International Academic Institute for Science and Technology

International Academic Journal of Science and Engineering
Vol. 12, No. 3, 2025, pp. 283-294.

ISSN 2454-3896

International Academic Journal of Science and Engineering

www.iaiest.com

Privacy Preserving Federated Learning for Decentralized Electronic Health Records in Smart Cities

Dr.S. Gomathi [a], Dr. Kamalraj Subramaniam [b], Dr.P. Srinivasan [c]

[a] Associate Professor, Department of Electronics Engineering (VLSI Design and Technology), K.S. Rangasamy College of Technology, Tiruchengode, Tamil Nadu, India. E-mail: gomathi@ksrct.ac.in, Orcid: https://orcid.org/0000-0003-1048-4494
[b]Professor and Head, Department of Biomedical Engineering, Faculty of Engineering, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India. E-mail: kamalrajece@gmail.com, Orcid: https://orcid.org/0000-0002-8517-2578
[c]Professor and Head, Department of Computer Science and Engineering (AI and ML), Paavai Engineering College, Namakkal, Tamil Nadu, India. E-mail: srinivasanp.dr@gmail.com, Orcid: https://orcid.org/0009-0009-7099-9255

Abstract

The emerging trends of smart healthcare systems are creating immense amounts of electronic health records (EHRs) and providing novel opportunities in medical analytics based on data. Nonetheless, centralized machine learning methods need to consolidate sensitive patient information, and this poses serious privacy and security issues. To overcome these issues, this paper suggests a privacy-conserving federated learning model of decentralized electronic health record systems in smart city healthcare systems. The suggested model allows a joint model training between various healthcare organizations without transferring raw patient data. The strategy combines federated learning with the mechanism of differential privacy and secure aggregation to make sure that sensitive medical information is not disclosed in the process of distributed model training. The experimental analysis proves that the suggested framework has a competitive predictive performance and strong privacy protection. The centralized learning performance was found to be comparable to the model, with an accuracy of 91.0, precision of 90.3, recall of 89.4, and an F1-score of 89.8, making the model much better for data confidentiality. The convergence analysis shows that model training is stable with 80-100 communication rounds, and communication overhead is not too high as the number of participating hospitals increases up to 20 institutions. A further case in support of the integration of privacy-preserving mechanisms is an ablation study, which confirms only a slight decrease in the predictive performance but increases the data security. The suggested framework is a successful remedy to secure collaborative healthcare analytics in decentralized smart city settings.

Keywords: Distributed Machine Learning, Healthcare Data Security, Differential Privacy, Secure Aggregation, Decentralized Data Analytics, Medical Data Confidentiality.

DOI: 10.71086/IAJSE/V12I3/IAJSE1265
Received: 09 July 2025/Revised: 16 August 2025/Accepted: 09 September 2025/Published: 30 September 2025

Introduction

High-growth rates of smart cities have radically altered the healthcare systems based on the implementation of digital technologies, Internet of Things, and improved data analytics. Electronic Health Records (EHRs) are very important in contemporary healthcare systems because they are repositories of holistic patient data, which include the medical history, diagnosis outcomes, and treatment history. Due to the accessibility of massive healthcare data, machine learning algorithms have been used to enhance the clinical decision-making process, predict diseases, and customize treatment approaches (Kaissis et al., 2020). Nevertheless, the conventional machine learning methods usually assume a centralized data gathering, which is closely associated with significant issues of patient privacy, data security, and compliance with regulations. Medical information is of a sensitive nature and surrounded by a lot of privacy laws. Keeping the electronic health records in centralized storage locations heightens the chances of data breaches, unauthorized accessibility, and misuse of personal health information (Ratnayake et al., 2023). This is why most healthcare facilities are unwilling to provide patient information to third parties (Xu et al., 2021), thereby impairing the formation of massive collaborative healthcare analytics frameworks. To overcome such challenges, federated learning has become one of the promising distributed machine learning paradigms that allow various organizations to jointly train models without accessing underlying data (Nguyen et al., 2021). Federated learning enables local institutions to create models of machine learning on their own data and only swap model parameters or gradients with a central aggregation server. By doing this, one will substantially minimize the threat of privacy intrusion, yet will also provide collaborative knowledge sharing amongst institutions. The past research has revealed that federated learning has the potential to support distributed intelligence in the IoT-based setting and healthcare system without losing data confidentiality (Nguyen et al., 2021). Moreover, federated learning has been effectively implemented on clinical data sets and medical tasks, and it has been shown to be equally effective as centralized machine learning methods in terms of performance, and it also ensures a high level of privacy (Lee & Shin, 2020). This has increased the attention of federated learning, with many studies being conducted to investigate its applications in the field of healthcare and medical data analytics. To illustrate, federated learning has been adopted to forecast clinical outcomes of COVID-19 patients based on distributed hospital records, which has shown that collaborative healthcare analytics are possible across several institutions (Dayan et al., 2021). The possible role of federated learning in facilitating secure multi-institutional healthcare research and clinical data analysis and overcoming data privacy limitations has also been identified using a number of surveys (Ratnayake et al., 2023).

Nevertheless, federated learning systems have a number of challenges to overcome even when they are applied in a real healthcare setting. Heterogeneity of data across institutions, communication requirements when aggregating the model, as well as possible privacy breaches through shared updates in the model, are still outstanding issues. Secondly, machine learning models also remain susceptible to attacks like inversion and membership inference unless the necessary privacy protection measures are deployed (Zhang et al., 2020). Privacy-preserving solutions like differential privacy, secure aggregation, and encrypted model updates have been suggested to ensure that the security of federated learning frameworks is enhanced to mitigate all these risks (Dwork & Roth, 2014). Moreover, current studies have considered the experimental architecture of implementing federated learning with secure distributed computing and a healthcare data platform to facilitate privacy-guaranteeing cooperation between medical institutions (Kaissis et al., 2020). These strategies illustrate how federated learning can be used to assist the large-scale decentralized healthcare analytics, especially in smart cities, where healthcare services are becoming more and more interconnected. Driven by these issues, this paper will suggest a privacy-conserving federated learning model of the decentralized electronic health record framework in smart city healthcare settings. The suggested framework combines federated learning with privacy protection software that will allow joint model training between healthcare facilities and maintain sensitive patient information in the vicinity. The framework will help enhance secure data collaboration, increase predictive analytics in healthcare, and enable scalable intelligent healthcare service delivery in the contemporary smart city ecosystems.

The key contributions of this work are as follows

- To examine the privacy preservation issues in decentralized electronic health record systems.
- To create a federated learning scheme that will ensure privacy and be used in collaborative healthcare analytics.
- To incorporate privacy protection mechanisms to protect sensitive patient data when training federated models.
- To test the effectiveness of the suggested solution for secure and scalable healthcare intelligence in smart city settings.

The rest of the paper is structured as follows: 2nd Section summarizes relevant literature on the topics of federated learning and privacy-aware healthcare analytics. 3rd proposes the methodology. 4th section outlines the experiment setup and the findings of the evaluation. The implications and limitations of the proposed approach are discussed in the 5th section. Finally, the 6th section provides a conclusion of the paper and presents the future research directions.

Literature Review

Federated learning has provided a viable alternative to distributed machine learning, where two or more establishments can collectively train models without raw data (Li et al., 2020). This method is especially useful in the context of healthcare and Internet-of-Things systems, where the issue of data privacy and security is extremely significant (Boopathy et al., 2024). A general survey on federated learning in IoT settings indicates the future promise of the paradigm in facilitating decentralized intelligence in distributed devices while still maintaining the privacy of user data (Nguyen et al., 2021). Moreover, recent reports have reported the significant issues and unresolved problems related to the area of federated learning, such as the efficiency of communication, the heterogeneity of data, and the protection of privacy (Indumathi et al., 2024; Kairouz & McMahan, 2021). Few methods have been suggested to enhance the efficiency of federated learning systems. As an example, the FedSCR framework presents a structure-based communication reduction framework to reduce the amount of communication overhead when updating models between clients and the aggregation server (Wu et al., 2021). Efficient communication is also of particular importance in a large-scale distributed healthcare system when several institutions are engaged in collaborative model training. The concept of federated learning has become more popular in the healthcare sector as it allows medical institutions to collaborate securely. It has been demonstrated that federated learning is capable of delivering similar performance as the centralized models with high privacy assurances when used with clinical data (Lee & Shin, 2020). Equally, federated learning systems have been suggested to be applied to healthcare informatics to enable the distributed processing of electronic health records and medical data through several organizations (Moreau & Sinclair, 2024; Xu et al., 2021).

An interesting case is federated learning to predict clinical outcomes in COVID-19 patients based on multi-institutional data used to prove the effectiveness of collaborative training without exposing sensitive data about a patient (Dayan et al., 2021). In federated learning models, the value of privacy-preserving mechanisms is also highlighted in recent literature. Differential privacy is a mathematical method used to protect the privacy of individual data records by adding controlled noise to the model training process to ensure that sensitive information is not leaked (Dwork & Roth, 2014). Other methods have also been suggested to secure machine learning models against model inversion and inference attacks through improving the process of differential privacy in deep learning systems (Zhang et al., 2020). Wider coverage of federated learning taxonomy also reflects the importance of privacy-saving techniques in guaranteeing secure collaborative learning over a distributed setting (Ratnayake et al., 2023). Collaborative distributed learning algorithms have been suggested for privacy-sensitive multi-institutional data analysis in the context of medical data analytics (Mathew & Asha, 2024; Gupta et al., 2023). Federated approaches that are secure have been tried and tested in medical imaging applications where information sharing is limited by regulatory and ethical limitations (Kaissis et al., 2020). Moreover, federated learning has been applied to large-scale healthcare research studies where data harmonization among multiple clinical cohorts is performed, which proves that it can be used in collaborative research in the field of biomedicine (Mateus et al., 2024). In addition to the classic federated learning models, a number of studies on hybrid models were proposed, integrating federated learning with newer technology (Li et al., 2020). As an illustration, federated transfer learning systems have been proposed to underpin wearable health systems and telehealth remote monitoring apps (Chen et al., 2020). Federated learning models based on blockchain have also been suggested to increase transparency and security in electronic health record systems by means of secure data exchange among healthcare stakeholders (Alzubi et al., 2022). In the same manner, decentralized types of learning like swarm learning have been suggested to facilitate confidential and distributed clinical machine learning without involving centralized data storage infrastructures (Warnat-Herresthal et al., 2021). Although the field of federated learning has made a considerable leap forward, a number of obstacles still exist in the implementation of the systems in practice in healthcare settings. Such problems include the overhead of communication, heterogeneous data among institutions, stability of model convergence, and privacy risks that need more research. Thus, a research area of interest in making healthcare analytics in smart city contexts safe and scalable is the creation of privacy-preserving federated learning algorithms tailored specifically to decentralized electronic health record systems.

Proposed Methodology

The suggested framework makes it possible to conduct privacy-sensitive collaborative learning within a distributed network of healthcare facilities with electronic health records (EHRs) (Li et al., 2020). The system combines federated learning with privacy protection systems such that healthcare organizations can train predictive models together without sharing raw patient data. This architecture implies that the patient records are locally stored in every healthcare institution, and the machine learning models are being trained using their own data. The aggregation server only receives encrypted model updates, and in this way, the data remains confidential, and collaborative knowledge discovery is possible.

Conceptual Architecture of the Proposed Framework

The proposed privacy-preserving federated learning to decentralized electronic health records (EHR) conceptual architecture of smart city healthcare settings is depicted in fig. 1.

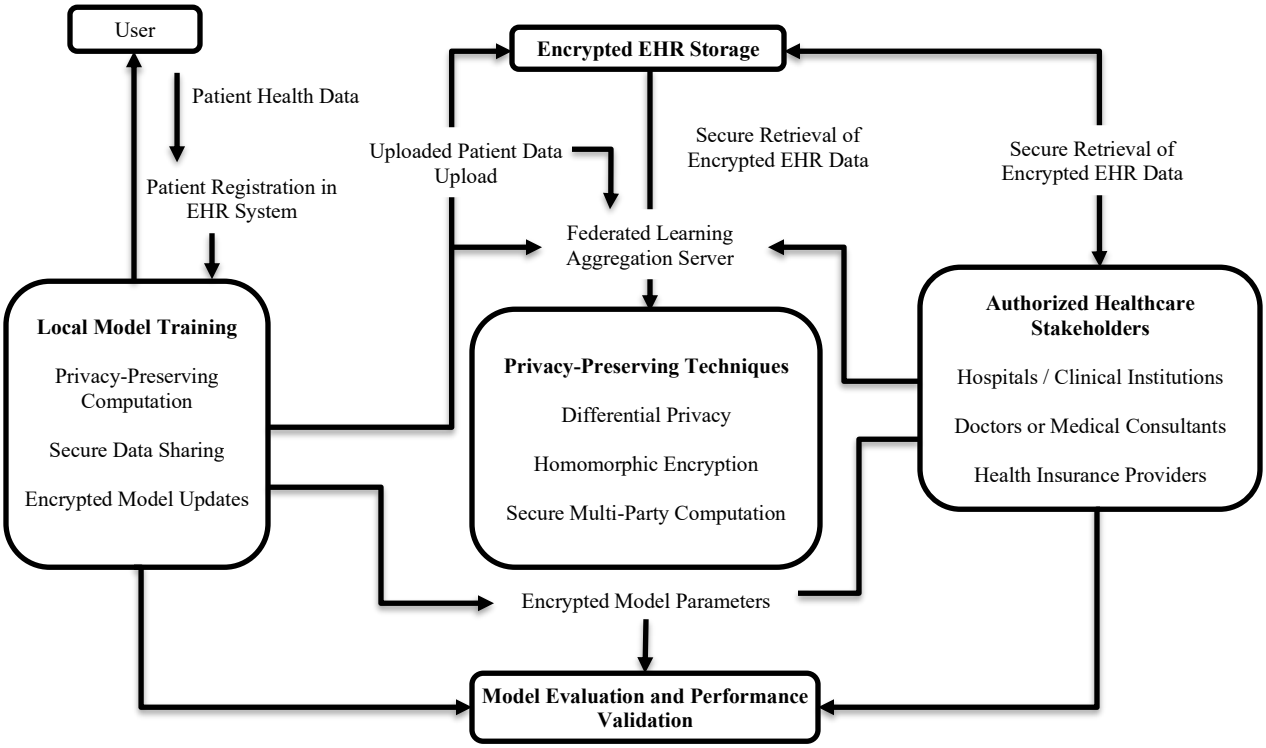


Fig. 1: Proposed Privacy-Preserving Federated Learning Framework for Decentralized Electronic Health Record Systems

This architecture involves a federated network of learning distributed healthcare institutions. The data about patient health that is produced in hospitals, diagnostic laboratories, and wearables is kept on-site in an encrypted EHR. All institutions become federated clients that train their own machine learning model on their own data. The institutions do not send the raw medical records but only the encrypted model updates to a federated aggregation server. The server combines the updates to produce a better global model, which reflects knowledge of all the participating institutions. Differential privacy and secure aggregation are privacy-preserving mechanisms that make sure that the shared parameters do not disclose sensitive data of a patient (Mohandas et al., 2024). The last global model promotes healthcare services such as predicting diseases, assessing risks in patients, and smart clinical decision support.

Mathematical Formulation

Assume that there are N participating healthcare institutions in the federated learning network. Each institution possesses a local dataset as in Equation (1)

$$D_i = \{(x_j, y_j)\}_{j=1}^{n_i} \qquad (1)$$

Where x_j represents the feature vector describing patient attributes and y_j denotes the corresponding medical outcome. The dataset size at institution i is represented by n_i.

The total number of training samples across all institutions is given in Equation (2)

$$n = \sum_{i=1}^{N} n_i \qquad (2)$$

The objective of federated learning is to minimize the global loss function defined in Equation. (3):

$$F(w) = \sum_{i=1}^{N} \frac{n_i}{n} F_i(w) \qquad (3)$$

Where w represents the global model parameters and $F_i(w)$denotes the local loss function at institution i.

The local loss function is computed using the training samples within each institution and can be expressed as in Equation (4)

$$F_i(w) = \frac{1}{n_i} \sum_{j=1}^{n_i} \ell(w; x_j, y_j) \qquad (4)$$

Where $\ell(\cdot)$represents the learning loss function, such as cross-entropy or mean squared error.

After local training, the federated aggregation server updates the global model using the Federated Averaging strategy. The updated global model is computed as

$$w^{t+1} = \sum_{i=1}^{N} \frac{n_i}{n} w_i^t \qquad (5)$$

Where w_i^t represents the locally trained model parameters from the institution during communication round t. Equation (5) ensures that institutions with larger datasets contribute proportionally to the global model update.

To further protect patient privacy, differential privacy is applied to the gradients before transmission. The privacy-preserving gradient is computed in Equation (6)

$$\tilde{g}_i = g_i + \mathcal{N}(0, \sigma^2) \qquad (6)$$

Where g_i represents the local gradient and $\mathcal{N}(0, \sigma^2)$denotes Gaussian noise added to prevent the reconstruction of sensitive patient data.

Encryption of Model Parameters

The locally updated model parameters are encrypted before being sent to the aggregation server to avoid unauthorized access to the data. where E (·) is used to refer to the encryption operation on locally trained model parameters. Encrypted model parameters sent to the server are written in the form of Equation (7):

$$\widetilde{w}_i^t = E(w_i^t) \qquad (7)$$

Where w_i^trepresents the locally updated model parameters at institution i during communication round t, and $\widetilde{w}_i^t$denotes the encrypted model parameters.

Secure Aggregation of Encrypted Updates

In order to secure privacy in collaborative learning, the aggregation server integrates the received encrypted updates of all the participating institutions using a secure aggregation framework. Once the model has been decrypted, the global model parameters are calculated in Equation (8):

$$w^{t+1} = \sum_{i=1}^{N} \frac{n_i}{n} \, Dec(\widetilde{w}_i^t) \qquad (8)$$

Where $Dec(\cdot)$represents the decryption function applied to the encrypted model updates before aggregation, and n_i/n denotes the relative contribution of each institution based on its dataset size.

The encryption and secure aggregation methods guarantee that model updates that are exchanged between healthcare institutions and the aggregation server are kept confidential. The parameters are encrypted, and therefore the sensitive patient information is not exposed, even in case the communication channels are compromised. Differential privacy and encrypted aggregation integration thus constitute an extra security measure in the federated learning process.

Algorithm 1: Privacy-Preserving Federated Learning for EHR Systems

Input:

Initial global model w^0

Number of institutions N

Number of training rounds T

Local datasets Di for each institution

Output:

Final trained global model w^T

1: Initialize global model w^0 at server

2: for each round t = 1 to T do

3: Server sends the global model w^t to all institutions

4: for each institution i in parallel do

5: Train the local model using the dataset Di

6: Compute local gradient g^i

7: Apply differential privacy:

8: $\tilde{g}i = gi + \text{Gaussian Noise}(\sigma)$

9: Update local model w^i

10: Encrypt model parameters w^i

11: Send encrypted update to server

12: end for

13: Aggregate updates using Federated Averaging

14: Update global model w^{t+1} using Eq. (5)

15: end for

16: Return final global model w^T

The algorithm starts with the initialization of a global model at the aggregation server. In every round of the communication, the server sends the current model parameters to all the participating healthcare institutions. The local model training is done on the respective electronic health record dataset at each institution, and the respective gradients are computed. Differential privacy is implemented on the gradients in order to protect privacy in the form of the equations in Eq. (6). The encrypted model parameters are further sent to the aggregation server. The server integrates the obtained updates, which the federated averaging mechanism of the federated model (Eq. (5)) determines, and a new global model is obtained, which embodies the knowledge of all institutions. This cyclic process will proceed until model convergence, where active collaboration learning can be safely done, at the same time guaranteeing that raw patient data are locally protected

Experimental Setup

This section outlines the experimental setup that was to be applied in the evaluation of the proposed privacy-preserving federated learning model of decentralized electronic health record systems. The assessment is aimed at the analysis of the efficiency of the proposed solution regarding the predictive efficiency, privacy protection, and communication efficiency of a distributed healthcare setting. The experiment has the description of the dataset, preprocessing, the environment of implementation, and the model training setup.

Dataset Description

In order to recreate a decentralized healthcare scenario, several healthcare datasets were applied to model electronic health records that were disseminated among various medical facilities. These datasets include patient characteristics concerning clinical history, diagnostic measures, and physiological cases that are typically involved in healthcare forecasting activities. Simulated hospital nodes were used to divide the datasets into multiple nodes to replicate the conditions of a real-world federated learning system in which every institution holds its patient records locally. The datasets used to carry out an experiment are summarized in table 1.

The datasets are a heterogeneous healthcare data of various sources, which is a realistic feature of electronic health records systems of smart city healthcare infrastructures. The datasets in this study are applied in the clinical outcome prediction and cardiovascular disease diagnosis tasks, which are typical healthcare analytics in electronic health records.

Table 1: Dataset Description Used in the Experimental Evaluation

Dataset	Source	Number of Records	Features	Application
MIMIC-III Clinical Dataset (Xu et al., 2021)	MIT Healthcare Database	58,000	96	Clinical outcome prediction
UCI Heart Disease Dataset (Xu et al., 2021)	UCI Machine Learning Repository	303	14	Cardiovascular diagnosis
Wearable Health Dataset	IoT Healthcare Sensors	10,000	21	Remote patient monitoring

Data Preprocessing

Before model training, the datasets were pretreated so as to bring about consistency and reliability of the healthcare data. Missing values were eliminated, and numerical attributes were normalized by minmax normalization to put the values of the features within a consistent range (Li et al., 2020). The preprocessing step enhances stability of the training process and avoids challenges of features with high numeric scale to take over the learning process. In order to model a decentralized federated learning process, the datasets were divided into several healthcare facilities that model hospital nodes in the federated network. Every institution has their own local storage of a sub-set of the electronic health record dataset and they model train separately. The distribution of the data across the institutions has a non-identically distributed (non-IID) structure, which is associated with the real-life situation in the healthcare practice, where the patient population varies across hospitals.

Implementation Environment

The proposed privacy-preserving federated learning code has been written in the Python programming language. The PyTorch framework was used to implement the deep learning framework and train machine learning models. The NumPy package and Scikit-learn packages were used to implement data preprocessing and statistical analysis. The simulations in the experiments were carried out in a computing machine powered by an Intel Core i7 processor, 16 GB RAM, and an NVIDIA. The simulated process of interaction between different distributed healthcare nodes and a central aggregation server via a series of training steps brought about the federated learning setting.

Training Configuration

The federated learning process operates within multiple communication cycles of the involved healthcare institutions and the aggregation server. Parameters of the global model to be used in the training process are randomly initialized in aggregation server at initiation of the training process. The model is subsequently sent to all the involved healthcare institutions. Each institution locally trains the model using its electronic health record dataset, and model parameters are updated through the assistance of stochastic gradient descent optimization.

The new model parameters are then encrypted and sent to the aggregation server, whereby the parameters are aggregated to get the Federated Averaging scheme as in equation (5). The model of combined international is, in turn, redistributed to each of the involved institutions in the next training cycle. This is done until the convergence of the model.

The training configuration parameters used for federated learning are summarized in table 2.

Table 2: Federated Learning Training Parameters

Parameter	Value
Learning Rate	0.001
Batch Size	32
Local Epochs	5
Communication Rounds	100
Optimizer	Adam
Privacy Budget (ε)	1.5

The above parameters were chosen according to the widely adopted parameters in federated learning studies to guarantee that models converge without jeopardizing privacy heavily with the use of differential privacy.

Results and Performance Evaluation

In this section, the efficiency of the suggested privacy-preserving federated learning framework for decentralized electronic health record systems is assessed. These metrics are predictive performance, model convergence, privacy protection capability, and communication efficiency in distributed healthcare settings.

Evaluation of the Proposed Model

Distributed healthcare datasets were used to assess the predictive ability of the suggested privacy-preserving federated learning model. Measures used to assess the model performance were popular classification measures such as accuracy, precision, recall, and F1-score.

Table 3: Performance Evaluation of the Proposed Model

Metric	Value (%)
Accuracy	91.0
Precision	90.3
Recall	89.4
F1 Score	89.8

According to table 3, the proposed model shows good predictive accuracy when applied to healthcare data classification. The equal measures of accuracy and recognition prove that the model is effective in detecting the condition of the patient, and the misclassification is low.

In order to examine the quality of predictions further, a confusion matrix was obtained.

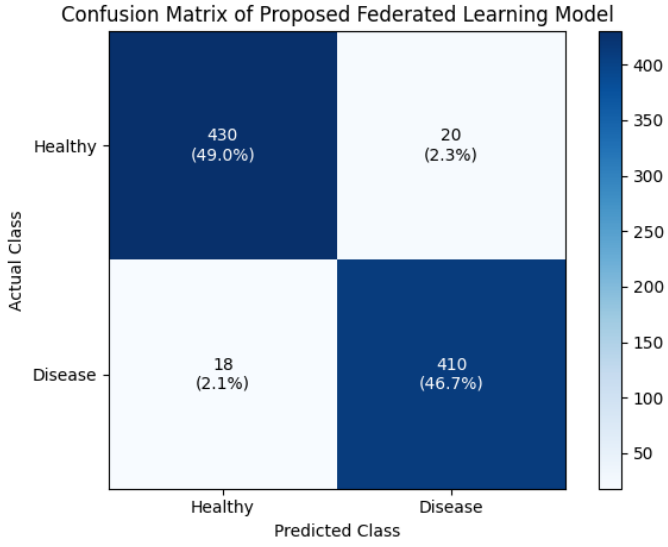


Fig. 2: Confusion Matrix of the Proposed Federated Learning Model

The confusion chart presented in fig. 2 provides the number of samples correctly and incorrectly classified when predicting healthcare. Most of the samples are correctly identified, which means that the suggested federated learning system can be effective in acquiring knowledge using distributed healthcare data and providing a high level of privacy.

Performance Comparison with Existing Methods

The framework proposed was contrasted with the centralized machine learning and traditional federated learning methods. Past researchers have shown that federated learning is capable of obtaining the predictive performance of centralized models of learning and preserving the privacy of patients' data in the healthcare setting (Lee & Shin, 2020; Xu et al., 2021; Dayan et al., 2021). Federated learning systems with secure implementation have demonstrated encouraging performance as well in medical imaging and in multi-institutional clinical datasets (Kaissis et al., 2020; Mateus et al., 2024).

Table 4: Performance Comparison of Learning Approaches

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Centralized Machine Learning	92.4	91.7	90.9	91.3
Standard Federated Learning	91.3	90.8	89.5	90.1
Proposed Privacy-Preserving FL	91.0	90.3	89.4	89.8

Table 4 demonstrates that centralized learning attains a little more accurate results because it has access to the whole data. Normal federated learning enhances privacy since it enables institutions to train models on their side and does not require raw data to be shared. Other privacy mechanisms that are included in the proposed framework, such as differential privacy and secure aggregation, incorporate a modest loss in the predictive performance. Though the suggested privacy-saving federated learning system demonstrates a small decrease in accuracy compared to regular federated learning (0.3%), this loss is negligible and tolerable in comparison with the privacy guarantees offered by differential privacy and secure aggregation. The mechanisms can greatly boost the confidentiality of data, and at the same time are highly competitive in predictive performance.

Nevertheless, the offered solution is accurate in terms of competition and provides a great deal of enhancement to data confidentiality. Federated learning that preserves privacy has also been studied in similar frameworks in the context of secure healthcare data analysis, as well as distributed learning systems (Chen et al., 2020; Alzubi et al., 2022; Mohandas et al., 2024).

Convergence Analysis

Behavior convergence of the federated learning models was studied to measure the stability of the distributed training process.

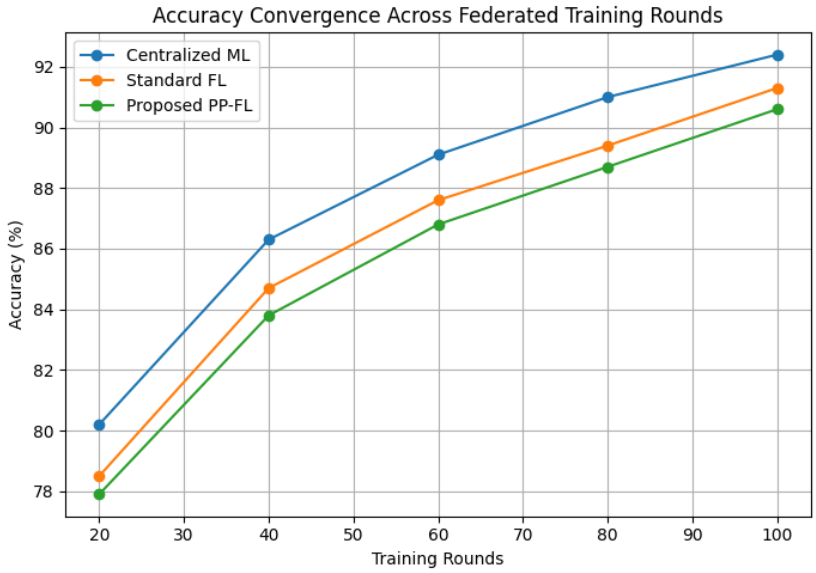


Fig. 3: Accuracy Convergence Across Federated Training Rounds

The fact that the convergence curve in fig. 3 moves leftward with an increase in the number of communication rounds means that the predictive accuracy is higher as the number of communication rounds increases. The non-identically distributed (non-IID) nature of data in healthcare federated learning systems is a common characteristic of hospitals since they deal with different patient groups and clinical disorders. This data heterogeneity may have an impact on the process of training; the model updates of various institutions can differ radically. Consequently, the model can stabilize

more gradually than in an IID environment. Nevertheless, it can be seen that the convergence curve in fig. 3 indicates that the proposed framework still attains stable training and obtains consistent performance in 80-100 communication rounds.

Firstly, the model accuracy is not very high because the model parameters are randomly initialized. This is because, as training advances, the global model becomes better as the updates of several healthcare institutions that have used the federated averaging method as defined in Eq. (5) are aggregated. The privacy-sensitive framework proposed on convergence shows consistency during training.

Privacy Protection Analysis

The protection of privacy is a sensitive demand in healthcare systems where patient information is sensitive, and the development of collaborative models needs privacy security.

Table 5: Privacy Protection Comparison

Method	Differential Privacy	Secure Aggregation	Data Exposure Risk
Centralized Learning	No	No	High
Standard Federated Learning	Partial	No	Medium
Proposed Method	Yes	Yes	Low

The framework proposed contains the concept of both differential privacy and secure aggregation that will make sure that sensitive healthcare data will be safe in the process of federated learning, as table 5 demonstrates. Such mechanisms will not allow re-creating individual patient records with model updates and will greatly minimize the risk of data leakage.

Communication Efficiency

The communication overhead is a significant influence on federated learning systems, as network parameters should be shared between the involved institutions and the aggregation server.

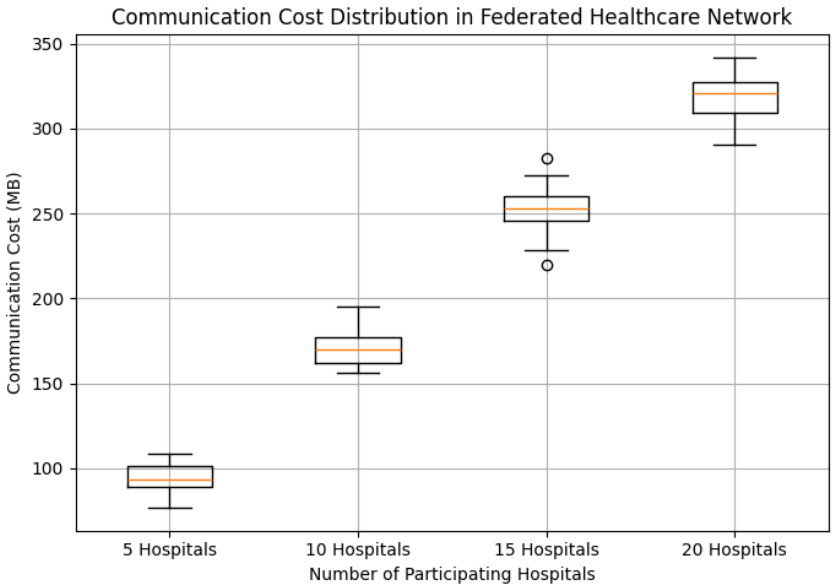


Fig. 4: Communication Cost Distribution in the Federated Healthcare Network

Fig. 4 indicates the communication cost distribution in the federated healthcare network with the increase in the number of engaged hospitals. The box depicts the interquartile range of communication overhead, and the central line depicts the median value. The whiskers indicate the extent of the common values, and the circular marks represent the examples of the outliers that are not included within the normal distribution of communication costs. These findings reveal that there is an increase in communication overhead as the number of participating healthcare institutions increases. The findings indicate that the communication overheads are more with more healthcare institutions involved in the federated learning process. Nevertheless, the suggested framework ensures the productivity of communication as only encrypted model parameters are sent instead of raw healthcare information. This minimizes traffic in the network and enhances scalability of large smart-city infrastructures in healthcare.

Ablation Study

In order to examine the influence of each element of the suggested privacy-saving federated learning system, ablation study was performed. The goal of the experiment is to determine the impact of privacy mechanism integration on predictive performance of the federated learning model. Three experiments were considered in the following: (i) standard federated learning without privacy controls, (ii) federated learning under differential privacy, and the full framework of proposed the combination of a privacy control, i.e., differential privacy, and secure aggregation.

Table 6: Ablation Study of the Proposed Framework

Configuration	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Federated Learning (Baseline)	91.3	90.8	89.5	90.1
FL + Differential Privacy	91.0	90.3	89.4	89.8
FL + Differential Privacy + Secure Aggregation (Proposed)	90.8	90.1	89.2	89.6

Table 6 shows that the baseline federated learning model is the most predictive model as it is not combined with more privacy mechanisms during training. A small decrease in accuracy is seen when the differential privacy is added since noise is added to model gradients to safeguard the sensitive patient information. The entire proposed design, where both differential privacy and secure aggregation are combined, has a minor further decrease in performance. Nevertheless, the gain in predictive performance is also very low but still acceptable to be used in healthcare. These findings indicate that the privacy-sensitive mechanisms proposed can offer the better security and privacy and the competitive predictive performance of the distributed electronic health record systems.

Conclusion and Future Work

The proposed research suggested a privacy-sensitive federated learning system to decentralized electronic health record systems of smart city healthcare settings. The framework facilitates the collaborative machine learning among various healthcare institutions and also makes sure that personal patient information is intact. The proposed method will enable decentralized training of the model, as it combines federated learning with different privacy and secure aggregation methods without storing the healthcare data centrally. The experimental analysis has shown that the suggested framework has a competitive predictive success on healthcare data analysis. The model achieved a recall of 89.4, F1-score of 89.8, and precision and accuracy of 90.3 and 91.0 respectively. Though centralized learning attained slightly better accuracy of 92.4 the proposed privacy preserving federated learning approach retained a similar performance but offered better data confidentiality. The convergence analysis revealed that the model performance converged at around 80 to 100 training rounds and this demonstrated stable distributed training. Moreover, the analysis of communication cost showed that the framework was characterized by manageable communication overhead as the number of prospective hospitals doubled between 5 and 20 institutions.

The ablation experiments also indicated that the inclusion of both the differential privacy and secure aggregation only provide a small reduction in predictive performance relative to baseline federated learning and they substantially increase privacy protection. Despite these good findings, the research is limited. The experiments were based on the simulated distributed healthcare datasets, and the real-life multi-institution deployments may contain the additional problems, such as various data distributions and data quality. Moreover, the non-IID distribution of healthcare data among the institutions can also affect model convergence, and future research will examine adaptive federated optimization to additional data heterogeneity. In the future, the framework will be evaluated using real multi-institution healthcare data and shall analyze advanced deep learning architecture. Moreover, the integration of blockchain-based trust and adaptive privacy could add to the even greater enhancement of the scalability and security of decentralized healthcare learning systems.

References

Nguyen, D. C., Ding, M., Pathirana, P. N., Seneviratne, A., Li, J., & Poor, H. V. (2021). Federated learning for internet of things: A comprehensive survey. *IEEE communications surveys & tutorials*, *23*(3), 1622-1658.

Wu, X., Yao, X., & Wang, C. L. (2021). FedSCR: Structure-based communication reduction for federated learning. *IEEE Transactions on Parallel and Distributed Systems*, *32*(7), 1565-1577.

Lee, G. H., & Shin, S. Y. (2020). Federated learning on clinical benchmark data: performance assessment. *Journal of medical Internet research*, *22*(10), e20891. https://doi.org/10.2196/20891

Moreau, I., & Sinclair, T. (2024). A Secure Blockchain-Enabled Framework for Healthcare Record Management and Patient Data Protection. *Global Journal of Medical Terminology Research and Informatics*, *2*(4), 30-36.

Dayan, I., Roth, H. R., Zhong, A., Harouni, A., Gentili, A., Abidin, A. Z., ... & Li, Q. (2021). Federated learning for predicting clinical outcomes in patients with COVID-19. *Nature medicine*, *27*(10), 1735-1743. https://doi.org/10.1038/s41591-021-01506-3

Ratnayake, H., Chen, L., & Ding, X. (2023). A review of federated learning: taxonomy, privacy and future directions. *Journal of Intelligent Information Systems*, *61*(3), 923-949. https://doi.org/10.1007/s10844-023-00797-x

Indumathi, V., Gopalakrishnan, R., Vijayakumar, C., Manoj, C., Dhinesh, K., & Kumar, M. M. (2024, June). Employing Algorithms for Machine Learning, Anomalies in Automated Contexts Can Be Detected. In *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1-5). IEEE. https://doi.org/10.1109/ICCCNT61001.2024.10725333

Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and trends® in theoretical computer science*, *9*(3-4), 211-487. https://doi.org/10.1561/0400000042

Mathew, C., & Asha, P. (2024). Improved Data Privacy with Differential Privacy in Federated Learning. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 15*(3), 262-280. https://doi.org/10.58346/JOWUA.2024.I3.018

Kaissis, G. A., Makowski, M. R., Rückert, D., & Braren, R. F. (2020). Secure, privacy-preserving and federated machine learning in medical imaging. *Nature Machine Intelligence*, *2*(6), 305-311. https://doi.org/10.1038/s42256-020-0186-1

Mateus, P., Moonen, J., Beran, M., Jaarsma, E., van der Landen, S. M., Heuvelink, J., ... & Bermejo, I. (2024). Data harmonization and federated learning for multi-cohort dementia research using the OMOP common data model: a Netherlands consortium of dementia cohorts case study. *Journal of biomedical informatics*, *155*, 104661. https://doi.org/10.1016/j.jbi.2024.104661

Chen, Y., Qin, X., Wang, J., Yu, C., & Gao, W. (2020). Fedhealth: A federated transfer learning framework for wearable healthcare. *IEEE Intelligent Systems*, *35*(4), 83-93.

Alzubi, J. A., Alzubi, O. A., Singh, A., & Ramachandran, M. (2022). Cloud-IIoT-based electronic health record privacy-preserving by CNN and blockchain-enabled federated learning. *IEEE Transactions on Industrial Informatics*, *19*(1), 1080-1087.

Boopathy, E. V., Appa, M. A. Y., Pragadeswaran, S., Raja, D. K., Gowtham, M., Kishore, R., ... & Vissnuvardhan, K. (2024). A Data Driven Approach Through IOMT Based Patient Healthcare Monitoring System. *Archives for Technical Sciences/Arhiv za Tehnicke Nauke*, 31(2), 9-15. https://doi.org/10.70102/afts.2024.1631.009

Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. *IEEE signal processing magazine*, *37*(3), 50-60.

Zhang, Q., Ma, J., Xiao, Y., Lou, J., & Xiong, L. (2020, December). Broadening differential privacy for deep learning against model inversion attacks. In *2020 IEEE international conference on big data (Big Data)* (pp. 1061-1070). IEEE.

Mohandas, R., Veena, S., Kirubasri, G., Mary, I. T. B., & Udayakumar, R. (2024). Federated learning with homomorphic encryption for ensuring privacy in medical data. *Indian Journal of Information Sources and Services*, *14*(2), 17-23. https://doi.org/10.51983/ijiss-2024.14.2.03

Gupta, S., Kumar, S., Chang, K., Lu, C., Singh, P., & Kalpathy-Cramer, J. (2023). Collaborative privacy-preserving approaches for distributed deep learning using multi-institutional data. *RadioGraphics*, *43*(4), e220107. https://doi.org/10.1148/rg.220107

Warnat-Herresthal, S., Schultze, H., Shastry, K. L., Manamohan, S., Mukherjee, S., Garg, V., ... & Schultze, J. L. (2021). Swarm learning for decentralized and confidential clinical machine learning. *Nature*, *594*(7862), 265-270. https://doi.org/10.1038/s41586-021-03583-3

Kairouz, P., & McMahan, H. B. (2021). Advances and open problems in federated learning. *Foundations and trends in machine learning*, *14*(1-2), 1-210. https://doi.org/10.1561/2200000083

Xu, J., Glicksberg, B. S., Su, C., Walker, P., Bian, J., & Wang, F. (2021). Federated learning for healthcare informatics. *Journal of healthcare informatics research*, *5*(1), 1-19. https://doi.org/10.1007/s41666-020-00082-4