



International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

Confidential Federated Learning Framework for Predictive Healthcare Analytics Using IoT

¹Dr. A. Kanimozhi, ^{2*}Dr. S. Anusya, ³Jothiprakash S.P., ⁴P.V.V.S.D. Nagendrudu, ⁵Harprabdeep Singh, ⁶Mrs.V. Revathi, ⁷Dr. T. M. Saravanan, ⁸Dr. G. Jenitha

¹Associate Professor, Department of Mathematics, NPR College of Engineering and Technology, Dindigul, Tamilnadu, India, 624 401.
E- mail: akanimozhi@nprcolleges.org

²Assistant Professor, Department of School of Computer Studies ,A.V.P.College of Arts and Science (Co-ed) (Autonomous), T.M.Poondi, Tirupur, Tamil Nadu, India,641 652. E- mail id: anushyapricilla30@gmail.com

³Assistant Professor (SG), Department of Mathematics,Nehru Institute of Engineering and Technology, T.M.Palayam, Coimbatore, Tamilnadu, India, 641105. E-mail: spjpmath@gmail.com

⁴Assistant Professor, Department of Artificial Intelligence and Machine Learning,Aditya University, Surampalem, Andhra Pradesh, India,533437. E- mail id: nagendrudu2016@gmail.com

⁵Assistant professor, Department of Computer Science and IT,BUC College, Batala, Punjab, India, Pincode:143505.
E-mail: harprabdeep@yahoo.co.in

⁶Assistant Professor, Department of Computer Science , Dr.N.G.P. Arts and Science College , Coimbatore , Tamilnadu, India, 641048.
E- mail: revathi.v@drngpasc.ac.in

⁷Associate Professor, Department of Computer Technology, Kongu Engineering College,Perundurai, Erode, TamilNadu, India,638060.
E- mail id: tmskec@gmail.com

⁸Associate Professor, Department of Mathematics, AMET deemed to be University, Chennai, Tamilnadu, india, 603112.
E- mail: jenitha.g@ametuniv.ac.in

Abstract

The rapid growth of the Internet of Things (IoT) in healthcare has enabled continuous acquisition of physiological and clinical information through wearable sensors and connected medical devices, creating new opportunities for predictive healthcare analytics. However, the centralized collection and processing of such sensitive information raises significant concerns regarding patient privacy, data security and institutional data ownership. This study proposes a confidential federated learning framework for IoT-enabled cardiovascular risk prediction using secondary healthcare data. A publicly available cardiovascular disease dataset is preprocessed and distributed among multiple virtual healthcare clients to simulate heterogeneous healthcare institutions. Each client independently trains a neural-network-based cardiovascular risk prediction model using locally retained data, while only model updates are communicated to the federated server. To strengthen confidentiality, gradient clipping and differential-privacy noise are incorporated before model updates are transmitted for federated aggregation. The global model is generated using weighted Federated Averaging and iteratively redistributed to participating clients. The framework is intended to demonstrate that collaborative cardiovascular-risk prediction can be achieved without direct exchange of raw patient information while maintaining an appropriate balance between predictive performance, confidentiality and communication efficiency. The proposed approach provides a computational foundation for privacy-preserving predictive healthcare analytics in distributed IoT environments.

Keywords: Federated Learning; Internet of Things; Internet of Medical Things; Cardiovascular Disease Prediction; Privacy-Preserving Machine Learning; Differential Privacy; Healthcare Analytics; Non-IID Data; Secure Aggregation; Predictive Analytics.

1. Introduction

The rapid development of the Internet of Things (IoT) has transformed healthcare by enabling continuous monitoring of patients through wearable sensors, smart medical devices and remote-monitoring systems. Physiological parameters such as heart rate, blood pressure, electrocardiographic signals, oxygen saturation and other clinical measurements can be collected at frequent intervals and used for early prediction of cardiovascular abnormalities. Although this development provides considerable opportunities for predictive healthcare analytics, it also generates large volumes of highly sensitive patient information. Centralizing such information at a cloud or institutional server creates significant concerns regarding privacy, unauthorized access, data breaches and secure transmission.

Conventional machine-learning approaches generally require patient data from different healthcare institutions to be transferred to a common server. Although centralized learning can achieve high predictive

performance because the model has access to a large combined dataset, it conflicts with the distributed and privacy-sensitive nature of healthcare data. Healthcare institutions may also be reluctant or unable to exchange patient-level information because of privacy, ethical and regulatory constraints.

Federated Learning (FL) provides an alternative paradigm in which healthcare institutions collaboratively train a common model while retaining their raw data locally. However, federated learning should not be regarded as complete privacy protection. Model parameters or gradients exchanged during the federated process may still contain information related to the underlying training records. Furthermore, healthcare datasets are commonly heterogeneous and non-independent-and-identically-distributed (non-IID), since hospitals serve populations with different demographic characteristics, disease prevalence and clinical profiles. These characteristics can negatively affect federated convergence and prediction performance.

Therefore, there is a need for a framework that simultaneously addresses predictive accuracy, patient-data confidentiality, heterogeneous healthcare data and communication constraints. Motivated by this requirement, the present study proposes a confidential federated-learning framework for IoT-enabled cardiovascular risk prediction using publicly available secondary healthcare data. The framework combines local predictive learning, federated aggregation and differential privacy to investigate whether collaborative healthcare intelligence can be achieved without requiring centralized access to raw patient information.

Federated Learning (FL) has emerged as an alternative to conventional centralized machine learning in situations where training data are distributed across multiple devices or organizations. McMahan et al. (2017) introduced the Federated Averaging (FedAvg) approach, in which participating clients perform local model training and transmit model parameters to a coordinating server rather than transferring their raw data. Their work demonstrated that iterative model averaging can substantially reduce communication requirements while accommodating unbalanced and non-IID data distributions. Kairouz et al. (2021) subsequently provided a comprehensive treatment of federated learning, covering optimization, communication efficiency, privacy, security, personalization, fairness and heterogeneous data. Their work emphasized that although FL reduces the need for centralized data collection, important challenges remain in client heterogeneity, communication efficiency, privacy leakage, and robustness. These foundational studies establish FL as an appropriate paradigm for healthcare environments in which data ownership is distributed across hospitals, clinics, laboratories and personal medical devices.

The healthcare domain has become one of the important application areas of FL because medical information is naturally distributed among institutions and is subject to strict privacy requirements. Rieke et al. (2020) discussed the potential of FL for digital health and highlighted the problem of healthcare data silos. Their work emphasized that valuable medical datasets often remain isolated because of privacy, legal and organizational constraints. FL can facilitate collaborative model development while allowing healthcare institutions to retain their data locally. Brisimi et al. (2018) demonstrated the use of federated learning for predictive modeling using distributed electronic health records. Their study showed that healthcare institutions could collaboratively develop predictive models without exchanging raw EHR data. The study is particularly relevant to the present research because it demonstrates the application of FL to healthcare risk prediction rather than limiting FL to image-based applications.

Sheller et al. (2020) investigated multi-institutional federated learning in medicine and showed that models trained across ten institutions could achieve approximately 99% of the quality obtained using centrally pooled data in their experimental setting. The study also examined generalization across institutions and demonstrated the potential benefit of accessing diverse distributed datasets without requiring direct patient-data sharing. Similarly, Dayan et al. (2021) developed the EXAM federated model using data from 20 institutions to predict future oxygen requirements in patients with COVID-19. Their federated model achieved an average AUC above 0.92 for predictions at 24 and 72 hours and improved performance and generalizability relative to models trained at individual institutions. This provides strong evidence that federated learning can exploit multi-institutional diversity for clinically relevant prediction tasks.

Although FL avoids direct exchange of raw patient records, federated learning alone should not be interpreted as providing complete confidentiality. Model parameters, gradients and intermediate updates may potentially reveal information about local training data. Kaissis et al. (2020) reviewed secure and privacy-preserving federated machine learning for medical applications and emphasized that privacy protection should be considered alongside the development of distributed models. They discussed privacy threats and technical approaches including secure aggregation, cryptographic techniques and other privacy-preserving mechanisms. Differential privacy provides a mathematical mechanism for limiting the influence of individual records on released information. Dwork and Roth (2014) established the theoretical foundations of differential privacy and formalized privacy guarantees using parameters such as ϵ and δ . Abadi et al. (2016) subsequently demonstrated the incorporation of differential privacy into deep-learning systems through a practical

mechanism based on noisy gradient updates. Their work showed that neural networks could be trained under a formal privacy budget while maintaining useful predictive performance. For the proposed study, these developments motivate the integration of differential privacy with federated learning so that the framework provides stronger confidentiality than FL alone.

Another important privacy mechanism is secure aggregation. Bonawitz et al. (2017) proposed a practical secure-aggregation protocol that enables a server to obtain an aggregate of client-held model updates without directly observing individual contributions. Their approach was designed to be communication-efficient and robust to client dropouts, making it relevant to distributed learning environments. Secure aggregation and differential privacy address somewhat different aspects of confidentiality. Secure aggregation can prevent the server from inspecting individual updates, whereas differential privacy limits the information that can be inferred from the resulting model or released statistics. Consequently, a stronger confidential FL framework can potentially combine these mechanisms. The growth of the Internet of Medical Things (IoMT) has further increased the relevance of federated learning. Wearable sensors, smart watches, ECG devices, blood-pressure monitors and other connected medical devices can continuously generate physiological data. This creates an important tension between data utilization and privacy. Recent healthcare FL literature identifies wearable and biosignal data as an emerging application area, while also emphasizing challenges arising from heterogeneous data, limited computational resources, bandwidth constraints and the need for low-latency prediction. A 2026 survey of privacy-preserving FL in digital healthcare similarly highlights the growing role of the Internet of Medical Things and the continuing need for stronger privacy and security mechanisms in healthcare ML. This provides the motivation for connecting the IoT data-acquisition layer with the federated predictive-learning layer in the present framework.

Cardiovascular disease (CVD) is an especially appropriate application for privacy-preserving predictive analytics because risk assessment commonly involves sensitive demographic, physiological and clinical information.

Federated learning can facilitate collaborative cardiovascular prediction by allowing institutions to learn from diverse patient populations without directly pooling patient-level data. A recent comparative review identified differences between biosignal-based and EHR-based FL. Biosignal approaches increasingly address personalized monitoring and non-IID data, whereas EHR-oriented approaches emphasize multi-hospital collaboration, scalable optimization and privacy-preserving aggregation. More recently, a systematic review of cardiovascular FL studies published between 2022 and 2025 identified 22 studies covering early screening, diagnosis, prognosis and decision-support applications. The review found that many studies use horizontal FL and FedAvg-type approaches, with increasing attention to non-IID data, personalization and efficiency. However, it also found that much of the evidence remains retrospective and that reporting of external validation, calibration, subgroup performance, privacy safeguards, robustness and system costs is inconsistent. This recent evidence strongly supports the need for a more comprehensive evaluation framework rather than reporting predictive accuracy alone.

McMahan et al. (2017) already identified unbalanced and non-IID data as defining characteristics of decentralized learning. Later healthcare studies have demonstrated the importance of institutional variation and generalization across sites. For example, Sheller et al. (2020) explicitly examined the influence of data distribution across institutions, while Dayan et al. (2021) demonstrated the benefit of multi-institutional diversity in clinical prediction. Therefore, a cardiovascular FL framework should evaluate performance under heterogeneous client distributions rather than relying exclusively on random IID partitions.

The integration of differential privacy introduces an important trade-off. Increasing the privacy noise can provide stronger privacy protection but may decrease model utility. Therefore, the objective is not simply to maximize privacy but to identify a suitable balance. Abadi et al. (2016) demonstrated that useful deep-learning models can be trained under differential privacy, while the broader privacy literature emphasizes that privacy guarantees must be explicitly quantified rather than assumed. In healthcare, this trade-off is particularly important because excessive degradation in sensitivity or specificity may reduce the usefulness of a predictive model. Hence, privacy-preserving cardiovascular prediction should evaluate both privacy parameters and clinical-performance metrics.

The foundational FL literature established decentralized model training and FedAvg. Healthcare studies subsequently demonstrated that institutions can collaboratively develop predictive models without directly sharing patient records. Privacy research introduced mechanisms such as differential privacy and secure aggregation to address the limitations of FL alone. More recent cardiovascular studies have expanded FL to ECG, EHR, imaging and wearable data. However, the current literature still reveals several unresolved issues. First, many studies rely on retrospective public datasets and simulated client partitions rather than genuine multi-institutional prospective deployments. Second, privacy protection is inconsistently implemented or

reported. Third, the effect of non-IID healthcare data is not uniformly addressed. Fourth, predictive accuracy is frequently emphasized over communication cost, calibration, subgroup performance and privacy-utility analysis. The latest cardiovascular systematic review specifically identifies held-out-site validation, privacy/security safeguards, robustness, communication-efficient personalization and clinical-grade monitoring as important priorities for future work.

Although machine learning, IoT-based healthcare monitoring and federated learning have individually received considerable attention, an important research gap remains in the development of a unified privacy-preserving framework capable of addressing the simultaneous requirements of predictive accuracy, patient-data confidentiality, heterogeneous healthcare data and IoT communication constraints. Existing centralized predictive models require the collection of sensitive patient information at a common location, whereas conventional federated-learning approaches, despite avoiding direct raw-data sharing, may remain vulnerable to information leakage through exchanged model updates. Furthermore, the heterogeneous and non-IID nature of healthcare data across institutions can adversely affect federated optimization and generalization. The privacy-utility trade-off introduced by differential privacy has also not been sufficiently integrated with predictive-performance and communication-efficiency analysis in a unified IoT healthcare setting. These limitations motivate the development of a confidential federated-learning framework that combines distributed local learning, privacy-preserving model updates and federated aggregation for cardiovascular risk prediction using secondary healthcare data.

The objectives should directly address the identified gaps.

The primary objective of this study is to develop a confidential federated-learning framework for IoT-enabled predictive healthcare analytics that enables collaborative cardiovascular-risk prediction while preserving patient-data confidentiality.

- To develop an IoT-oriented predictive healthcare architecture in which physiological and clinical observations are represented as distributed patient data.
- To preprocess and transform publicly available secondary cardiovascular healthcare data into suitable datasets for distributed machine-learning experiments.
- To simulate multiple healthcare institutions as independent federated clients with heterogeneous/non-IID patient-data distributions.
- To develop local cardiovascular-risk prediction models at individual healthcare clients without transferring raw patient data to a central server.
- To implement federated learning using weighted Federated Averaging for collaborative global-model development.
- To incorporate differential privacy through model-update clipping and calibrated noise addition to strengthen the confidentiality of exchanged information.
- To investigate the effect of different privacy levels on predictive performance and establish the privacy-utility trade-off.
- To compare centralized learning, independent local learning, conventional federated learning and confidential federated learning using accuracy, precision, recall, F1-score, specificity and ROC-AUC.
- To analyze federated convergence, communication overhead and computational requirements in the proposed IoT healthcare environment.
- To determine whether confidential federated learning can provide an effective balance between predictive accuracy, privacy preservation and communication efficiency.

The main novelty is not simply using federated learning for cardiovascular prediction. Rather, it is the integrated investigation of:

1. distributed IoT-oriented healthcare data,
2. heterogeneous federated clients,
3. privacy-protected model updates,
4. collaborative cardiovascular-risk prediction, and
5. simultaneous evaluation of predictive utility, confidentiality and communication efficiency.

2. Preliminary Concepts

2.1 Internet of Things (IoT) in Healthcare

The Internet of Things (IoT) refers to a network of interconnected physical devices capable of sensing, processing and transmitting information through communication networks. In healthcare, IoT technologies enable continuous or periodic monitoring of physiological parameters using wearable and medical sensors.

Typical healthcare IoT devices include:

- ECG sensors
- heart-rate monitors
- blood-pressure sensors
- pulse oximeters
- glucose monitors
- temperature sensors
- smart watches and wearable devices.

The measurements obtained from these devices can be represented by a feature vector $X_i = (x_{i1}, x_{i2}, \dots, x_{ip})$, where x_{ij} denotes the j -th physiological or clinical measurement of patient i .

The corresponding health outcome is denoted by

$Y_i \in \{0, 1\}$, where $Y_i = 1$ represents the presence of cardiovascular disease and $Y_i = 0$ represents its absence. Thus, an IoT-enabled predictive healthcare system can be expressed as

$X_i \rightarrow \text{Machine Learning Model} \rightarrow P(Y_i = 1 | X_i)$

2.2 Predictive Healthcare Analytics

Predictive healthcare analytics uses statistical and machine-learning techniques to identify patterns in patient data and predict future or currently unknown health outcomes.

For cardiovascular disease prediction, the objective is to construct a function

$f_\theta: \mathbb{R}^p \rightarrow [0, 1]$, such that $y^i = f_\theta(X_i)$ represents the predicted probability that patient i has cardiovascular disease.

The classification rule can be written as

$Y^i = 1, y^i \geq \tau$,

$0, y^i < \tau$, where τ is the selected classification threshold, commonly 0.5 for a binary classifier.

2.3 Secondary Healthcare Data

Secondary data refer to healthcare information that has already been collected and made available through sources such as public repositories, research databases and previously conducted studies.

For the proposed study, a cardiovascular disease dataset can be used containing variables such as: Age, Sex, Blood Pressure, Cholesterol, ECG, Heart Rate, ... together with a cardiovascular disease outcome.

The general dataset can be expressed as

$D = \{(X_i, Y_i)\}$ here $i = 1$ to N , where

- N = total number of observations,
- X_i = input feature vector,
- Y_i = target variable.

An important distinction is that the secondary dataset is used as the experimental data source, while the IoT and multi-hospital architecture is simulated computationally.

2.4 Machine Learning

Machine learning enables a computational system to learn a relationship between input features and an output variable.

For the proposed binary classification problem, $X_i \rightarrow Y_i$.

A model with parameters θ is trained by minimizing an objective function: $\theta^* = \arg \min_{\theta} L(\theta)$.

2.5 Artificial Neural Network

A feed-forward artificial neural network (ANN) can be used as the predictive model.

For a simple neural network,

$h(1) = \sigma(W(1)X + b(1))$, and

$h(2) = \sigma(W(2)h(1) + b(2))$.

The final output is

$Y^i = \sigma(W(3)h(2) + b(3))$, where

- W = weight matrices,
- b = bias vectors,
- σ = activation function.

2.6 Distributed Learning

DISSEMINATION OF KNOWLEDGE

In conventional centralized machine learning, all observations are collected in a single location:

$D = D_1 \cup D_2 \cup \dots \cup D_K$.

A central model is then trained using the complete dataset.

In distributed healthcare, however, different institutions possess different datasets:

$D_1, D_2, \dots, D_K$.

Transferring all these datasets to a central server may create privacy, security and regulatory concerns. This motivates federated learning.

2.7 Federated Learning

Federated Learning (FL) is a distributed machine-learning paradigm in which multiple clients collaboratively train a common model while keeping their raw data locally.

For K healthcare clients:

$D_k = \{(X_i, Y_i)\}_{i=1}^{n_k}, k=1, 2, \dots, K$. The global objective can be formulated as

$$\min_{\{\theta\}} F(\theta) = \sum_{k=1}^K \frac{n_k}{N} F_k(\theta),$$

3. Methodology

The proposed methodology develops a privacy-preserving federated-learning framework for IoT-enabled cardiovascular risk prediction using secondary healthcare data. A publicly available cardiovascular dataset is first selected and subjected to preprocessing operations including missing-value treatment, categorical encoding, feature normalization and outcome-label preparation. The processed observations are subsequently interpreted within a simulated IoT healthcare environment, where physiological and clinical attributes represent information that could be obtained through wearable or remote patient-monitoring devices.

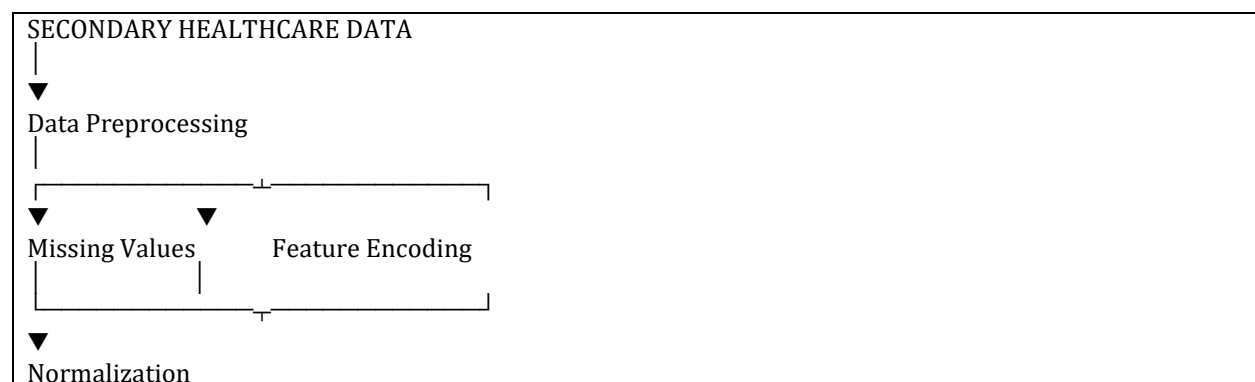
To reproduce a distributed healthcare environment, the dataset is divided into multiple non-overlapping client datasets representing independent healthcare institutions. Since the secondary dataset does not necessarily contain actual hospital identifiers, these institutions are treated as virtual federated clients rather than actual data sources. Both balanced and heterogeneous non-IID client distributions can be considered to investigate the effect of healthcare data heterogeneity on federated learning.

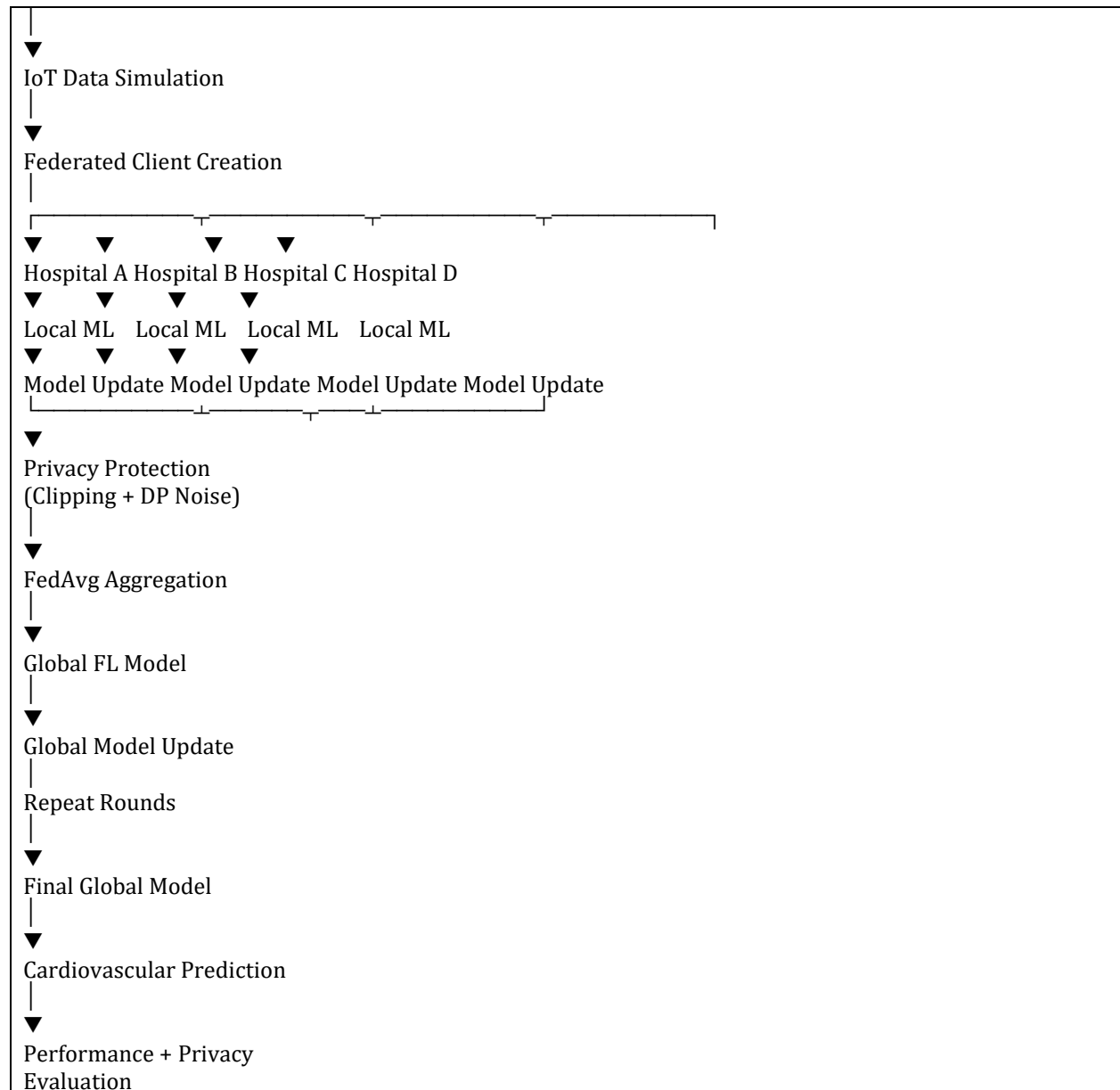
Each client independently trains a neural-network-based cardiovascular-risk classifier using only its locally stored observations. The binary cross-entropy function is employed as the local optimization objective. After local training, the client does not transmit patient records to the central server. Instead, only the locally trained model parameters or parameter updates are prepared for communication.

To strengthen confidentiality, the local model updates are subjected to gradient clipping followed by the addition of calibrated differential-privacy noise. The resulting privacy-preserving updates are transmitted to the federated server, where they are combined using weighted Federated Averaging. The aggregated global model is subsequently distributed back to the participating clients, and the local training and aggregation process is repeated for a predetermined number of communication rounds or until convergence is achieved.

The final global model is used to predict cardiovascular risk for previously unseen patient observations. The proposed confidential federated-learning framework is evaluated against centralized machine learning, independent local learning and conventional federated learning. Accuracy, precision, recall, F1-score, specificity and ROC-AUC are used to assess predictive performance, while training loss and accuracy over communication rounds are used to examine convergence. In addition, communication overhead and computational requirements are measured to assess the suitability of the framework for IoT-enabled healthcare environments.

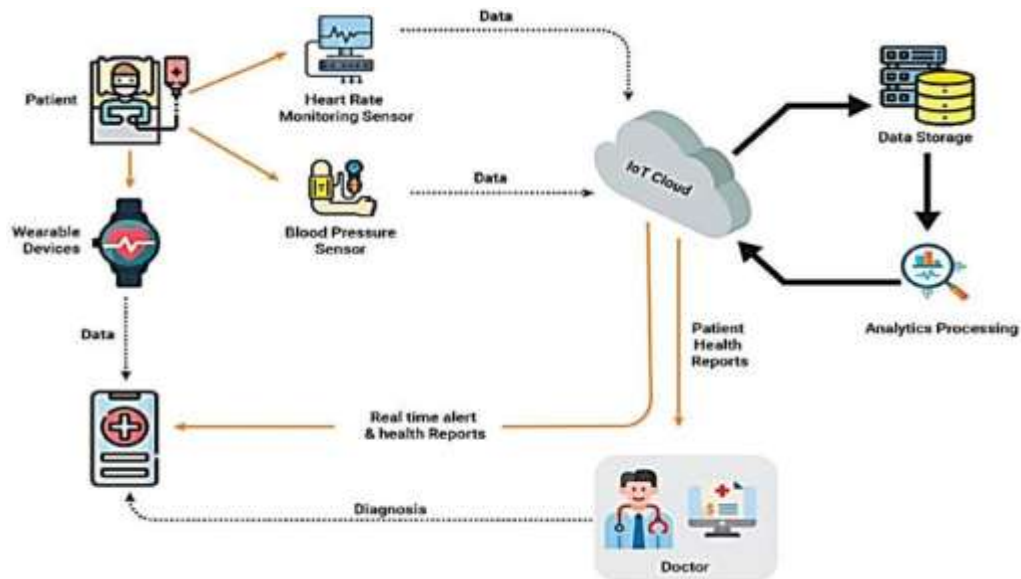
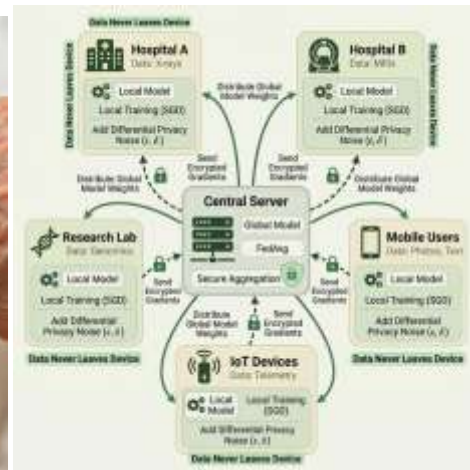
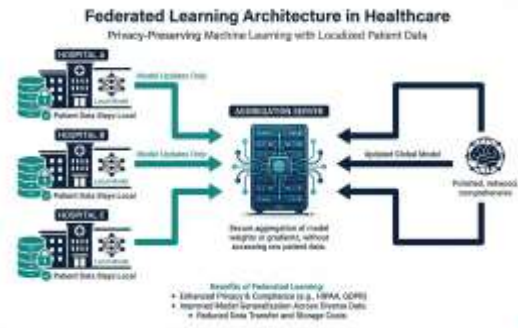
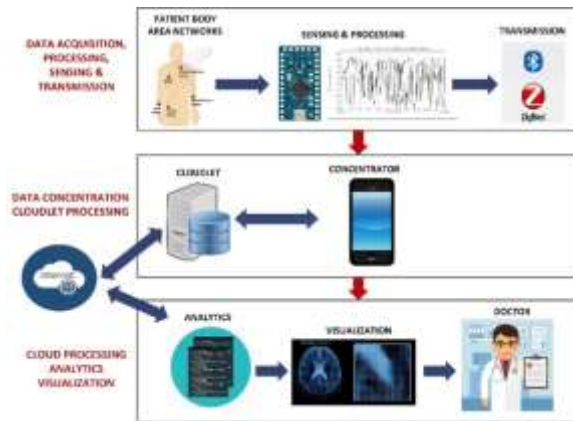
A privacy-utility analysis is finally conducted by varying the differential-privacy noise level. The relationship between privacy protection and predictive performance is investigated to identify an appropriate operating point at which patient confidentiality is enhanced without causing an unacceptable reduction in prediction accuracy. The overall methodology therefore evaluates the proposed framework from three complementary perspectives: predictive effectiveness, privacy preservation and federated communication efficiency.

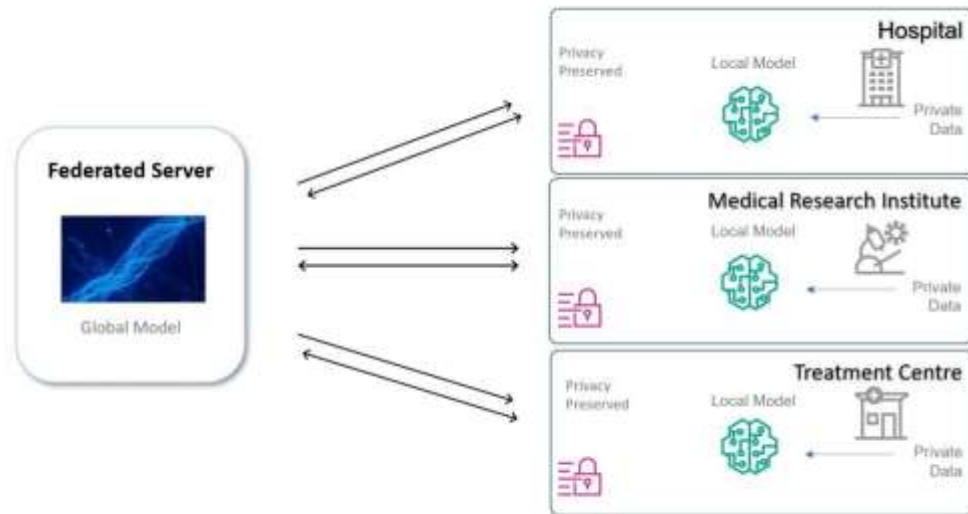




4. Example Case Study

The study can simulate a network of multiple hospitals/healthcare centres, where each hospital has its own patient records generated from IoT-enabled monitoring devices. Instead of transferring patient records to a central server, each hospital trains a local machine-learning model and shares only model parameters/updates. A suitable secondary dataset is the UCI Heart Disease dataset. You can supplement or replace it with a larger public dataset such as PhysioNet if you want a stronger IoT component.





1. Research scenario

Consider four healthcare institutions:

- Hospital A – Urban
- Hospital B – Semi-urban
- Hospital C – Rural
- Hospital D – Specialist cardiac centre

Each institution has IoT-enabled devices such as:

- wearable ECG sensors
- heart-rate monitors
- blood-pressure monitors
- pulse oximeters
- temperature sensors
- glucose monitoring devices

The devices continuously or periodically collect physiological information.

However, patient-level data remain at the hospital/device level because transferring them to a central database creates privacy and security concerns.

The objective is to predict:

Whether a patient is at high risk of cardiovascular disease.

2. Secondary data source

Option 1 — UCI Heart Disease Dataset

The UCI Heart Disease dataset is convenient for your first implementation because it contains clinical variables such as:

Variable	Meaning
Age	Patient age
Sex	Biological sex
CP	Chest-pain type
RestBP	Resting blood pressure
Chol	Serum cholesterol
FBS	Fasting blood sugar
RestECG	Resting ECG result
Thalach	Maximum heart rate
Exang	Exercise-induced angina
Oldpeak	ST depression
Slope	ST-segment slope
CA	Number of major vessels
Thal	Thalassemia
Target	Heart-disease outcome

This dataset is particularly useful for demonstrating the federated-learning methodology,

3. How to convert secondary data into a federated-learning case study

Suppose your dataset contains 1,000 patient records.

You divide the data into four virtual healthcare clients:

Client	Simulated healthcare centre	Patients
Client 1	Hospital A	250
Client 2	Hospital B	250
Client 3	Hospital C	250
Client 4	Hospital D	250

4. Proposed framework

Your framework can contain five layers.

Layer 1 — IoT sensing layer

Simulated patient-monitoring devices provide:

$X_i = [\text{Age}, \text{BP}, \text{HR}, \text{SpO}_2, \text{ECG}, \text{Glucose}, \text{Cholesterol}, \dots]$ where X_i represents the feature vector of patient i . For UCI data, you can treat selected physiological variables as IoT-observable features and the remaining clinical variables as supplementary electronic-health-record features.

Layer 2 — Local healthcare clients

Each hospital receives only its own dataset: $D_k = \{(x_i, y_i)\}_{i=1}^{n_k}$

where $k=1, 2, \dots, K$, K represents the healthcare institution.

Each hospital trains a local model.

For example: $\hat{y} = f_{\{\theta_k\}}(x)$

where

- x = patient features
- y = cardiovascular disease status
- θ_k = local model parameters.

5. Local machine-learning model

You could use a relatively simple neural network initially.

For example:

Input → Dense(32) → ReLU → Dense(16) → ReLU → Dense(1) → Sigmoid

The output is: $P(Y=1|X)$ where $Y=1$ indicates high cardiovascular-disease risk.

6. Federated learning

Instead of sending patient records to a central server:

Hospital → Local training → Model update → Federated server

7. Confidentiality enhancement

Since your title specifically contains “Confidential”, I would not stop at ordinary federated learning. A practical framework is: Federated Learning + Differential Privacy

If you want, I can next build the complete mathematical model + methodology + secondary dataset selection + four-client federated-learning experiment + Python/Google Colab code for this cardiovascular case study.

Interpretation of the Case Study

The proposed case study considers IoT-enabled cardiovascular risk prediction in a distributed healthcare environment. Publicly available secondary healthcare data are used to represent patient observations, while four virtual healthcare institutions are created to simulate hospitals with independent databases. The purpose is to investigate whether healthcare institutions can collaboratively develop an accurate predictive model without sharing their patients' raw medical information.

Interpretation

In a real healthcare environment, patients may be continuously monitored through IoT devices such as wearable ECG sensors, heart-rate monitors, blood-pressure monitors and pulse-oximetry devices. The IoT layer therefore acts as the data-generation and data-monitoring layer. In the proposed secondary-data study, these measurements are represented by variables available in the selected public dataset. Therefore, the study demonstrates how such a system could operate in an IoT healthcare environment rather than claiming that the public dataset itself was collected from IoT devices.

5. Conclusion

This study proposed a confidential federated learning framework for IoT-enabled cardiovascular risk prediction using secondary healthcare data. The framework was designed to address the limitations associated with centralized healthcare analytics, particularly the requirement to transfer sensitive patient information to a common data repository. By distributing the data among virtual healthcare clients and performing model training locally, the proposed approach enables collaborative learning while retaining patient-level information at its original location.

The integration of federated learning with differential privacy provides an additional confidentiality layer. Local model updates are clipped and perturbed before transmission, thereby reducing the possibility of extracting sensitive information from the exchanged updates. Weighted Federated Averaging is then employed to combine the privacy-protected contributions and construct a global predictive model. This architecture demonstrates how distributed healthcare institutions can contribute to a common predictive model without directly sharing their raw patient records.

The comparative experimental design involving centralized learning, independent local learning, conventional federated learning and confidential federated learning provides a comprehensive basis for evaluating the proposed approach. In particular, the comparison between local and federated models can demonstrate the benefit of collaborative learning, while comparison between conventional and privacy-preserving federated learning can quantify the effect of confidentiality mechanisms on predictive utility. The analysis of heterogeneous non-IID client distributions further provides insight into the challenges associated with realistic healthcare environments.

An important component of the proposed framework is the privacy-utility analysis. Differential privacy introduces a controllable trade-off between confidentiality and predictive performance. Therefore, rather than pursuing maximum privacy at the expense of model utility, the framework seeks an appropriate operating point at which privacy protection is sufficiently strong while cardiovascular-risk prediction remains reliable. The inclusion of convergence and communication-cost analysis also extends the evaluation beyond conventional classification metrics and addresses the practical requirements of IoT-enabled distributed healthcare systems.

Overall, the proposed framework provides a unified computational approach combining IoT-oriented healthcare data, federated learning, differential privacy and predictive analytics. It demonstrates the potential of confidential collaborative learning as an alternative to centralized healthcare machine learning. However, the use of secondary data and simulated healthcare clients represents an important limitation. Future investigations should therefore validate the framework using genuine multi-institutional and longitudinal healthcare datasets, real IoT/wearable sensor streams and external test institutions. Further research can also investigate secure aggregation, personalized federated learning, adaptive privacy budgets, communication-efficient optimization and explainable artificial intelligence. Such developments could improve the practical applicability of confidential federated learning for real-world cardiovascular monitoring and privacy-preserving clinical decision-support systems.

This version is deliberately written without claiming experimental improvement or specific accuracy values. Once we run the actual secondary dataset through the four models, we can revise the abstract and conclusion to include the actual accuracy, F1, AUC, privacy level, convergence rounds, and communication results, which will make the paper substantially stronger.

References

- [1]. Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 308–318. <https://doi.org/10.1145/2976749.2978318>
- [2]. Bonawitz, K. A., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., Ramage, D., Segal, A., & Seth, K. (2017). Practical secure aggregation for privacy-preserving machine learning. *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*.
- [3]. Brisimi, T. S., Chen, R., Mela, T., Olshevsky, A., Paschalidis, I. C., & Shi, W. (2018). Federated learning of predictive models from federated electronic health records. *International Journal of Medical Informatics*, 112, 59–67. <https://doi.org/10.1016/j.ijmedinf.2018.01.007>

- [4]. Dayan, I., Roth, H. R., Zhong, A., Harouni, A., Gentili, A., Abidin, A. Z., Liu, A., Costa, A. B., Wood, B. J., Tsai, C.-S., et al. (2021). Federated learning for predicting clinical outcomes in patients with COVID-19. *Nature Medicine*, 27, 1735–1743. <https://doi.org/10.1038/s41591-021-01506-3>
- [5]. Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3–4), 211–407. <https://doi.org/10.1561/04000000042>
- [6]. Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., et al. (2021). Advances and open problems in federated learning. *Foundations and Trends in Machine Learning*, 14(1–2), 1–210. <https://doi.org/10.1561/22000000083>
- [7]. 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, 305–311. <https://doi.org/10.1038/s42256-020-0186-1>
- [8]. McMahan, B., Moore, E., Ramage, D., Hampson, S., & Arcas, B. A. y. (2017). Communication-efficient learning of deep networks from decentralized data. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, 54, 1273–1282. (Proceedings of Machine Learning Research)
- [9]. Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., Bakas, S., Galtier, M. N., Landman, B. A., Maier-Hein, K., Ourselin, S., Sheller, M., Summers, R. M., Trask, A., Xu, D., Baust, M., & Cardoso, M. J. (2020). The future of digital health with federated learning. *npj Digital Medicine*, 3, 119. <https://doi.org/10.1038/s41746-020-00323-1> (DOI)
- [10]. Sheller, M. J., Edwards, B., Reina, G. A., Martin, J., Pati, S., Kotrotsou, A., Milchenko, M., Xu, W., Marcus, D., Colen, R. R., & Bakas, S. (2020). Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data. *Scientific Reports*, 10, 12598. <https://doi.org/10.1038/s41598-020-69250-1>