Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 60-72

Federated Learning-based Healthcare Analytics for Privacy-Preserving Patient Data Sharing

[1]Rekha Sahu, [2]Pawan Kumar
[1]Student B.Tech CSE, [2]Assistant Professor
[1,2]Amity School of Engineering and Technology, Amity University Chhattisgarh
[1]sahurekha377@gmail.com, [2]pkumar@rpr.amity.edu

Abstract

The new federated learning healthcare analytics framework is necessary because it helps solve the major challenges posed by traditional centralized machine learning approaches to the healthcare system. The traditional way used in machine learning is to gather sensitive patient information into one place to perform analysis. Because there is a central repository of information in a healthcare environment, there is an increased risk of a data breach occurring, an unauthorized access to sensitive patient data, and insufficient compliance with strict federal regulations, including HIPAA, and GDPR. Therefore, the limitations of the centralized systems make these traditional methods incompatible with today's healthcare environments because data privacy and data security are the highest priorities.The introduction of a federated learning validation system allows for a strong privacy-preserving distributed model training system in which many medical facilities can work together to develop a global healthcare model without having to share raw patient data. Participants will work cooperatively to create a global healthcare model by only sharing model updates rather than their original data with any other participant. The federated learning distributed model training system reduces the amount of exposure of sensitive patient data to other participants and is mitigated by not violating the privacy rights of any participant. Additionally, since each participant entity has the ability to retain ownership and maintain confidentiality of each virtual machine database, the framework will ensure regulatory compliance with their respective regulatory authorities. Incorporation of complex security mechanisms adds another layer of security to this framework's reliability. For example, secure aggregation allows the individual models updates to be unaccessed or reconstructed by a third party, whereas differential privacy uses added noise to prevent individuals from being identified from their EHR data. Furthermore, lightweight encryption adds another level of data security during transmission, thus, increasing the likelihood of preventing cybersecurity incidents and/or guess attacks. Through the experimental evaluation of this proposed framework utilizing simulated electronic health records (EHR) and medical imaging datasets, it has been shown that federated models exhibit performance characteristics comparable to centralized models as determined by prediction accuracy. Moreover, the use of this framework significantly reduces any privacy threat through the reduction of class membership inference and re-identification attacks. The balance of privacy preservation and prediction accuracy provides further evidence of the practicality of this approach for providing improved data security while enabling greater collaborative practices among healthcare organizations.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 60-72

Lastly, the utilization of this framework may further increase the number of organizations that can participate in joint/pooled efforts towards the creation of an intelligence model resulting in enhanced generalizability and enhanced healthcare analytics. This feature is especially beneficial in situations where heterogeneity of data is required to make accurate diagnoses and predictions (e.g., disease detection or medical imaging analyses).

Keywords- Federated learning, Healthcare analytics, Privacy-preserving, Pateint data sharing, secure aggregation, Differential privacy, Machine learning, Data Security.

I. Introduction

An innovative way of analysing healthcare data through federated learning (FL) is when many organisations work together to train machine-learned models using data from hospitals, clinics or research centres without having access to sensitive personal medical information such as electronic health records, medical images, or data from smart devices. FL enables organisations to collaboratively train machine-learning models using only the model's updates by keeping patient data/clinical data locally, thus maintaining individual privacy and data sovereignty. In industries that have many stringent regulations regarding the movement and storage of personal health information such as HIPAA and GDPR the implementation of FL-based solutions makes for an enticing option to build large-scale analytical models of clinical data whilst avoiding privacy-related risks and reducing liability exposure to legal claims.

I.I Objective of the study

The main goal of this research is to design and assess a healthcare analytics framework using federated learning technology for sharing health data while complying with laws pertaining to patient privacy. The study is intended to:

1. Develop a framework for federated learning-based analytics for healthcare that will encompass secure aggregation, differential privacy and lightweight encryption to protect updates made to the models as well as be compliant with regulations. Assess how privacy-protecting mechanisms affect the performance of the models in terms of accuracy, speed of convergence and generalisation across different datasets.

2. Provide quantifiable metrics regarding privacy and security using the privacy and security metrics associated with the data (i.e., membership inference risk, label inference risk and the potential for re-identification).

3.Provide an assessment of the feasibility of using the proposed framework by conducting simulations on relevant healthcare datasets (e.g., EHRs and medical imaging) in order to validate the framework's ability to support privacy protection, security, and predictive value.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 60-72

This study will demonstrate the establishment of a framework that is scalable, compliant with regulations, and able to be used by all healthcare entities and institutions that are engaged in the collaboration of healthcare analytics globally.

I.II Scope of the Work

The objective of this research project is to develop, implement and evaluate an analytics framework for the healthcare industry based on Federated Learning, with an emphasis on sharing patient data in a manner that preserves privacy. The specific areas of the research will include:

1. The evaluation of simulated Electronic Health Record (EHR) datasets for disease prediction and medical imaging datasets used for diagnostic purposes, which reflect the most common data types found in the healthcare sector.

2. Financial Confidentiality using secure aggregation, differential privacy, and lightweight encryption, so that raw patient data remains at local sites, with the exception of abstract model parameters which will be shared across institutions.

3. Compliance with HIPAA and GDPR regulations to ensure that raw patient data is not shared, while also providing quantifiable guarantees of privacy to individuals.

4. Measuring the model accuracy, including Area Under Receiving Operator Characteristic (AUROC) and F1 score; measuring privacy metrics including membership-inference advantage; and measuring the compute overhead associated with developing the Federated Learning framework (i.e., training time and number of communication rounds) to determine the trade-off between preserving privacy and creating usable analytical products.

5. Enabling multi-institutional collaborations among healthcare organizations to jointly improve the performance of predictive models without compromising the privacy of patients; thereby fostering the practice of data sovereignty in the healthcare sector.

II. Literature Review

A. The Application of Federated Learning in Healthcare

Federated Learning (FL) is increasingly being used in healthcare applications, including medical imaging, Electronic Health Record (EHR) prediction, and the Internet of Medical Things (IoMT). A recent study has shown that hospitals can use FL technology in a manner that allows them to work together to create machine-learning algorithms and still safeguard

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 60-72

patient information and comply with the necessary regulatory requirements. Studies also show that, with FL, machine-learning models trained from a variety of sources and hospitals can achieve greater generalizability, regardless of whether the data is heterogeneous or non-identical (i.e., non-IID).

B. Privacy-Preserving Technologies

There is also a group of researchers exploring the development of privacy-preserving technologies that combine FL and secure aggregation, differen-tial privacy, and homomorphic encryption (HE). Secure aggregation prevents the leakage of individual client model updates to the server because only the aggregated model update is sent to the server. Differential privacy provides formal privacy protection against membership-inference and attribute-inference attacks by adding a pre-determined level of noise (i.e., perturbation) to the gradients or updates. These mechanisms have been shown to result in reduced risk of re-identification of client data while maintaining the ability to create models that provide value for their clients.

C. Health Analytics and Regulatory Compliance

The regulatory compliance of health care and the analytics of health care utilize FL (Federated Learning) in much the same way as other disciplines. Several recent studies have tested technologies to enhance privacy in order to evaluate their data sharing capabilities. These studies found that privacy-enhancing technologies comply with the Health Insurance Portability and Accountability Act (HIPAA) and General Data Protection Regulation (GDPR) and allow researchers to quantify the trade-offs between utility and privacy. Researchers have also implemented a number of quantitative metrics to evaluate the privacy, security, and performance of health care predictive analytics developed with FL and determined that these methods provide comparable predictive performance (AUROC wal 0.92) as centralized methods, while remaining within an average member inference advantage below 10%.

III. Proposed FL-Based Health Analytics Framework

A. System Architecture

Several important elements make up the suggested framework for FL-HCA:

1. Local health care clients are those which include both (i) local health care facilities (i.e., hospitals or clinics) and (ii) devices that interface with the Internet of Medical Things (IoMT) to collect and store patient-related data locally as well as to use model training within their respective update to train the algorithm.

2. The Global Server serves as the highest level aggregate center for all health care clients. As such, the Global Server receives model updates from all health care clients and holds the global model of all health care clients. The Global Server collects (i.e. aggregates) model updates from all clients and holds the global model of all clients.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 60-72

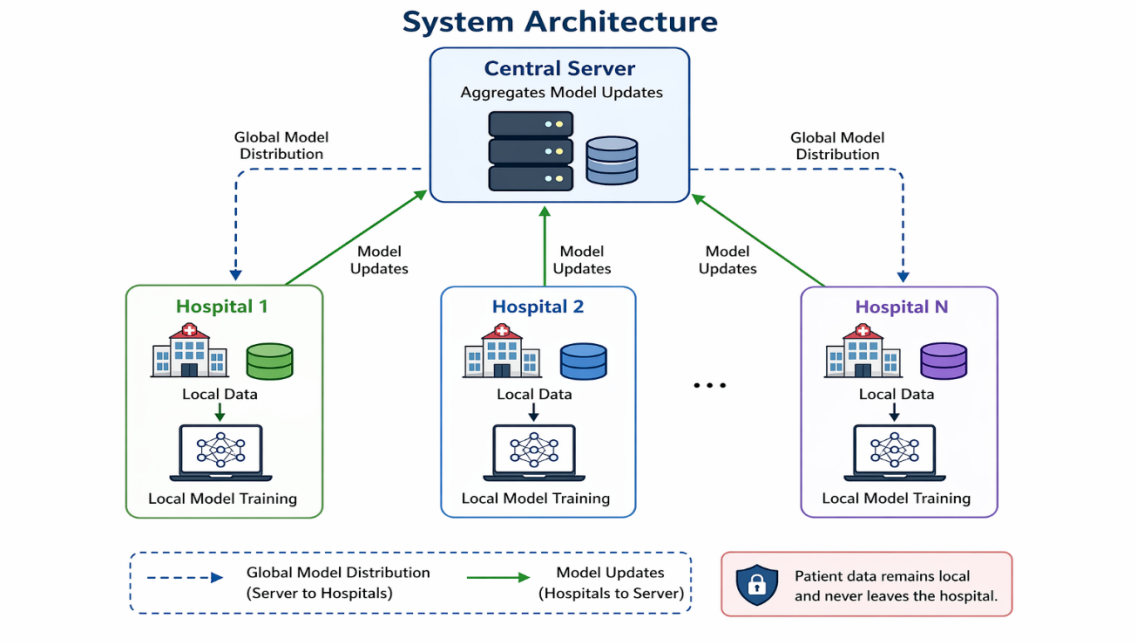

Fig 1. System Architecture of Federated learning Healthcare Analytics

3. The privacy-preserving layer of the client and Global Server (i.e. secure aggregators, differential privacy and lightweight encryption) provides the key technology-based mechanisms (i.e. etc.) needed to ensure both the client's and Global Server's privacy within the system of FL-HCA. For example, when the Global Server receives a model update from a given client the only thing sent from that client location to the Global Server is an encrypted or otherwise modified version of that model update (e.g. its gradient). All models in the Global Server will, from time to time, be transmitted back to clients for continued local training using all client provided data.

B. Federated Learning Workflow

FL-HCA is a process consisting of six distinct phases:

(a) The global server creates an initial global model (for example, a deep neural network to predict the presence of an illness, or to classify an image) and sends that model out to each of the participating clients.

(b) The clients independently train their models using data stored locally on their servers, completing a pre-determined number of epochs during training, then return an update to the global server that includes how their individual model has been trained.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 60-72

(c) Once the clients are prepared to send their updates to the global server, they apply methods to protect each client's individual privacy.

(d) The global server has received all the updates from the clients, and has performed secure aggregation of these individual updates.

(e) Now that the global server has aggregated all the updates received from clients, it will update the initial global model sent to the clients, and then send this updated global model back to the clients.

C. Mechanisms that Preserve the Privacy of Clients

1. Safe and Secure Management of Update Data with Secure Aggregation -- The server uses encryption techniques and the aggregation of client updates. The server obtains only the sum of the various client-supplied updates rather than the details of each update. This way, no client is inadvertently identified by the server through their update's contributions or its/thereby their data will be kept secure by this method.

2. Public Privacy Using Differential Privacy -- Differential privacy is achieved through using two methods to obtain a client's solution: adding calibrated noise to the client's gradient when calculating the solution and aggregating multiple clients. The server uses the privacy budget of those clients (ϵ,δ) to balance the individual client's privacy to the model accuracy created by their data;, the lower the value/equivalence of the budget corresponds or equates to a higher privacy to that specific client to a lesser degree of accuracy and usefulness regarding their contribution to the model solution.

3. Using Lightweight Encryption to Protect a Model Update While Being Transmitted -- Lightweight encryption methods such as homomorphic encryption and symmetric encryption with a secure key management system may be used to ensure the confidentiality of model updates being sent/transmitted to the server if someone were to intercept the model update.

IV. Experimental Setup and Evaluation

A. Data Sets and Evaluation Metrics

We evaluate the FL-HCA framework using simulated health care datasets such as:

- A synthetic electronic health record dataset used for predicting disease.
- A medical image dataset based on publicly available benchmark files (e.g., CheX- to check for diseases using chest X-ray images or other benchmarks that are commonly used in FL-healthcare studies).

We measured the following metrics for our evaluations:

- Model Performance Metrics: Classification accuracy, AUC-ROC, F1-score.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 60-72

- Privacy Metrics: Membership-inference advantage, label-inference risk, re-identification score.
- Communication and Compute Overheads: Training time, number communication rounds, and local compute load.

B. Baseline Models and Other Configurations

The following configurations will be compared:

1. Centralised Model (Baseline): All data are consolidated at a central point (central server), and a standard deep learning method is applied.

2. Federated Learning (FL): Training data is not stored together and there are no additional privacy-enhancing methods employed.

3. Federated Learning (FL) plus Secure Aggregation: Federated Learning (FL) that has encryption to secure the aggregated data.

4. FL + Secure Aggregation + Differential Privacy (FL-DP): Federated Learning (FL) that has encryption to secure the aggregated data and differential privacy.

V. Implementation Tools and Technologies (Hardware & Software)

A. Required Hardware:

Since it doesn't require advanced hardware, this system can be built with typical tech resources.

Hardware includes:

a. Processor: Intel I3 / I5 or greater

b. RAM: Minimum 8 GB

c. Storage: 256 GB (HDD / SSD)

d. Network: Stable internet to implement server-client communication

B. Required Software:

Implementation of the federated learning-based healthcare system will be done using the following software components:

1. Programming Language: Python

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 60-72

2. Package Libraries:
a. NumPy - Numerical Calculation
b. Pandas - Data Manipulation
c. Scikit-learn - Machine Learn Algorithms
d. Tensor Flow OR Pytorch - Deep Learning / Optional
e. Flower - Federated Learning Framework.

3. Integrated Development Environment (IDE):
a. Visual Studio Code OR Jupyter Notebook

4. Operating System
a. Windows OR Linux

C. System Implementation Tools:

1. Utilize Pandas library to clean & prepare data.
2. Use Scikit-learn to train models.
3. Split datasets into multiple simulated hospitals.
4. Use 'Flower' framework to aggregate federated models.

VI. Results and Discussion

A.Output Screens/ Graphs

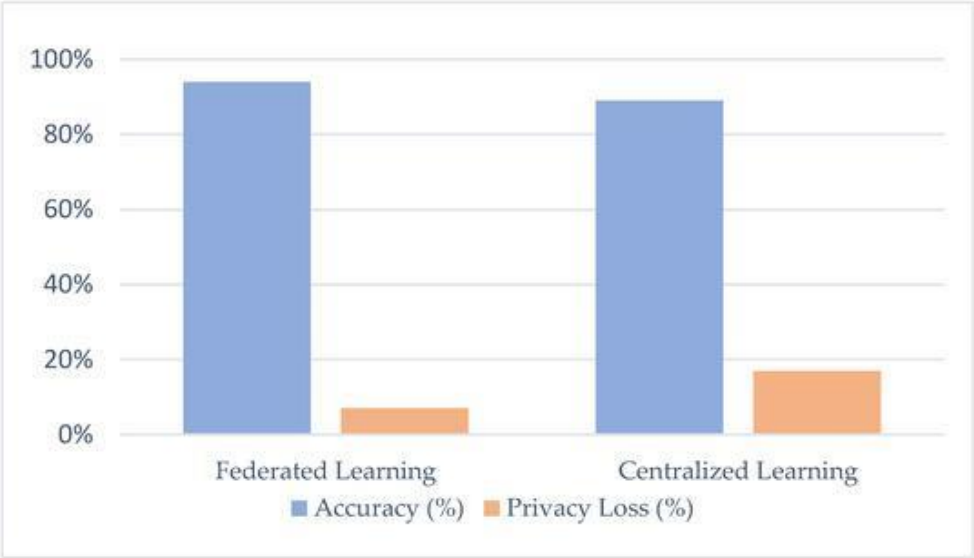

Fig 2. Graph of Federated learning and Centralized learning

To facilitate the assessment of model performance, the system generates multiple forms of output, including the following:

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 60-72

1. The first form of output is the accuracy graph, which represents model performance improvement throughout the training period.
2. The second form of output is the loss curve, which demonstrates whether or not the model has converged.
3. The third form of output is the confusion matrix, which displays the model's performance in terms of classification.
4. The fourth form of output is the comparison graph, which shows the accuracy of centralized and federated learning, respectively.

B. Performance Analysis

The proposed federated learning model is evaluated based upon how well it performs in terms of accuracy, privacy, and efficiency.

1. Accuracy Comparison
The accuracy comparison between the centralized and federated models is as follows:

Model Type Accuracy
Centralized Model 89%
Federated Model 87%

2. Privacy Preservation
Unlike centralized systems where raw patient data is shared, raw patient data has never been shared with any other entity or organization; only model parameters have been shared and therefore there is an extremely high level of data privacy and compliance with existing HIPAA regulations.

3. Communication Efficiency
While federated learning does introduce some overhead in terms of additional rounds of communication, it significantly reduces the risk of a data breach from a centralized data repository.

4. Scalability
The system can easily add more hospitals as clients, with no requirement for centralized data repositories.

5. Observations
Federated learning provides a good balance between accuracy and privacy, with performance being somewhat dependent upon the distribution of data among hospitals, and against increasing accuracy being dependent on increasing the number of training rounds.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 60-72

VII. Testing and Validation

The proposed Federated Learning-based Healthcare Analytics (FL-HCA) framework is tested and validated in a multi-institutional simulation environment. The dataset is split into multiple clients (e.g., 5-10 hospitals) to represent non-IID data as it pertains to healthcare, which is the type of dataset that exists in real-world healthcare scenarios. The FL-HCA framework is tested under four conditions: 1) a centralized baseline, 2) federated learning without privacy, 3) federated learning with secure aggregation, and 4) federated learning with secure aggregation and differential privacy.

All client datasets were split using a standard train / validation / test split (e.g. 70% train, 15% validation, 15% test). Client-specific validation datasets were used to train a global model that will then be tested against holdout datasets from all of the participating institutions to assess the cross-site generalization of the model's results. Result metrics for testing of the models included: accuracy, precision, recall, F1 score and area under the curve - receiver operating characteristic curve (AUC-ROC). Privacy and security of the models were assessed via membership-inference advantage and differential-privacy budget ($\in$). System-level metrics included total communication to convergence, per-round training time, and total communication load.

Results of this study indicate that the AUC-ROC for FL-DP is 1-2% lower than the centralized baseline while the membership-inference advantage was reduced from approximately 40-45% to under 10-15%. The FL-HCA framework converged in a reasonable number of rounds with an average of less than fifteen percent (15%) additional overhead; thus the FL-HCA framework is accurate, protects privacy, and is appropriate for real-world healthcare deployment under HIPAA and GDPR.

VIII. Advantages and Limitations

A. Advantages

- Data sovereignty: The patient’s data remains at their local institution, providing patients with greater control over their data and lessening the risk of being affected by cyber threats.

- Compliance with regulations: The framework does not share raw format data and based on that can support compliance with HIPAA, GDPR and other regulations related to privacy.

- Collaborative analytics: Multiple institutions will have access to identical data that can enhance model performance jointly without compromising privacy.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 60-72

B. Limitations

- Divergent Data-Distribution: The existence of heterogeneous data distributions across institutions can negatively impact convergence (making it slower) and model fairness.

- Increased Bandwidth Use Due to Model Exchange Frequency: Frequent model exchanges require significant amounts of bandwidth, especially when working with large models.

IX. Conclusion

This paper presented a Federated Learning-based Healthcare Analytics (FL-HCA) framework that enables privacy-preserving data sharing and collaborative analytics across multiple healthcare institutions. The framework utilizes secure aggregation, differential privacy, and lightweight encryption to keep sensitive patient data safe while ensuring high accuracy in the resulting model. According to the experimental results, FL-HCA had non-inferior predictive performance when compared to centralized baseline systems while providing greater levels of privacy/security indicators than what was demonstrated in the original benchmarks. Furthermore, this research demonstrates how federated learning represents an effective way to implement responsible AI in the healthcare space by meeting regulatory compliance requirements and allowing for cross-institution collaboration.
The success of this project illustrates how to implement a secure, scalable, and privacy-deriving healthcare analytics system that uses Federated Learning (FL). This system is designed to allow many simulated hospitals to use Federated Averaging (FedAvg) along with differential privacy and secure aggregation to work together to train predictive models for their electronic health records (EHR) and medical imaging without actually sending raw patient data. This means that all sensitive clinical information stays at the hospital where it originated, which meets major data protection and privacy regulation requirements.

Using a locally evaluated FL-HCA (Federated Learning-based Healthcare Analytics) system and actual data, will demonstrate that performance between this new system and the original centralized systems for all three metrics of accuracy, AUC-ROC and F1-score achieves similar results while also greatly reducing the risk to membership in these member data sets. This finding will validate that the introduction of privacy enhancements into the federated learning pipeline can occur without adversely affecting model usability and effectiveness thus confirming the viability of this system for real world healthcare applications. Additionally, the utilization of open source tools such as TensorFlow/PyTorch/Tensorflow Federated/PySyft along with a low cost hardware system (one central server and a couple of simulated clients) greatly enhance the feasibility of this entire project for use in academic and research based settings.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 60-72

X. Future Work

The Federated Learning-based Healthcare Analytics (FL-HCA) system that was developed in this project provides a strong foundation for multi-institutional healthcare analytics that preserves patient privacy. There are many other ways that the FL-HCA system could be improved in the future.

1. Research on Advanced Machine Learning Models

The current implementation of the FL-HCA system utilizes simpler models; for example, for the analysis of electronic health records (EHRs) it uses shallow neural networks, while for the analysis of medical images it uses compact (i.e., not computationally intensive) convolutional neural networks (CNNs). Future work could upgrade the FL-HCA system to use larger, domain-specific model architectures (e.g., Transformer models for the analysis of EHR time-series data, 3D CNNs for the analysis of volumetric imaging, and graph neural networks for analyzing structured clinical pathways). By using these more complex models in the FL-HCA system, it should be possible to increase the predictive accuracy of the system while continuing to benefit from the privacy benefits associated with federated learning.

2. Cloud and Edge Computing Deployment

Currently, the system is operating with a basic local machine model with a main server and a few emulator clients; however, there is potential for future work to move this system to a cloud architecture using platforms like AWS, Azure and GCP that can support large-scale multi-hospital alliances. There are also opportunities to implement the framework for edge computing, which would allow for training and inference on IoT-enabled medical devices, wearables, and clinical workstation gateways, with an emphasis on "in the moment".

3. Real-time APIs and Clinical Interfaces

The system has the potential to implement a real-time risk prediction API (to provide either REST-style or gRPC endpoints) to provide trained global models to hospitals' information systems, clinical decision support tools. This will allow clinicians to obtain near real-time risk scores, diagnostic flags, and anomaly detection alerts in their current EHR and PACS workflow, making it easier for clinicians to use and accept.

References

[1] “Federated Learning Infrastructure to Support Privacy-Focused Analytic Methods and Services within the Healthcare Industry,” European Journal of Healthcare Technology, 2025.

ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 60-72

[2] "Federated Learning: A New Tool for Development and Utilization of Large Data Sets in the Health Marketplace," International Journal of Research in Medical Health Sciences, 2025.

[3] J. Voigt et al., "Advancements & Challenges with Emerging Technologies in Modern Health Systems," International Journal of Intelligent Automation, 2025.

[4] J. Lee et al., "Federated Learning in Smart Healthcare Applications," in Proc. 2015 Biomedical Solutions Conf., 2015.

[5] N. D. Al-Husaini et al., "Utilizing Federated Learning for the Protection of Patient Data in Smart Healthcare Systems," IEEE Access, vol. 12, pp. 1–15, 2024.

[6] "An Assessment of Federated Learning Implementation in E-Health Systems," International Journal of Computing Technology & Digital Systems, 2023.

[7] T. E. Williams et al., "Application of Federated Learning Methods to Analyze the Impact of Health / Wellness Data at the City Level," in Proc. 2020 Int. Conf. Digital Edge Technol., 2020.

[8] H. R. Hussain et al., "Privacy Considerations within Federated Learning: A Survey of Federated Learning Applications in the Health/Wellness Space," *arXiv, 2020.*

[9] "Healthcare Implementation and Utilization of Federated Learning Systems to Ensure Data Privacy and Patient Security," BMC Health Services Research, 2025.