International Journal of Creative and Open Research in Engineering and Management
ISSN: 3108-1754 (Online)
Volume 02 Issue 09 September-2026 | Impact Factor: 3.5

An Adaptive Personalized Federated Learning Framework for Secure and Robust Pregnancy Risk Prediction

Pranita Sanjay Waikar[1], Dr. Pallavi Jha[2]

[1] Research Scholar and Assistant Professor ,Department of Computer Engineering ,

Modern Education Society's Wadia College of Engineering, Pune

[2] Research Guide & Professor ,School of Engineering and Technology ,ALARD University, Pune

How to Cite this Article:

Waikar, P. S. (2026). An Adaptive Personalized Federated Learning Framework for Secure and Robust Pregnancy Risk Prediction. International Journal of Creative and Open Research in Engineering and Management, 02(9), 1-9. https://doi.org/10.55041/ijcope.v2i9.008

License:

This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

© The Author(s). Published by International Journal of Creative and Open Research in Engineering and Management.

https://doi.org/10.55041/ijcope.v2i9.008

Abstract

Pregnancy-related complications continue to be an important concern in maternal healthcare, particularly in situations where early identification of high-risk pregnancies can support timely medical intervention. Machine learning has shown promising results in predicting maternal health risks from physiological and clinical parameters. However, most conventional machine-learning approaches depend on centralized healthcare datasets, requiring sensitive patient information to be transferred to a common location. Such centralized data management can introduce concerns related to privacy, security, data ownership, regulatory compliance, and institutional data-sharing policies.Federated Learning (FL) offers an alternative by allowing healthcare institutions to collaboratively train a machine-learning model while keeping patient records within their respective institutions. However, healthcare data rarely follow identical distributions across hospitals. Differences in patient demographics, geographical conditions, disease prevalence, clinical practices, and available healthcare facilities can lead to highly non-independent and identically distributed (non-IID) data. Under such conditions, a single global model may not provide equally effective predictions for all participating institutions.To address this limitation, this study proposes an **Adaptive Personalized Federated Learning (APFL)** framework for privacy-preserving pregnancy risk prediction. The proposed framework combines personalized local models, adaptive client aggregation, differential privacy, and secure aggregation. Participating healthcare institutions act as decentralized clients and train the prediction model using locally available pregnancy-related data. Instead of sharing patient records, protected model updates are transmitted to a coordinating server. The server adaptively combines these updates, while personalized model components allow individual institutions to retain knowledge specific to their local patient population.The proposed framework will be evaluated using accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC). Additional evaluation will consider communication overhead and privacy-related performance. The study aims to demonstrate that adaptive personalization can improve the robustness of federated pregnancy-risk prediction when healthcare data are heterogeneous, while simultaneously reducing the exposure of sensitive patient information. The proposed framework can provide a foundation for developing scalable and trustworthy artificial intelligence solutions for decentralized maternal healthcare.

Keywords: Federated Learning, Pregnancy Risk Prediction, Maternal Healthcare, Non-IID Data, Privacy-Preserving Machine Learning

1. Introduction

Maternal healthcare remains a major public-health concern because pregnancy-related complications can adversely affect both mothers and newborns. Early identification of high-risk pregnancies can support timely monitoring, referral, counselling, and clinical intervention. In recent years, machine learning (ML) has shown promising potential in maternal-risk prediction by identifying patterns in physiological and clinical parameters such as age, blood pressure, blood glucose, body temperature, and heart rate (Ahmed et al., 2020; Khadidos et al., 2024). Various approaches, including decision trees, ensemble models, and neural networks, have been explored for classifying maternal-risk levels (Al Mashrafi et al., 2024).Despite these advances, most ML systems rely on centralized healthcare data. Collecting patient records from multiple institutions in a central repository raises concerns regarding privacy, security, data ownership, and regulatory compliance. Federated Learning (FL) provides an alternative by enabling institutions to train models locally while sharing model updates rather than raw patient data (Kanauzia et al., 2026).However, healthcare data are typically non-IID because hospitals differ in patient demographics, disease prevalence, clinical practices, and healthcare resources. Such heterogeneity can reduce the effectiveness of conventional FedAvg and lead to client drift and unequal model performance.

To address these challenges, this study proposes an **Adaptive Personalized Federated Learning (APFL)** framework that combines adaptive aggregation, personalized learning, Differential Privacy, and Secure Aggregation for privacy-preserving pregnancy-risk prediction under heterogeneous healthcare data.

2. Background of the Study

2.1 Pregnancy Risk Prediction Using Machine Learning

Machine learning has increasingly been investigated for supporting maternal-risk assessment. The underlying idea is to learn relationships between physiological measurements and clinically defined risk categories from historical data. Once trained, a classification model can estimate the risk category associated with a new patient record.The **UCI Maternal Health Risk dataset** is one of the publicly accessible datasets commonly used in maternal-health prediction research. The dataset contains **1,013 instances and six predictive variables**, namely age, systolic blood pressure, diastolic blood pressure, blood sugar, body temperature, and heart rate. The target attribute, **RiskLevel**, represents three maternal-risk categories: Low, Mid, and High.Studies using this dataset have demonstrated that machine-learning methods can identify meaningful relationships between physiological characteristics and maternal-risk levels. Ensemble algorithms, decision-tree approaches, and hybrid learning architectures have all been explored for this task.

Khadidos et al. (2024), for example, investigated ensemble learning techniques for maternal-health risk prediction and reported encouraging results. Other comparative studies have also identified Random Forest as a competitive method for maternal-risk classification. Hybrid approaches combining artificial neural networks and Random Forest have further demonstrated the potential of combining different learning strategies.Although these studies establish the usefulness of machine learning for maternal-risk prediction, they generally assume that the complete dataset can be accessed centrally. This assumption may not hold in real-world healthcare environments, where institutions may be unwilling or unable to share individual-level patient records.

2.2 Federated Learning in Healthcare

Federated Learning is a distributed learning paradigm in which multiple clients jointly train a machine-learning model without directly transferring their raw datasets to a central server.

In a conventional centralized architecture, the learning process can be represented as:

Patient Data → Central Repository → Model Training

In contrast, a federated architecture follows the pattern:

Hospital A → Local Training → Model Update
Hospital B → Local Training → Model Update
Hospital C → Local Training → Model Update
Hospital D → Local Training → Model Update

The server receives model updates from participating clients and combines them to produce an improved global model. The updated model is then returned to the clients for another round of local training.

A basic FedAvg aggregation process can be expressed as:

wt+1=∑k=1KnkNwktw^{t+1}=\sum_{k=1}^{K}\frac{n_k}{N}w_k^t

where wktw_k^t denotes the model parameters obtained from client kk at training round tt, nkn_k represents the number of training samples available at that client, and NN denotes the total number of training samples across participating clients.The distributed nature of FL makes it particularly attractive for healthcare applications. Medical institutions can potentially collaborate on machine-learning development without creating a centralized repository containing all patient records. Nevertheless, practical healthcare FL systems must address additional challenges such as communication cost, client availability, model heterogeneity, privacy attacks, interoperability, and governance (Kanauzia et al., 2026).

2.3 Non-IID Data in Healthcare

A fundamental challenge in federated learning is the presence of non-IID data. In a conventional machine-learning setting, training data are often assumed to originate from a similar underlying distribution. In federated healthcare environments, this assumption is frequently unrealistic.Consider four healthcare institutions participating in a pregnancy-risk prediction system. Hospital A may have a larger proportion of low-risk cases. Hospital B may treat more high-risk pregnancies. Hospital C may serve an older population, while Hospital D may primarily serve patients from rural communities.

Therefore, the probability distributions associated with two clients may differ:

Pk(X,Y)≠Pj(X,Y)P_k(X,Y)\neq P_j(X,Y)

where Pk(X,Y)P_k(X,Y) and Pj(X,Y)P_j(X,Y) represent the data distributions of clients kk and jj, respectively.

Such differences can produce several problems during federated training:

- **Client drift:** local models move toward different optimization directions.
- **Slower convergence:** the global model may require more communication rounds.
- **Training instability:** aggregated updates can conflict with one another.
- **Reduced global performance:** the global model may not represent all clients effectively.
- **Poor minority-client performance:** smaller or unusual client populations may receive less benefit.
- **Unequal prediction quality:** model performance may vary considerably between healthcare institutions.

These challenges make non-IID learning particularly important when designing federated systems for maternal healthcare.

2.4 Personalized Federated Learning

Conventional FL attempts to develop a common global model that can be shared among all participating clients. Although this strategy is effective when client distributions are relatively similar, it may be less suitable when institutions have substantially different patient populations.Personalized Federated Learning addresses this limitation by allowing a client to retain a portion of its model specifically adapted to local data.

The proposed framework represents the personalized model as:

Mk=S+PkM_k=S+P_k

where:

- SS represents the **shared model component**,
- PkP_k represents the **personalized component** for client kk, and
- MkM_k represents the complete model used by client kk.

The shared component is expected to capture common maternal-risk patterns that are useful across institutions. The personalized component, in contrast, captures local characteristics that may differ from one institution to another.

For example, a rural healthcare institution may develop personalized parameters reflecting the characteristics of its local patient population, while an urban hospital may learn a different personalized representation. This allows collaboration without requiring all institutions to follow an identical model.

2.5 Privacy Preservation

Federated Learning reduces the need to transfer raw healthcare records, but it should not be considered a complete privacy solution by itself. Model updates can potentially reveal information about the data used during local training under certain adversarial conditions.For this reason, additional privacy-preserving mechanisms can be incorporated into the federated architecture. Differential Privacy is one such technique. It introduces controlled randomness into model updates so that the contribution of an individual record becomes more difficult to infer.

A simplified representation of a noisy gradient is:

g~=g+N(0,σ2)\tilde{g}=g+N(0,\sigma^2)

where gg represents the original gradient and N(0,σ2)N(0,\sigma^2) represents Gaussian noise with variance σ2\sigma^2. Another important mechanism is **Secure Aggregation**. Instead of allowing the coordinating server to inspect each client's update separately, secure aggregation enables the server to obtain only the combined contribution of participating clients. This provides an additional layer of protection against exposure of individual client updates.Recent research has examined combinations of differential privacy, secure multiparty computation, homomorphic encryption, trusted execution environments, and other privacy-enhancing techniques for healthcare federated learning (Alankamony & Nels, 2026).

3. Literature Review

The existing literature relevant to this study can be broadly organized into three interconnected areas: **machine learning for maternal-risk prediction, federated learning for healthcare, and privacy-preserving personalized federated learning**.Research in maternal-health prediction has demonstrated that physiological characteristics can provide useful information for classifying pregnancy-related risk. The UCI Maternal Health Risk dataset has

consequently been used in several machine-learning studies because it contains clinically meaningful variables together with three maternal-risk categories.

Ahmed et al. (2020) demonstrated the applicability of machine-learning approaches to maternal-health risk prediction. Subsequent studies have investigated ensemble and decision-tree-based approaches for improving predictive performance. Khadidos et al. (2024) examined ensemble learning techniques and demonstrated the potential of computational models for maternal-risk classification. Other studies have reported promising results using Random Forest and hybrid neural-network approaches.Despite the predictive capability demonstrated by these studies, most of the existing approaches depend on centralized datasets. Consequently, they do not directly address the challenge of enabling several healthcare institutions to collaboratively develop a model while retaining patient data locally.

Federated Learning provides an alternative learning architecture. Rather than transferring patient records to a central repository, each institution performs local training and communicates model information to a coordinating server. Recent healthcare FL research has identified applications in diagnosis, clinical prediction, monitoring, and personalized healthcare while also highlighting challenges associated with privacy, communication, interoperability, governance, and heterogeneous data (Kanauzia et al., 2026).However, healthcare FL is particularly affected by non-IID data. Different institutions may have substantially different patient populations, clinical environments, and data volumes. Consequently, a conventional global model may not perform equally well for all clients.

Personalized Federated Learning has therefore received increasing attention as a strategy for addressing client-specific variations. By separating shared knowledge from locally adapted knowledge, personalized approaches can potentially improve client-level performance without abandoning the collaborative benefits of FL.Privacy protection is another important area of research. Recent work emphasizes that FL should be combined with additional mechanisms when sensitive healthcare information is involved. Differential Privacy, Secure Aggregation, encryption-based approaches, and secure multiparty computation have been investigated as complementary privacy mechanisms (Alankamony & Nels, 2026).

The literature therefore indicates a clear opportunity to integrate **maternal-risk prediction, non-IID federated learning, personalization, adaptive aggregation, and privacy protection** within a single framework. The proposed study is designed to address this intersection.

4. Research Gap

Existing maternal-risk prediction studies largely rely on centralized datasets, limiting privacy-preserving collaboration across healthcare institutions.Non-IID healthcare data can negatively affect conventional FedAvg performance and cause client drift.A single global model may not adequately capture institution-specific maternal-health patterns.Federated Learning alone may not provide sufficient protection against privacy attacks on model updates.

5. Research Objectives

1. To develop a privacy-preserving federated learning framework for pregnancy-risk prediction.
2. To investigate the impact of non-IID data distributions on maternal-risk prediction across heterogeneous healthcare clients.
3. To design an adaptive aggregation mechanism that can assign appropriate importance to participating healthcare clients.
4. To develop personalized model components capable of capturing institution-specific maternal-health patterns.
5. To incorporate Differential Privacy into the federated training process to reduce the risk of information leakage.
6. To investigate Secure Aggregation as an additional mechanism for protecting individual client updates.

7. To compare the proposed APFL framework with centralized machine learning, FedAvg, FedProx, and existing personalized federated-learning approaches.
8. To evaluate the proposed framework using prediction performance, communication overhead, and privacy-related measures.

6. Methodology

The proposed research follows an experimental methodology designed to examine pregnancy-risk prediction under decentralized and heterogeneous healthcare data conditions. The overall process consists of **dataset acquisition, data preprocessing, non-IID client generation, local model training, privacy protection, adaptive aggregation, personalized learning, and performance evaluation**.

6.1 Dataset

The initial experimental evaluation will use the **Maternal Health Risk dataset available through the UCI Machine Learning Repository**.

The dataset contains **1,013 records and six predictive attributes**:

- Age
- Systolic Blood Pressure
- Diastolic Blood Pressure
- Blood Sugar
- Body Temperature
- Heart Rate

The target variable is **RiskLevel**, which consists of three classes:

- Low Risk
- Mid Risk
- High Risk

The dataset provides a suitable starting point for evaluating the proposed framework because it contains physiological attributes directly related to maternal-risk classification.Since the publicly available dataset is not naturally divided among hospitals, virtual healthcare clients will be constructed during experimentation to simulate decentralized healthcare institutions.

6.2 Data Preprocessing

Before federated training, the dataset will undergo preprocessing to improve data quality and model stability.

The preprocessing stage will include:

1. Identification of missing values.
2. Detection of duplicate observations.
3. Examination of inconsistent records.
4. Identification of extreme observations.
5. Feature scaling.
6. Encoding of target classes.
7. Division into training and testing datasets.

Numerical features will be standardized using:

$X' = X - \mu \sigma$X'=\frac{X-\mu}{\sigma}

where ,XX represents the original feature value, μ\mu represents the mean of the feature, and σ\sigma represents its standard deviation.

The target variable will be encoded as:

$Low = 0, Mid = 1, High = 2$Low=0,\quad Mid=1,\quad High=2

The test dataset will remain isolated from federated training so that final model performance can be evaluated without information leakage.

6.3 Non-IID Client Construction

Because the initial dataset is publicly available rather than naturally distributed across hospitals, the research will simulate multiple healthcare clients.

For example, the dataset may be divided among several virtual hospitals with intentionally different distributions of:

- maternal-risk classes,
- age groups,
- blood-pressure ranges,
- blood-sugar levels, and
- other physiological characteristics.

Three levels of heterogeneity will be investigated:

Scenario 1: Low Heterogeneity

Client datasets will have relatively similar distributions.

Scenario 2: Moderate Heterogeneity

Clients will contain noticeable differences in risk categories and patient characteristics.

Scenario 3: High Heterogeneity

Clients will have strongly different distributions, representing challenging real-world healthcare conditions.This experimental design will allow the robustness of the proposed APFL framework to be examined under progressively more difficult non-IID conditions.

6.4 Local Prediction Model

Each healthcare client will train a lightweight neural-network classifier using its locally available data.

The proposed local model will consist of:

Input Layer → Dense Layer → ReLU → Dropout → Dense Layer → Softmax Output

The input layer will receive the six maternal-health features. Hidden dense layers will learn relationships between physiological variables, while ReLU activation will introduce non-linearity. Dropout will be used to reduce overfitting.The final Softmax layer will generate probabilities for the three maternal-risk classes:

P(Low|x),P(Mid|x),P(High|x)P(Low|x),\quad P(Mid|x),\quad P(High|x)

The predicted class will be determined using:

y^=arg⁡max⁡yP(y|x)\hat{y}=\arg\max_y P(y|x)

where P(y|x)P(y|x) represents the predicted probability of class yy for input xx.

6.5 Federated Training

The federated training process will operate over multiple communication rounds.

During each round:

1. The server distributes the current shared model to participating clients.
2. Each client trains the model using its local dataset.
3. The client calculates the updated model parameters.
4. Differential Privacy is applied to protect the transmitted update.
5. Protected updates are securely aggregated.
6. The server applies the adaptive aggregation mechanism.
7. The updated shared parameters are distributed to clients.
8. Each client updates its personalized component.

The process continues until the predefined number of communication rounds is completed or the model reaches the required convergence criterion.

6.6 Adaptive Aggregation

Standard FedAvg assigns client contributions primarily according to the number of local training samples. However, dataset size alone may not be sufficient in highly heterogeneous healthcare environments.The proposed APFL framework will therefore investigate an adaptive aggregation strategy. Client contributions can be influenced by factors such as:

- local dataset size,
- local model performance,
- validation loss,
- data heterogeneity,
- update quality, and
- training stability.

A generalized adaptive aggregation formulation can be represented as:

wt+1=∑k=1Kαktwktw^{t+1}=\sum_{k=1}^{K}\alpha_k^t w_k^t

subject to:

∑k=1Kαkt=1\sum_{k=1}^{K}\alpha_k^t=1

where αkt\alpha_k^t represents the adaptive aggregation weight assigned to client kk during round tt.The purpose of this mechanism is to prevent the global model from being dominated solely by clients with larger datasets and to encourage more balanced learning across heterogeneous institutions.

6.7 Personalized Model Learning

The proposed model will separate common and client-specific knowledge using:

Mk=S+PkM_k=S+P_k

Here, SS represents shared parameters learned collaboratively across clients, whereas PkP_k represents the personalized parameters maintained by client kk.The shared component is intended to capture general maternal-risk patterns that are common across institutions. The personalized component will adapt the model to characteristics specific to each healthcare institution.This separation is expected to be particularly useful under strong non-IID conditions because clients can benefit from collaborative knowledge without losing the ability to model their own local characteristics.

6.8 Differential Privacy

Differential Privacy will be incorporated into the communication process to reduce the possibility of reconstructing sensitive information from model updates.Before a client sends an update to the coordinating server, the update will be subjected to a privacy-preserving mechanism. A simplified representation is:

g~=g+N(0,σ2)\tilde{g}=g+N(0,\sigma^2)

where gg is the original gradient or update, g~\tilde{g} is the protected update, and σ\sigma controls the magnitude of the added Gaussian noise.Different privacy settings will be examined to study the trade-off between prediction performance and privacy protection.

6.9 Secure Aggregation

Secure Aggregation will provide an additional protection layer during federated communication. Instead of exposing each individual client's model update to the server, the protocol will allow the server to obtain the aggregated update.The objective is to ensure that the coordinating server cannot directly inspect the contribution of a specific healthcare institution.The combination of **Federated Learning + Differential Privacy + Secure Aggregation** is expected to provide stronger protection than relying on federated learning alone.

6.10 Evaluation Metrics

The proposed APFL framework will be evaluated using both predictive and system-level measures.

Classification Metrics

The following metrics will be considered:

- Accuracy
- Precision
- Recall
- F1-score
- AUC-ROC

For multi-class classification, macro-averaged and weighted metrics will be considered where appropriate.

Federated Learning Metrics

The experimental evaluation will additionally consider:

- Number of communication rounds
- Communication overhead
- Convergence behaviour
- Global model performance
- Client-level performance
- Performance variation across clients

Privacy Evaluation

The study will investigate the effect of different privacy settings on model performance and will analyse the trade-off between privacy protection and predictive accuracy.

6.11 Comparative Evaluation

To establish the effectiveness of the proposed framework, APFL will be compared against several baseline approaches:

Model	Purpose
Centralized ML	Establishes conventional centralized prediction performance
FedAvg	Baseline federated-learning approach
FedProx	Evaluates FL under heterogeneous client data
Personalized FL	Evaluates the benefit of personalization
APFL	Evaluates the proposed adaptive personalized approach

The comparison will be performed under different levels of non-IID data heterogeneity.The experimental results will help determine whether adaptive aggregation and personalization provide measurable improvements over conventional federated learning.

6.12 Proposed Experimental Workflow

The complete research workflow can be summarized as:

Maternal Health Risk Dataset

↓

Data Cleaning and Preprocessing

↓

Non-IID Data Partitioning

↓

Virtual Healthcare Clients

↓

Local Model Training

↓

Differential Privacy

↓

Secure Aggregation

↓

Adaptive Client Aggregation

↓

Shared Model Update

↓

Personalized Client Models

↓

Prediction

↓

Performance and Privacy Evaluation

The proposed architecture is intended to create a balance between three important requirements: **predictive effectiveness, personalization under heterogeneous data, and protection of sensitive healthcare information**.

7. Results and Expected Outcomes

The proposed experimental framework has not yet been implemented and evaluated; therefore, numerical performance results are not reported in the present study. Instead, this section defines the evaluation strategy and the outcomes that will be investigated after implementation.The proposed APFL framework will be compared with five learning approaches: **Centralized Machine Learning, FedAvg, FedProx, Personalized Federated Learning, and APFL**. The comparison will be conducted using Accuracy, Precision, Recall, F1-score, and AUC-ROC. In addition to overall predictive performance, client-level performance will be examined to understand how effectively each method handles differences among healthcare institutions.A particular focus will be placed on **recall for the High-risk class**. In maternal healthcare, correctly identifying women who may require additional clinical attention is especially important.

Therefore, a model with high overall accuracy but poor detection of high-risk cases may not be clinically desirable. The proposed evaluation will consequently examine both overall classification performance and class-specific results.The experiments will be conducted under different levels of non-IID data distribution, including low, moderate, and high heterogeneity. This will help determine whether personalization and adaptive aggregation can reduce the negative effects of client drift. Convergence behaviour across communication rounds will also be monitored to assess training stability and efficiency.

The effect of Differential Privacy will be studied by varying the privacy level and observing the resulting change in predictive performance. This analysis is expected to demonstrate the **privacy–utility trade-off**, in which stronger privacy protection may introduce additional noise and consequently affect model performance.Based on the design of the proposed framework, APFL is expected to perform more consistently than conventional FedAvg when client data are highly heterogeneous. FedProx and Personalized FL are also expected to provide improvements over standard FedAvg by reducing the effect of local model divergence. The proposed adaptive personalization strategy is expected to further improve client-level consistency by considering both shared knowledge and institution-specific characteristics.These expectations will be verified experimentally rather than assumed as confirmed findings. The final results will determine whether APFL provides a measurable improvement in the balance among **prediction accuracy, high-risk detection, personalization, convergence, and privacy protection**.

8. Discussion

The proposed APFL framework addresses key challenges in maternal-risk prediction, including data heterogeneity, privacy, and decentralized learning.
Federated Learning enables healthcare institutions to collaborate without directly sharing sensitive patient records.
Adaptive aggregation is designed to reduce client drift and provide better performance under non-IID healthcare data.
The personalized model ($M_k=S+P_k$) combines shared knowledge with institution-specific patterns.
Differential Privacy and Secure Aggregation provide additional protection for model updates while requiring a balance between privacy and predictive utility.
The evaluation will focus on Accuracy, Recall, F1-score, AUC, and particularly High-risk Recall to assess the practical effectiveness of the proposed framework.

9. Limitations

 The study relies on a publicly available maternal-health dataset with simulated non-IID clients; therefore, it may not fully represent the complexity, diversity, and operational challenges of real multi-hospital healthcare environments.
 The dataset contains limited clinical attributes, and real-world deployment would require larger and more diverse datasets, additional clinical variables, healthcare-professional validation, fairness assessment, security evaluation, and regulatory approval.

10. Conclusion

This study proposes an **Adaptive Personalized Federated Learning (APFL) framework for privacy-preserving pregnancy-risk prediction under non-IID healthcare data**. The framework is motivated by the limitations of conventional centralized machine learning and standard federated learning when healthcare information is distributed across institutions and exhibits substantial heterogeneity.The proposed approach combines four key components: **adaptive aggregation, personalized model learning, Differential Privacy, and Secure Aggregation**. Adaptive aggregation is intended to provide a more suitable contribution mechanism for heterogeneous clients, while personalization allows individual healthcare institutions to retain knowledge specific to their local patient populations. Differential Privacy and Secure Aggregation provide additional protection for model updates exchanged during collaborative training.

The experimental methodology will compare APFL with Centralized Machine Learning, FedAvg, FedProx, and Personalized Federated Learning. Evaluation will consider Accuracy, Precision, Recall, F1-score, AUC-ROC, convergence behaviour, communication overhead, and privacy–utility trade-offs. Particular attention will be given to High-risk Recall because reliable identification of high-risk pregnancies is especially important in maternal-health applications.

Since the experimental implementation has not yet been completed, the study does not claim numerical improvements at this stage. Instead, it establishes a research framework through which the effectiveness of adaptive personalization under different levels of non-IID data can be systematically investigated.

The proposed work provides a foundation for future multi-institutional maternal-health AI systems in which collaborative learning can take place without requiring direct centralization of patient records. Future research should extend the evaluation to larger and clinically diverse datasets, real hospital environments, stronger privacy and security assessments, fairness analysis, and prospective clinical validation. Such extensions will be essential before the proposed framework can be considered for practical healthcare deployment.

References

1.Ahmed, M., Kashem, M. A., Rahman, M., & Khatun, S. (2020). *Maternal health risk* [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5DP5D

2.Coelho, K. K., Nogueira, M., Vieira, A. B., Silva, E. F., & Nacif, J. A. M. (2023). A survey on federated learning for security and privacy in healthcare applications. *Computer Communications, 207*, 113–127. https://doi.org/10.1016/j.comcom.2023.05.012

3.Khadidos, A. O., Saleem, F., Selvarajan, S., Ullah, Z., et al. (2024). Ensemble machine learning framework for predicting maternal health risk during pregnancy. *Scientific Reports, 14*, 21483. https://doi.org/10.1038/s41598-024-71934-x

4.Kanauzia, R., Singh, M., Gulati, S., Singh, B. M., Arora, N., & Gola, K. K. (2026). A comprehensive survey on federated learning for privacy preservation in digital healthcare applications. *Knowledge and Information Systems, 68*, 55. https://doi.org/10.1007/s10115-025-02673-2

5.Lee, et al. (2024). PPFL: A personalized progressive federated learning method for leveraging different healthcare institution-specific features. *Scientific Reports*.

6.Hossain, M. M., Nayan, N. M., & Kashem, M. A. (2026). A comprehensive maternal health risk prediction dataset from IoT-enabled medical cyber-physical systems in developing countries: Supporting machine learning and deep learning applications for clinical decision support. *BMC Medical Informatics and Decision Making, 26*, 79. https://doi.org/10.1186/s12911-026-03343-1

7.Prediction of maternal health risk factors using machine learning algorithms. (2025). *Procedia Computer Science, 258*, 2713–2722. https://doi.org/10.1016/j.procs.2025.04.532

8.Enhancing maternal health risk prediction: Addressing class imbalance with oversampling and machine learning. (2026). *Procedia Computer Science, 283*, 5536–5546. https://doi.org/10.1016/j.procs.2026.06.603

9.A review of privacy enhancement methods for federated learning in healthcare systems. (2023). *Healthcare, 11*.

10.Privacy preservation for federated learning in health care. (2024). *Patterns, 5*(7), 100974. https://doi.org/10.1016/j.patter.2024.100974

11.A systematic review of privacy-preserving federated learning in decentralized healthcare systems. (2025). *Franklin Open, 13*, 100440. https://doi.org/10.1016/j.fraope.2025.100440

12.Federated learning in healthcare: Recent progress and challenges. (2026). *Computers and Electrical Engineering, 131*, 110924. https://doi.org/10.1016/j.compeleceng.2025.110924

13.Federated learning in healthcare: From research to real-world deployment. (2026). *Annual Review of Biomedical Engineering*.

14.Demography-aware personalized federated learning for fair, private, and efficient clinical risk prediction. (2026). *Biomedical Signal Processing and Control, 114*, 109312.