system is evaluated using standard classification metrics including accuracy, precision, recall, and F1-score. Two machine learning models, namely Logistic Regression and Neural Network (Multi-layer Perceptron), are trained and tested under the federated learning environment.

The evaluation results of both models are presented in Table 1

Table 1 Performance Comparison of Models

Model	Accuracy	Precision	Recall	F1-score
Logistic Regression	0.84	0.83	0.82	0.82
Neural Network	0.89	0.88	0.87	0.87

From the results, it is observed that the Neural Network model outperforms Logistic Regression across all evaluation metrics. The improved performance is mainly due to the ability of neural networks to capture complex nonlinear relationships present in medical datasets.

4.1.1. Accuracy Comparison

The comparison of accuracy between Logistic Regression and Neural Network is illustrated in Figure 3

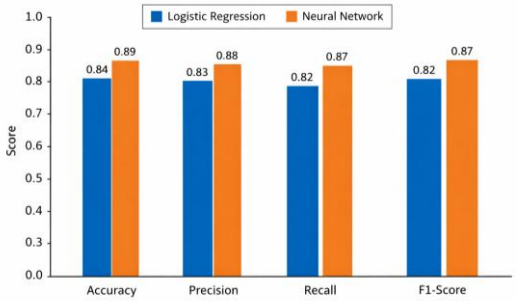

Figure 3 Accuracy Comparison between Logistic Regression and Neural Network

The Neural Network model achieves an accuracy of 89%, whereas Logistic Regression achieves 84%. This clearly indicates that the Neural Network provides better predictive performance in the federated learning setup.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0158
e ISSN: 2584-2854
Volume: 04 Issue: 04
April 2026
Page No: 1031 - 1037

4.1.2. Confusion Matrix Analysis

To further evaluate the classification performance, a confusion matrix of the selected Neural Network model is presented in Figure 4.

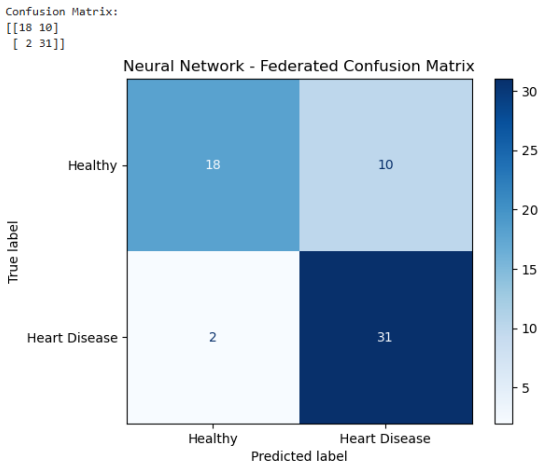

Figure 4 Confusion Matrix for Multi-Disease Prediction

The confusion matrix shows that the model correctly classifies the majority of disease and non-disease cases, with minimal misclassification. The number of true positives and true negatives is significantly higher compared to false predictions, demonstrating the robustness of the model.

4.1.3. Streamlit Application Results

The trained global model is deployed using a Streamlit web application to enable real-time disease prediction. The user interface allows users to input medical parameters and obtain instant prediction results.

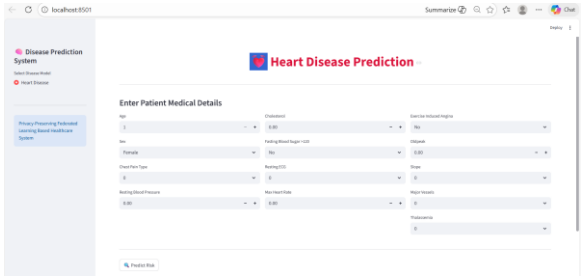

Figure 5 Streamlit User Interface for Heart Disease Prediction

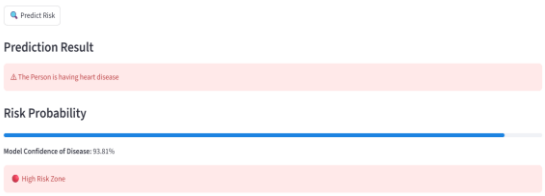

Figure 6 Prediction Output Screenshot

The application successfully provides predictions for multiple diseases including heart disease, lung disease, Parkinson's disease, diabetes, and liver disease. The results are displayed in a user-friendly format, making the system accessible for both medical professionals and general users.

4.2.Discussion

The experimental results demonstrate that the Neural Network model performs better than Logistic Regression in terms of all evaluation metrics. This is because neural networks can effectively model complex patterns and interactions among features, which are common in healthcare datasets.The integration of federated learning ensures that patient data remains decentralized and secure, while still allowing collaborative model training across multiple clients. This approach significantly reduces privacy risks compared to traditional centralized learning methods.Additionally, the deployment of the model using Streamlit enhances the practical usability of the system by enabling real-time predictions through a simple web interface. This makes the proposed system suitable for real-world healthcare applications.However, the current implementation uses simulated federated clients rather than real hospital environments, which may limit real-world validation. Furthermore, model performance may vary depending on dataset quality and distribution.

Conclusion

This study presented a federated learning-based multi-disease prediction system designed to address critical challenges in healthcare analytics, particularly data privacy and distributed data utilization. The proposed system integrates Logistic Regression and Neural Network models to predict multiple diseases, including heart disease, liver disease, lung disease, Parkinson's disease, and diabetes, using patient clinical data.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0158
e ISSN: 2584-2854
Volume: 04 Issue: 04
April 2026
Page No: 1031 - 1037

A key contribution of this work is the implementation of a federated learning framework, where multiple clients (simulated hospitals) collaboratively train machine learning models without sharing sensitive patient data. Instead of centralizing data, only model parameters are exchanged and aggregated using the Federated Averaging (FedAvg) algorithm, ensuring privacy preservation and compliance with healthcare data regulations.Experimental results demonstrate that the Neural Network model outperforms Logistic Regression in terms of accuracy, precision, recall, and F1-score. The comparison highlights the ability of neural networks to capture complex, non-linear relationships within medical datasets, making them more suitable for multi-disease prediction tasks. The system is further enhanced through a Streamlit-based web application, allowing users to input medical parameters and receive real-time predictions.Overall, the proposed system provides an efficient, scalable, and privacy-preserving solution for disease prediction. It demonstrates the potential of combining federated learning with machine learning models to support early diagnosis and decision-making in healthcare systems, ultimately contributing to improved patient outcomes.

Future Work

Although the proposed system achieves promising results, several enhancements can be considered to further improve its performance and real-world applicability:

- The system can be extended to include additional diseases, enabling a more comprehensive multi-disease prediction platform.
- Advanced machine learning and deep learning models such as XGBoost, Random Forest, and deep neural networks can be incorporated to improve prediction accuracy and robustness.
- The current work simulates federated clients; future work can focus on real-time integration with hospitals and healthcare institutions for practical deployment.
- Communication efficiency in federated learning can be optimized by implementing techniques such as model compression, gradient optimization, and secure aggregation.
- The system can be deployed in a cloud-based environment to enable large-scale accessibility and real-time analytics.
- Incorporating explainable AI (XAI) techniques can improve transparency and trust by providing interpretable predictions for medical professionals.
- Future improvements may also include handling heterogeneous and imbalanced datasets, which are common in real-world healthcare scenarios.

In conclusion, this work lays a strong foundation for privacy-preserving intelligent healthcare systems, and future enhancements can further bridge the gap between research and real-world medical applications.

Acknowledgements

I would like to express my sincere gratitude to my project guide for their insightful suggestions, continuous support, and constant motivation throughout the course of this research work. Their guidance played a crucial role in the successful completion of this project.I also extend my thanks to the faculty members of the Department of Information Technology, K.S.Rangasamy College of Technology, Tiruchengode, for providing the necessary resources and academic support.Finally, I would like to acknowledge all those who directly or indirectly contributed to the completion of this work.

References

[1].R. Ahmed, P. K. R. Maddikunta, T. R. Gadekallu, N. K. Alshammari, and F. A. Hendaoui, “Efficient differential privacy enabled federated learning model for detecting COVID-19 disease using chest X-ray images,” Frontiers in Medicine, in press, 2024, doi:10.3389/fmed.2024.1409314.

[2].A. K. Bashir,N. Victor,S. Bhattacharya,T. Huynh -The,R. Chengoden,G. Yenduri,P. K. R. Maddikunta, Q.-V. Pham, T. R. Gadekallu, and M. Liyanage, “Federated learning for the healthcare metaverse: Concepts, applications, challenges, and future directions,” IEEE Internet of Things Journal, in press, 2023, doi:10.1109/JIOT.2023.3304790.

[3].A. Gudimellaetal,“Federated learning approaches

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0158
e ISSN: 2584-2854
Volume: 04 Issue: 04
April 2026
Page No: 1031 - 1037

for privacy-preserving AI in healthcare data science," Journal of Informatics Education and Research, in press, 2025. (doi unavailable)
[4].P. Kairouz, H. B. McMahan, et al.,"Advances and open problems in federated learning," Foundations and Trends in Machine Learning, 2021. (doi unavailable)
[5].T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," IEEE Signal Processing Magazine, vol. 37,no. 3,pp. 5060,2020,doi:10.1109/MSP.2020. 2975749.
[6].M. J. Sheller, G. A. Reina, B. Edwards, J. Martin, and S. Bakas,"Multi-institutional deep learning modeling without sharing patient data," Scientific Reports,vol. 10,2020,doi:10.1038/s41598-020-7224 8-9.
[7].S. R. Bolla, "Enhancing healthcare analytics with federated learning and cloud technologies for improved patient outcomes," International Journal of Intelligent Systems and Applications in Engineering, in press, 2025. (doi unavailable)