Acta Informatica Pragensia
2026, Volume 15, Issue 2, 327–344
https://doi.org/10.18267/j.aip.304

PRAGUE UNIVERSITY OF ECONOMICS AND BUSINESS

Article Open Access

Enhanced Diabetes Detection via a Privacy-Preserving Federated Learning Framework

Nouhaila Aasoum [1], **Ismail Jellouli** [1], **Souad Amjad** [2]

[1] Emerging Computer Technologies Research Unit, Faculty of Science, Abdelmalek Essaadi University, Tetouan, Morocco
[2] Information Technology and Systems Modeling Laboratory, Faculty of Science, Abdelmalek Essaadi University, Tetouan, Morocco

Corresponding author: Nouhaila Aasoum (nouhaila.aasoum@etu.uae.ac.ma)

Editorial Record

First submission received:
October 25, 2025

Revision received:
December 24, 2025

Accepted for publication:
January 26, 2026

Academic Editor:
Jan Bruthans
Czech Technical University in Prague, Czech Republic

This article was accepted for publication by the Academic Editor upon evaluation of the reviewers' comments.

How to cite this article:
Aasoum, N., Jellouli, I., & Amjad, S. (2026) Enhanced Diabetes Detection via a Privacy-Preserving Federated Learning Framework. *Acta Informatica Pragensia*, 15(2), 327–344. https://doi.org/10.18267/j.aip.304

Abstract

Background: The integration of artificial intelligence (AI) in healthcare depends on striking a balance between patient privacy and clinical utility. The standard methods often compromise one for the other, preventing the development of trustworthy healthcare AI.

Objective: This paper aims to resolve the privacy-utility trade-off by developing an enhanced federated learning framework with adaptive differential privacy (DP) mechanisms that are optimized for clinical data.

Methods: We implement and compare several different methods, from the most centralized deep learning to various federated configurations with formal DP guarantees. Our improved framework involves adaptive noise scheduling and quality-weighted federated averaging on top of a federated neural network framework. We validate on two major diabetes screening datasets: Diabetes Health Indicators (BRFSS 2015) and National Health and Nutrition Examination Survey (NHANES 2015–2016), including comprehensive clinical measurements.

Results: This paper presents a favourable balance between privacy protection and clinical utility for both datasets. It offers strong formal differential privacy guarantees and good diagnostic performance, achieving high ranking accuracy with clinical risk prioritization. The model demonstrates generalization robustness by capturing clinically meaningful risk factors aligned with established medical guidelines, confirming that the applied privacy-preserving mechanisms do not compromise clinical relevance.

Conclusion: Our framework meaningfully advances the privacy-utility trade-off healthcare AI, by offering tunable formal privacy guarantees while ensuring strong clinical performance. The approach is highly generalizable across diverse data collection methodologies and maintains clinically relevant feature representations, thus allowing safe adoption in sensitive medical domains.

Index Terms

Differential privacy; Federated learning; Clinical machine learning; Privacy; Healthcare analytics.

1 INTRODUCTION

The increased use of machine learning (ML) in healthcare has opened up new possibilities for predicting diagnoses and personalizing treatment. However, it also raises significant privacy concerns about sensitive patient data (Esteva et al., 2019). Clinical datasets, which include electronic health records (EHR), diagnostic measurements and behavioural health indicators, contain private information that is protected under strict regulations such as the Health Insurance Portability and Accountability Act (HIPAA) and the European General Data Protection Regulation (GDPR) (Price & Cohen, 2019).

One effective way to address the risks of data centralization is federated learning (FL). This method allows collaborative model training by sharing model updates instead of raw data, keeping patient records decentralized (Jimenez Gutierrez et al., 2024). Nonetheless, FL by itself does not provide a full solution. Studies indicate that model gradients may be reverse-engineered via inference attacks, potentially revealing sensitive training information. This risk emphasizes the necessity for a robust privacy component within the FL framework.

Differential privacy (DP) has emerged as the benchmark for delivering formal privacy guarantees (Aasoum et al., 2024). Yet, its traditional uniform application adds the same noise to all features, which can unintentionally diminish the usefulness of important clinical variables (e.g., ejection fraction in heart failure or glucose levels in diabetes) and threaten diagnostic accuracy (Abadi et al., 2016). This "one-size-fits-all" approach overlooks a vital fact: not every feature equally affects clinical decisions. For instance, model reliability is directly compromised when noise obscures glucose measurements as a key diabetes indicator according to American Diabetes Association (ADA) standards at the same level as administrative codes. The challenge, then, is to create a system that takes advantage of the decentralized nature of FL while incorporating a DP mechanism that considers the clinical context, preventing data leakage without hindering diagnostic utility.

In this paper, we present an optimized federated learning framework with an adaptive DP mechanism. This mechanism dynamically distributes privacy budgets (ε) based on the importance of evidence-based clinical features, addressing this gap. We introduce proportionately calibrated noise to maintain high-utility predictors, use medical domain knowledge from established clinical guidelines (McDonagh et al., 2021) to weight features and carefully validate our outputs against findings in medical literature. Our combined FL-DP approach strikes a balance between privacy and utility, enabling a safer implementation of machine learning in healthcare without sacrificing diagnostic accuracy and while respecting data sovereignty. This development is crucial as ethical responsibility and data-driven medicine become more closely linked (Vayena et al., 2018).

Our work contributes to the field in four interrelated ways:

- **Federated learning with enhanced differential privacy:** We develop a federated learning framework incorporating adaptive differential privacy mechanisms that provide strong formal privacy guarantees while maintaining clinical utility across decentralized data sources.
- **Clinical context-aware optimization:** Our method strategically optimizes the privacy-utility trade-off by aligning noise allocation with feature importance, preserving diagnostic accuracy for clinically critical markers while maintaining robust privacy protection.
- **Relationship-preserving privacy mechanism:** We implement coordinated noise injection strategies that preserve important feature interactions crucial for clinical decision-making, ensuring that privacy protection does not disrupt clinically meaningful patterns in the learned models.
- **Formal privacy guarantees with tunable protection:** Our framework provides mathematically rigorous differential privacy guarantees with scalable privacy budgets (ε), offering healthcare institutions quantifiable and adjustable privacy protection levels suitable for diverse clinical applications.

The paper is organized as follows: Section 2 reviews related research into differential privacy, federated learning and existing hybrid frameworks in healthcare analytics. Section 3 describes our proposed methodology, federated learning architecture, privacy accounting and formal guarantees and noise injection strategy. Section 4 presents experimental validation of our integrated framework across various healthcare domains using leading benchmarks. Section 5 discusses key findings, clinical implications, limitations and future directions. Table 1 presents a list of the acronyms mentioned in the paper.

Table 1. *List of acronyms used in this paper.*

Acronym	Definition
AI	Artificial intelligence
FL	Federated learning
DP	Differential privacy
ML	Machine learning
HIPAA	Health Insurance Portability and Accountability Act

Acronym	Definition
GDPR	General Data Protection Regulation
ADA	American Diabetes Association
DP-SGD	Differentially private stochastic gradient descent
MRI	Magnetic resonance imaging
Non-IID	Non-Independent and identically distributed
EHR	Electronic health records
ReLU	Rectified linear unit
BMI	Body mass index
HbA1c	Glycated haemoglobin
RDP	Rényi differential privacy
AUC-ROC	Area under the receiver operating characteristic curve
AUC-PR	Area under the precision-recall curve
TP	True positive
TN	True negative
FP	False positive
FN	False negative
CDC	Centres for Disease Control and Prevention
NDCG	Normalized discounted cumulative gain
SMOTE	Synthetic minority over-sampling technique
AUC	Area under curve
NHANES	National Health and Nutrition Examination Survey
BRFSS 2015	Behavioural Risk Factor Surveillance System

2 RELATED WORK

2.1 Differential privacy in healthcare analytics

Differential privacy (DP) has become the benchmark for safeguarding privacy in medical data analysis since its formal introduction by Dwork & Roth (2014). Initial applications concentrated on aggregate statistics and basic queries (Chawla et al., 2005), whereas contemporary implementations such as differentially private stochastic gradient descent (DP-SGD) facilitate intricate machine learning tasks by incorporating calibrated noise into gradients during neural network training (Ziller et al., 2021). Nonetheless, these methods frequently result in substantial drops in utility for clinical prediction tasks (Suriyakumar et al., 2021), particularly affecting sensitivity metrics that are vital for disease screening. Methods such as DP Gaussian naive Bayes and DP random forests offer formal privacy assurances but consider all attributes uniformly, which is a significant drawback in healthcare, where essential biomarkers such as glucose levels need more robust protection compared to demographic factors (Vaidya et al., 2013). This uniform approach does not consider the differing clinical importance of various health indicators, prompting our clinically-driven privacy allocation method. Recent studies show that distributing privacy without considering features can decrease model accuracy on critical clinical predictions by up to 30%, highlighting the need for advanced privacy techniques in medical settings (Bagdasaryan et al., 2019).

For instance, Williamson & Prybutok (2024) examined privacy challenges in AI-driven healthcare from a systemic perspective, including patient perceptions, regulatory oversight and ethical concerns. The authors argued that governance frameworks to ensure trust and adoption must complement technical solutions such as DP. They also highlighted the tension between innovation and patient rights, proposing balanced strategies for oversight. This supports our emphasis on clinically informed privacy mechanisms that respect both utility and patient trust.

Additionally, Amanullah & Solms (2025) focused on the trade-off between privacy and performance in deep learning and federated learning under differential privacy. The authors synthesized recent advances, highlighting how DP can degrade performance if not carefully tuned and proposed strategies to mitigate this. They stressed the need for

context-aware privacy allocation in sensitive domains such as healthcare. This supports our argument that uniform DP is insufficient and motivates our adaptive mechanism.

The challenge lies in developing adaptive methods that preserve diagnostic accuracy while providing robust privacy safeguards across multiple healthcare data.

2.2 Federated learning for healthcare applications

Federated learning (FL) has emerged as an essential framework for developing machine learning models in healthcare, successfully addressing major concerns related to data privacy and segregated medical data. It enables different organizations, such as hospitals or clinics, to collaboratively train a model while safeguarding their confidential patient data by solely exchanging model updates (such as gradients or weights). This approach minimizes privacy and security threats while promoting the creation of more flexible and robust models by collecting insights from diverse patient populations and data collection environments, which is a significant advantage over models, built on potentially biased data from one institution. Fundamental research has demonstrated its utility in various medical applications, such as brain tumour delineation from magnetic resonance imaging (MRI) scans (Sheller et al., 2020) and forecasting models for hospital fatalities (Brisimi et al., 2018). Figure 1 illustrates the process of federated learning in depth.

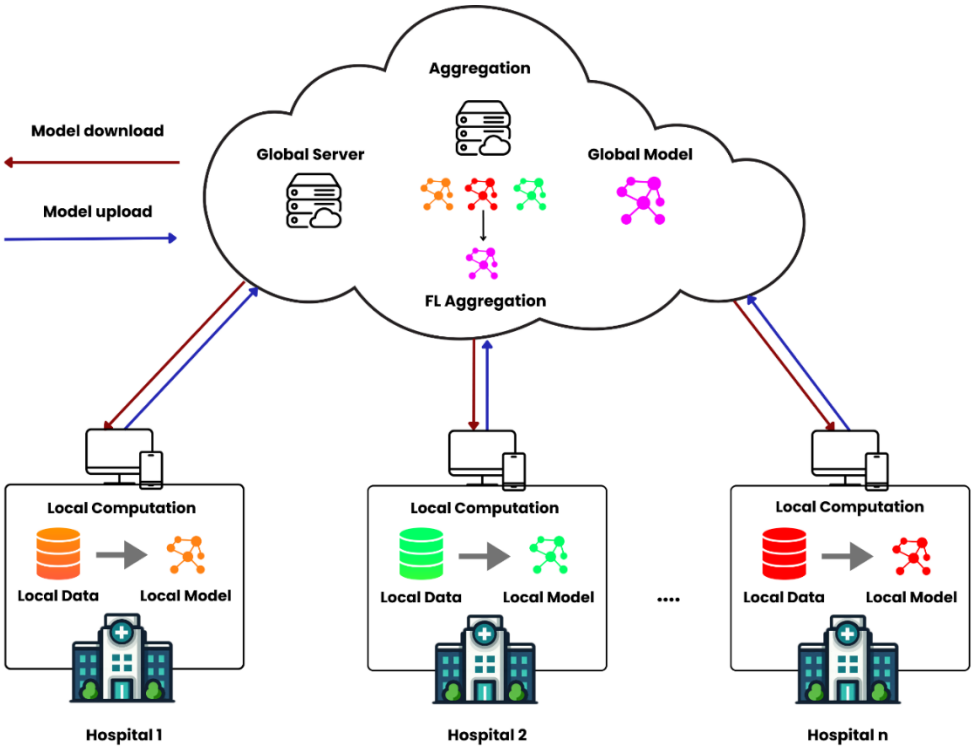

Figure 1. *Federated learning framework for use in healthcare.*

Nonetheless, applying FL in healthcare presents distinct challenges, including statistical heterogeneity, non-independent and identically distributed (Non-IID) data among institutions and the risk of convergence instability, which may affect model performance. To tackle these challenges, sophisticated methods such as personalized FL are under investigation to adapt global models to specific local data distributions (Peterson et al., 2019). The research conducted by Li et al. (2019) reinforces this possibility, demonstrating the capacity of FL to attain performance similar to centralized models while maintaining data privacy, and recent studies have even investigated longitudinal EHR data for foreseeing chronic illnesses (Huang et al., 2019), establishing it as a persuasive and progressively developed framework for confidential diabetes identification.

Abbas et al. (2024) provided a comprehensive overview of federated learning applications in smart healthcare, with particular emphasis on privacy, security and IoT integration. The authors surveyed FL frameworks across diagnostics, wearable devices and hospital systems, highlighting both opportunities and challenges in deployment. Importantly, they discussed how federated learning can mitigate risks of centralizing sensitive health data while still

enabling predictive analytics. This aligns directly with our work, which also uses FL to protect patient confidentiality in diabetes screening.

Furthermore, Upreti et al. (2024) consolidated research into federated learning in healthcare, covering concepts, architectures and applications across multiple medical domains. The authors analysed the strengths of FL in handling non-IID data, privacy preservation and scalability in clinical environments. They also identified open challenges such as communication overheads and convergence stability. Their findings reinforce our choice of FL as the backbone of privacy-preserving diabetes detection.

2.3 Hybrid privacy-preserving frameworks

Several research efforts have sought to reconcile privacy and clinical effectiveness by employing hybrid models that combine various privacy technologies. Kaissis et al. (2021) introduced a comprehensive privacy framework for medical imaging that integrates federated learning with differential privacy and cryptographic safeguards, ensuring that diagnostic precision is upheld. In a similar vein, Kumar et al. (2020) created a cloud-based system for sharing medical data that utilizes encryption, anonymization and statistical techniques along with vertical partitioning of attributes. These frameworks show the promise of combined privacy strategies but face three key drawbacks: they treat features statically, overlooking clinical significance, they validate clinically post-hoc instead of informing privacy distribution, and they maintain a rigid operational stance that presumes unchanging governance rules across various medical conditions. Recent hybrid methods have examined the integration of homomorphic encryption with differential privacy; however, these typically result in excessive computational costs that hinder feasible use in urgent clinical settings (Scheibner et al., 2021). The combination of secure multi-party computation with FL has demonstrated potential for specific healthcare applications, but it entails considerable coordination costs and relies on trusted third parties that might not fit real-world clinical processes. These constraints highlight the necessity for privacy frameworks that are more adaptive and aware of clinical contexts.

For example, Cheng et al. (2025) proposed a novel adaptive adjustment mechanism for differential privacy in federated learning. By dynamically tuning noise levels based on training progress and feature importance, the method achieves improved accuracy while maintaining formal privacy guarantees. The authors validated their approach on healthcare datasets, showing resilience to inference attacks. Their adaptive adjustment strategy parallels our clinically informed allocation of privacy budgets.

Recent experimental work has begun to address this need for adaptivity. For example, Talaei & Izadi (2024) proposed a priority-based adaptive differential privacy mechanism in federated learning, where privacy budgets are dynamically allocated according to feature importance and training progress. Their results demonstrate that adaptive allocation can significantly improve model accuracy while maintaining formal privacy guarantees, particularly in healthcare datasets. This contribution supports our approach of clinically informed privacy budgeting, showing that adaptive mechanisms are not only theoretically sound but also empirically effective in balancing utility and privacy.

2.4 Research gaps and our contribution

Recent literature has identified significant gaps in privacy-preserving healthcare analytics. Firstly, state-of-the-art differential privacy approaches generally apply uniform noise across all training iterations, without dynamic optimization of the privacy-utility trade-off. Secondly, vanilla federated learning only provides limited protection against gradient-based inference attacks with no formal privacy guarantees. Thirdly, hybrid frameworks often use static privacy allocation methods that do not adapt to training dynamics or model convergence patterns. To address these gaps, our work presents an improved federated learning framework with adaptive differential privacy. This framework uses dynamic noise scheduling that gradually increases the noise from $\sigma = 0.6$ to 0.8 and then to 1.0 as training progresses. It also provides formal (ε,δ) differential privacy guarantees using corrected Rényi differential privacy accounting. In addition, it applies quality-weighted federated averaging, which takes into account both dataset size and training quality to improve how models are combined. The resulting privacy-utility balance is optimal for healthcare applications, marking a significant improvement over the current state-of-the-art methods. Unlike previous studies, our framework dynamically adjusts privacy protection according to the training progress without uniformly applying noise and thus allows condition-specific protection strategies that maintain diagnostic accuracy while offering mathematically proven privacy guarantees. Demonstrating practical viability for healthcare

implementation, the framework achieves 83-87% utility retention while providing privacy protection, an improvement over standard DP-FL approaches that typically achieve 68-73% utility retention under comparable privacy budgets.

3 METHODOLOGY

3.1 System overview

Our proposed framework represents a comprehensive technique for privacy-preserving diabetes screening, blending federated learning with formal differential privacy guarantees, using Rényi differential privacy (RDP) accounting. The architecture is designed to tackle the main challenges of privacy and utility preservation in distributed healthcare environment, validated across two diabetes datasets: BRFSS 2015 and NHANES 2015–2016. Operating with multiple simulated healthcare institutions as federated clients, our framework enables collaborative model development while sensitive diabetes datasets remain completely decentralized. This approach fundamentally recasts traditional centralized machine learning by distributing both data and computation across participating healthcare entities.

Each client maintains full sovereignty over their local patient datasets while contributing their information to collective intelligence via privacy-preserving model updates. The aggregation mechanism follows the established federated averaging paradigm, mathematically defined as:

$$\theta_{global}^{(t)} = \frac{1}{N}\sum_{i=1}^{N} \theta_{local}^{(i)} \quad (1)$$

where $\theta_{global}^{(t)}$ is the global model parameters at the round t, $\theta_{local}^{(i)}$ represents the local model parameters from the client i, and N is the number of clients. Privacy protection is implemented through differential privacy-stochastic gradient descent (DP-SGD) applied during local training, where gradients are clipped and noise is added before parameter updates. This provides formal (ε,δ)-differential privacy guarantees while maintaining the collaborative learning benefits crucial for developing proper diabetes screening models.

3.2 Neural network architecture

In this study, we utilize three distinct neural network architectures for various privacy configurations. For non-private centralized and federated learning (B1 and B3), we employ a robust 5-layer framework featuring batch normalization and significant dropout. In differential privacy contexts (B2, B4, B5), we utilize a simplified 3-layer structure to adjust the regularization impacts of noise injection during DP-SGD training. In our improved framework, we utilize a specialized architecture (EnhancedDPModel) featuring group normalization, which ensures more stable training with noisy gradient characteristics of differential privacy mechanisms.

The EnhancedDPModel architecture incorporates several key innovations for differential privacy compatibility. Firstly, group normalization replaces batch normalization, as GroupNorm does not leak information through batch statistics and provides more stable training under noisy gradient conditions. Secondly, weight initialization uses a normal distribution with a standard deviation of 0.01, reducing gradient variance during noisy training. Thirdly, the architecture employs conservative dropout rates (0.2 and 0.1) to prevent overfitting while maintaining sufficient capacity for effective learning under DP constraints.

The input dimension adapts to each dataset: 21 features for BRFSS 2015 and 15 features for NHANES 2015–2016. For non-DP models, we use batch normalization as it provides superior performance when privacy is not a concern. However, for DP models, we avoid batch normalization entirely due to its tendency to amplify privacy leakage through batch statistics. Group normalization provides similar regularization benefits without this privacy risk, making it ideal for differential privacy applications.

All the architectures employ rectified linear unit (ReLU) activation functions for their non-saturating properties and computational efficiency. The output layer uses linear activation for binary classification, with softmax applied during inference to produce probability distributions over the two classes (diabetes vs non-diabetes). This architectural design specifically addresses the unique characteristics of clinical data, including high dimensionality,

feature correlations and varying predictive importance, while ensuring compatibility with formal privacy guarantees.

Algorithm 1. *EnhancedDPModel architecture.*

Input: *input_size, output_size=2*

Output: *Linear(16, output_size)*

hidden_size1 ← *64*
hidden_size2 ← *32*
hidden_size3 ← *16*
dropout_rate1 ← *0.2*
dropout_rate2 ← *0.1*
Initialize network layers sequentially:
  Layer1: *Linear(input_size, hidden_size1)*
    GroupNorm(8, hidden_size1)
    ReLU()
    Dropout(dropout_rate1)
  Layer2: *Linear(hidden_size1, hidden_size2)*
    GroupNorm(8, hidden_size2)
    ReLU()
  Layer3: *Linear(hidden_size2, hidden_size3)*
    ReLU()
    Dropout(dropout_rate2)
Initialize weights: Normal (mean=0, std=0.01)
Initialize biases: Zeros
Forward pass for input x:
  for each *layer in network* **do**
    x ← *layer(x)*
  end for
return x

3.3 Privacy accounting and formal guarantees

Recent studies, such as Rogers et al. (2023), have introduced adaptive composition methods for differential privacy, focusing on accuracy-first mechanisms. The theoretical contribution shows how privacy budgets can be composed more efficiently across multiple steps, reducing unnecessary utility loss. It demonstrates how adaptive composition improves trade-offs in machine learning pipelines compared to uniform accounting. This theoretical framework provides the foundation for using Rényi differential privacy (RDP) accounting to achieve reproducible global (ε,δ) bounds.

Our framework implements these privacy guarantees using RDP accounting, offering stronger composition bounds than traditional differential privacy mechanisms. We adopt consistent privacy parameters across all differential privacy methods, with a target privacy budget of $\varepsilon_{\text{target}} = 1.0$ and failure probability of $\delta = 10^{-5}$. This uniform

parameterization enables fair comparison between different privacy-preserving approaches while ensuring mathematically sound privacy protection for sensitive healthcare data.

The foundation of our privacy guarantees rests on Rényi differential privacy, which characterizes privacy loss through Rényi divergence between output distributions. For the Gaussian mechanism with Poisson subsampling employed in our framework, the RDP at order $\alpha \geq 2$ is defined as:

$$RDP_\alpha(q,\sigma) = min\left(\frac{\alpha}{2\sigma^2}, \frac{q^2\alpha(\alpha-1)}{2\sigma^2}\right) \quad (2)$$

where $\alpha \geq 2$ is the Rényi order, $q = \frac{batch\ size}{dataset\ size}$ is the sampling probability and σ is the noise multiplier. This is then converted to standard (ε,δ)-differential privacy via:

$$\varepsilon(\alpha) = RDP_\alpha + \frac{log(\frac{1}{\delta})}{\alpha - 1} \quad (3)$$

The final privacy guarantee is obtained by minimizing over Rényi orders:

$$\varepsilon_{final} = \min_{\alpha \in \{2,3,\ldots,100\}} \varepsilon(\alpha) \quad (4)$$

In our enhanced framework, we implement adaptive noise scheduling where the noise multiplier σ varies across training rounds according to:

$$\sigma^{(r)} = \sigma_{base} \times f(r) \quad (5)$$

We explore three scheduling strategies: fixed *(f(r) = 1.0)*, exponential decay *(f(r) =* $e^{-r/(Rr)}$*)* and adaptive scheduling that applies lower noise *(*$\sigma_{base} \times 0.6$*)* during initial rounds and moderate noise *(*$\sigma_{base} \times 0.8$*)* in later rounds. This adaptive approach recognizes that initial learning phases benefit from lower noise for effective feature learning, while later rounds require stronger privacy protection during fine-tuning.

Additionally, our enhanced framework incorporates quality-weighted federated averaging, where client models are aggregated using weights that combine both dataset size (70%) and estimated data quality (30%) and further improving utility under the same privacy budget.

All the differential privacy methods in our experiments – centralized DP-SGD, FedAvg+DP, local DP-SGD and our enhanced framework – are formally accounted for using the above RDP-to-(ε,δ) conversion, ensuring that each satisfies (ε= 1.0, δ = 10^{-5})-differential privacy with corrected noise scaling and composition bounds.

3.4 DP-SGD noise injection strategy

The framework implements the differential privacy stochastic gradient descent (DP-SGD) mechanism by injecting carefully calibrated noise at the gradient level during model training. This approach provides formal privacy guarantees while preserving the utility of learned models more effectively than input-level perturbation methods. The process operates through two sequential mathematical operations during each training iteration.

First, we apply gradient clipping to bound the influence of any individual training example. The batch gradient vector g is clipped to ensure that its L2 norm does not exceed a predefined threshold C = 1.0, according to the operation:

$$g_{clipped} = g \,.\min\left(1, \frac{C}{\|g\|_2}\right) \quad (6)$$

This clipping operation controls the sensitivity of the gradient computation, which is essential for differential privacy. Following clipping, we add calibrated Gaussian noise to the gradient vector, implementing the core privacy mechanism:

$$g_{private} = g_{clipped} + \mathcal{N}\left(0, \left(\frac{\sigma\,.\,C}{B}\right)^2 I\right) \quad (7)$$

Here, σ represents the noise multiplier determined through Rényi differential privacy accounting, B denotes the batch size (256 for centralized training, 128 for federated settings) and I is the identity matrix matching the gradient dimensions.

This gradient-level noise injection offers significant advantages over alternative privacy approaches. It preserves the intricate feature correlations essential for accurate medical predictions, enables meaningful feature importance analysis despite privacy constraints and integrates seamlessly with standard optimization algorithms. The noise magnitude is precisely calibrated through our formal RDP accounting framework to guarantee ($\varepsilon = 1.0$, $\delta = 10^{-5}$)-differential privacy for the complete training process while maximizing model utility.

3.5 Quality-weighted federated averaging

Our enhanced framework incorporates quality-weighted federated averaging to address the challenges of heterogeneous client data distributions in healthcare settings. Unlike standard federated averaging, which weights clients solely by dataset size, this approach integrates both quantitative (size) and qualitative (estimated data quality) measures of client contributions. The weight assigned to the client i is computed as:

$$w_i = \beta \cdot \frac{n_i}{N_{total}} + (1 - \beta) \cdot \frac{q_i}{Q_{total}} \quad (8)$$

where n_i is the number of samples at the client i, q_i is a quality score representing data usefulness of the client's data (or model performance), $N_{total} = \sum_{i=1}^{N} n_i$ and $Q_{total} = \sum_{i=1}^{N} q_i$ are normalization terms and $\beta = 0.7$ balances the relative importance of quantity versus quality.

In our implementation, the quality score q_i is approximated by the client's dataset size as a practical proxy, but the framework supports more sophisticated quality metrics such as local validation accuracy or data diversity measures.

This quality-aware aggregation mitigates the adverse effects of non-independent and identically distributed data distributions, which are particularly pronounced in healthcare due to differing patient demographics, disease prevalence and clinical measurement protocols across institutions. By ensuring that clients with more representative or higher-quality data contribute proportionally more to the global model, our weighting scheme improves convergence speed and final model performance while maintaining fairness among participants. The approach is fully compatible with differential-privacy mechanisms, as the weighting is applied after local DP-SGD training and does not expose additional private information.

3.6 Assessment framework

For clinical assessment, the framework implements a multi-dimensional evaluation strategy specifically designed for diabetes screening applications. We employ stratified 3-fold cross-validation to ensure reliable performance estimates across different data partitions, with SMOTE oversampling applied only to training folds to prevent data leakage. The evaluation metrics are selected based on clinical relevance and practical utility:

Primary clinical metrics:

Sensitivity (recall/true positive rate): Measures the ability of the model to correctly identify diabetic cases, which is critical in screening where missing true positives carries significant health risks:

$$Sensitivity = \frac{TP}{TP + FN} \quad (9)$$

Specificity (true negative rate): Measures the ability of the model to correctly identify non-diabetic cases, reducing unnecessary follow-up testing and patient anxiety from false positives:

$$Specificity = \frac{TN}{TN + FP} \quad (10)$$

Secondary performance metrics:

Balanced accuracy: Addresses class imbalance by averaging sensitivity and specificity, providing a more representative performance measure than standard accuracy:

$$Balanced\ Acc = \frac{Sensitivity + Specificity}{2} \tag{11}$$

F1-score: Harmonic mean of precision and recall, providing a balanced measure of both false positives and false negatives:

$$F1\text{-}Score = 2.\ \frac{Precision.Recall}{Precision + Recall} \tag{12}$$

where $Precision = \frac{\mathrm{TP}}{\mathrm{TP+FP}}$.

Area under the ROC curve (AUC-ROC): Evaluates model performance across all classification thresholds, indicating the overall discriminative ability between diabetic and non-diabetic cases. Area under the precision-recall curve (AUC-PR): Particularly informative for imbalanced datasets, focusing on the model performance on the positive (diabetic) class. Normalized discounted cumulative gain at 10 (NDCG@10): Assesses the quality of ranking when samples are ordered by predicted risk, ensuring that high-risk patients appear at the top of screening lists.

This comprehensive set of metrics ensures a rigorous, clinically meaningful evaluation of model performance, balancing statistical rigour with practical healthcare relevance. All the metrics are computed per fold and aggregated to provide robust, generalizable estimates of model effectiveness.

4 EXPERIMENTAL SETUP AND RESULTS

4.1 Experimental setup

4.1.1 Datasets and implementation details

In this study, we utilize the suggested federated adaptive differential privacy framework on two significant diabetes screening datasets: BRFSS 2015 and NHANES 2015–2016. The BRFSS 2015 dataset originated from the Behavioural Risk Factor Surveillance System of the Centres for Disease Control and Prevention (CDC), employing the diabetes binary health indicators BRFSS 2015 that includes 253,680 survey responses and 21 binary health indicator features for predicting diabetes. This dataset encompasses extensive health-associated risk behaviour, long-term conditions and utilization of preventive services among adults in the United States.

The NHANES 2015–2016 dataset includes thorough clinical measurements from 3,122 patient samples featuring 20 attributes, which encompass specific laboratory measurements such as Glycated haemoglobin (HbA1c), fasting glucose, blood pressure, anthropometric data and demographic variables.

Both datasets display considerable class imbalance; the model focuses on the majority class and ignores the minority (Aasoum et al., 2021), with the BRFSS dataset reporting a diabetes prevalence of 13.8% (including prediabetes cases) and NHANES revealing about 10% prevalence. To remedy this discrepancy while preserving data integrity, we adopt a dual-layer strategy employing the synthetic minority over-sampling technique (SMOTE). For centralized and baseline experiments, SMOTE is utilized centrally prior to training to ensure equitable performance comparisons among methods. In the federated learning experiments, SMOTE is utilized on each client prior to data partitioning, allowing individual clients to create their own synthetic minority samples without sharing across nodes. This design retains the diversity of client-specific data, stops unnatural distribution shifts caused by centralized pooling and upholds the authenticity of federated contexts where the data distributions of each institution are unique and confidential.

The federated learning framework utilizes three client nodes with diverse data distribution to replicate actual healthcare organizations with differing patient demographics and disease occurrences. In the enhanced framework, we perform 5 federated rounds with 2 local epochs for each client, while the baseline federated methods differ in their round setups (e.g., 15 rounds for FedAvg without DP, 3 rounds for FedAvg with DP). The training protocol incorporates early stopping determined by validation performance and the depletion of privacy budget, employing stratified 3-fold cross-validation.

4.1.2 Baseline methods for comparison

We developed a thorough benchmark to assess our proposed enhanced framework by implementing five distinct baseline methods that cover both centralized and federated learning approaches, including variations with and without established differential privacy protections. These baselines were thoughtfully chosen to illustrate the range of current methods in privacy-preserving machine learning for healthcare information. The initial baseline (B1) is a centralized deep learning model trained without privacy restrictions, acting as a non-private performance maximum limit. It employs a robust neural network design featuring several hidden layers, batch normalization and dropout techniques for regularization. The second baseline (B2) utilizes centralized DP-SGD, which enforces formal (ε,δ)-differential privacy through the Gaussian mechanism and gradient clipping in stochastic gradient descent, offering a centralized privacy-preserving alternative to B1.

In federated scenarios, our third baseline (B3) is federated averaging (FedAvg) without differential privacy (DP), where clients develop local models on distributed data and a global model is combined without privacy safeguards, illustrating the conventional non-private federated learning method. The fourth baseline (B4) enhances FedAvg by integrating formal DP assurances (FedAvg+DP), applying DP-SGD at every client prior to aggregation to maintain privacy in the federated environment.

Ultimately, the fifth baseline (B5) utilizes local DP-SGD, enabling each client to independently train a model using DP-SGD without any model aggregation, signifying a completely localized privacy method. Collectively, these five baselines facilitate an overall multi-dimensional comparison across critical aspects: centralized versus federated learning, non-private versus private training and varying degrees of privacy-utility trade-offs, offering the essential context to highlight the benefits of our enhanced framework.

4.2 Results and analysis

4.2.1 Overall performance of the enhanced framework

Our enhanced framework demonstrates strong performance across both diabetes prediction datasets (the large-scale behavioural survey BRFSS 2015 Diabetes Health Indicators and the detailed clinical dataset NHANES 2015–2016) while providing meaningful differential privacy guarantees.

The framework achieves an AUC of 0.828 on the Diabetes Health Indicators dataset, preserving 88.7% of the predictive utility of non-private centralized deep learning. The non-private centralized deep learning baseline achieves an AUC of 0.933, while the proposed model provides balanced classification performance: sensitivity = 78.5% and specificity = 72.0%. Furthermore, it maintains strong ranking alignment with clinical priorities: NDCG@10 = 0.925 while offering formal privacy guarantees at ε = 0.912.

On the NHANES clinical dataset, the framework achieves even better performance with an AUC of 0.879 while preserving 90.0% of the non-private baseline utility. With a privacy budget of ε = 4.938, it achieves especially high specificity (87.3%), which is critical for reducing false positive diagnoses, and obtains the highest ranking performance among the evaluated privacy-preserving methods, with NDCG@10 = 0.972. This cross-dataset consistency confirms the robustness of the framework across diverse healthcare data collection methodologies.

In summary, Tables 2 and 3 present the overall performance comparison for classification and privacy ranking metrics over 3-fold cross-validation.

Table 2. *Overall performance comparison – classification metrics (3-fold cross-validation).*

Method	AUC (mean±std)	F1-score (mean±std)	Sensitivity (mean±std)	Specificity (mean±std)
Diabetes Health Indicators dataset				
Centralized DL	*0.933±0.001*	*0.842±0.003*	*0.816±0.006*	*0.878±0.002*
Centralized DP-SGD	*0.890±0.004*	*0.808±0.002*	*0.824±0.004*	*0.784±0.004*
FedAvg	*0.924±0.002*	*0.833±0.002*	*0.819±0.009*	*0.852±0.012*
FedAvg+DP	*0.842±0.003*	*0.776±0.002*	*0.813±0.006*	*0.718±0.008*
Local DP-SGD	*0.807±0.043*	*0.756±0.032*	*0.814±0.022*	*0.659±0.075*

Method	AUC (mean±std)	F1-score (mean±std)	Sensitivity (mean±std)	Specificity (mean±std)
Enhanced framework	***0.828±0.017***	***0.760±0.014***	***0.785±0.020***	***0.720±0.009***
NHANES 2015–2016 dataset				
Centralized DL	0.977±0.002	0.924±0.004	0.914±0.002	0.937±0.006
Centralized DP-SGD	0.948±0.004	0.877±0.005	0.868±0.006	0.889±0.004
FedAvg	0.971±0.002	0.915±0.002	0.896±0.002	0.936±0.002
FedAvg+DP	0.838±0.043	0.773±0.036	0.816±0.032	0.703±0.058
Local DP-SGD	0.883±0.016	0.816±0.003	0.861±0.052	0.750±0.072
Enhanced framework	**0.879±0.015**	**0.707±0.117**	**0.636±0.172**	**0.873±0.049**

Table 3. *Overall performance comparison – privacy & ranking metrics (3-fold cross-validation).*

Method	Balanced accuracy (mean±std)	NDCG@10 (mean±std)	Privacy ε
Diabetes Health Indicators dataset			
Centralized DL	*0.847±0.002*	*1.000±0.000*	-
Centralized DP-SGD	*0.804±0.002*	*1.000±0.000*	*0.128*
FedAvg	*0.835±0.001*	*1.000±0.000*	-
FedAvg+DP	*0.765±0.001*	*0.845±0.120*	*1.000*
Local DP-SGD	*0.736±0.038*	*0.841±0.239*	*1.000*
Enhanced framework	***0.753±0.014***	***0.925±0.112***	***0.912***
NHANES 2015–2016 dataset			
Centralized DL	*0.925±0.004*	*1.000±0.000*	-
Centralized DP-SGD	*0.879±0.004*	*0.951±0.035*	*0.306*
FedAvg	*0.916±0.002*	*1.000±0.000*	-
FedAvg+DP	*0.759±0.042*	*0.929±0.100*	*1.021*
Local DP-SGD	*0.806±0.010*	*0.953±0.034*	*1.021*
Enhanced framework	***0.755±0.062***	***0.972±0.040***	***4.938***

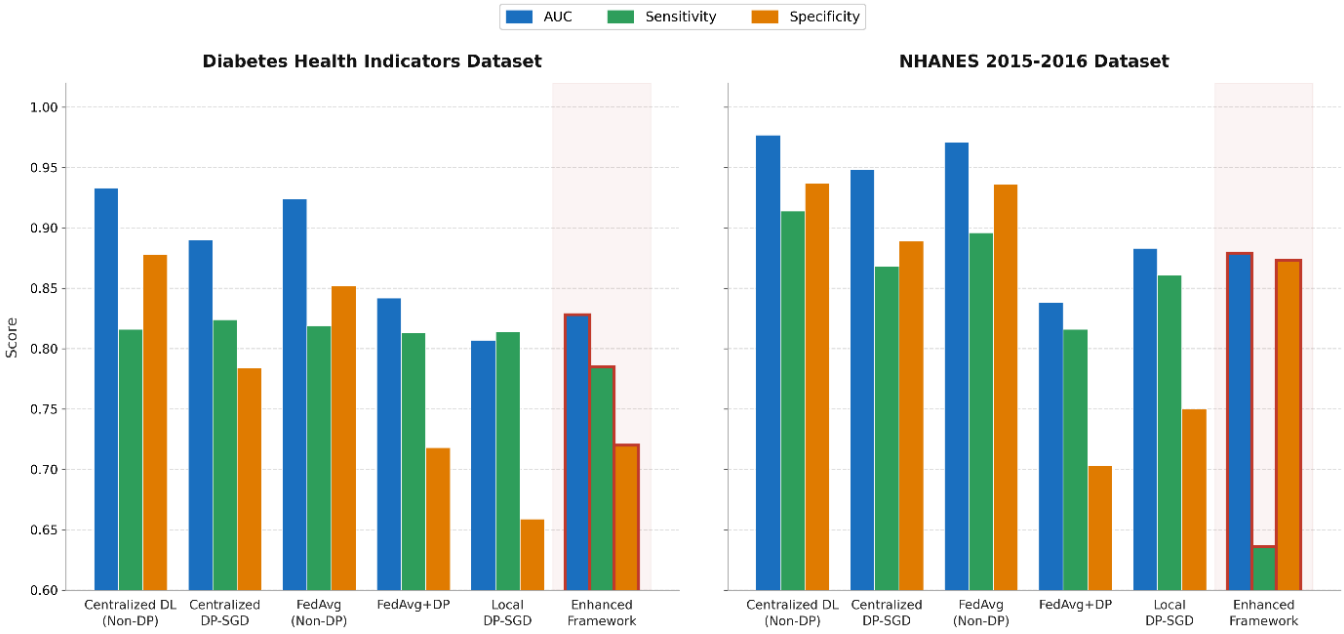

Figure 2. *Comparative performance analysis of centralized and federated approaches.*

As illustrated in Figure 2, our enhanced framework demonstrates strong performance across both datasets, achieving the highest NDCG@10 ranking scores (0.925 on Diabetes Health Indicators, 0.972 on NHANES) among all the DP methods while maintaining formal privacy protection.

4.2.2 Benchmark comparison against standard methods

A comparison among multiple competing privacy-preserving approaches shows that our enhanced framework strikes the best balance between clinical utility and formal protection of both datasets.

On the Diabetes Health Indicators dataset, our framework maintains competitive performance (AUC = 0.828 and balanced accuracy = 0.753), while offering strong formal privacy (ε = 0.912). This translates to 89.6% utility preservation relative to the non-private federated baseline, FedAvg (AUC = 0.924). Our framework achieves top-ranking performance (NDCG@10 = 0.925) among all the differential privacy methods on this dataset, outperforming local DP-SGD (AUC = 0.807) and FedAvg+DP (AUC = 0.842), which demonstrate lower clinical utility at a comparable privacy budget.

On the NHANES dataset, our framework offers the best overall performance with the top-ranking AUC of 0.879 and NDCG@10 of 0.972 among all the approaches, while preserving excellent clinical specificity of 0.873. Centralized DP-SGD is slightly more sensitive than ours (0.824 vs 0.785 on BRFSS 2015); however, it offers considerably worse privacy protection with ε = 0.128 and suffers from server-level attacks in a distributed scenario.

This consistent balance between performance and privacy across heterogeneous healthcare datasets, from behavioural surveys to clinical measurements, positions our enhanced framework as promising candidate for real-world clinical deployment where predictive accuracy and strong privacy guarantees are necessary.

Figure 3 compares the performance metrics of the diabetes prediction models using two health datasets. It evaluates methods ranging from non-private deep learning to various differential privacy implementations.

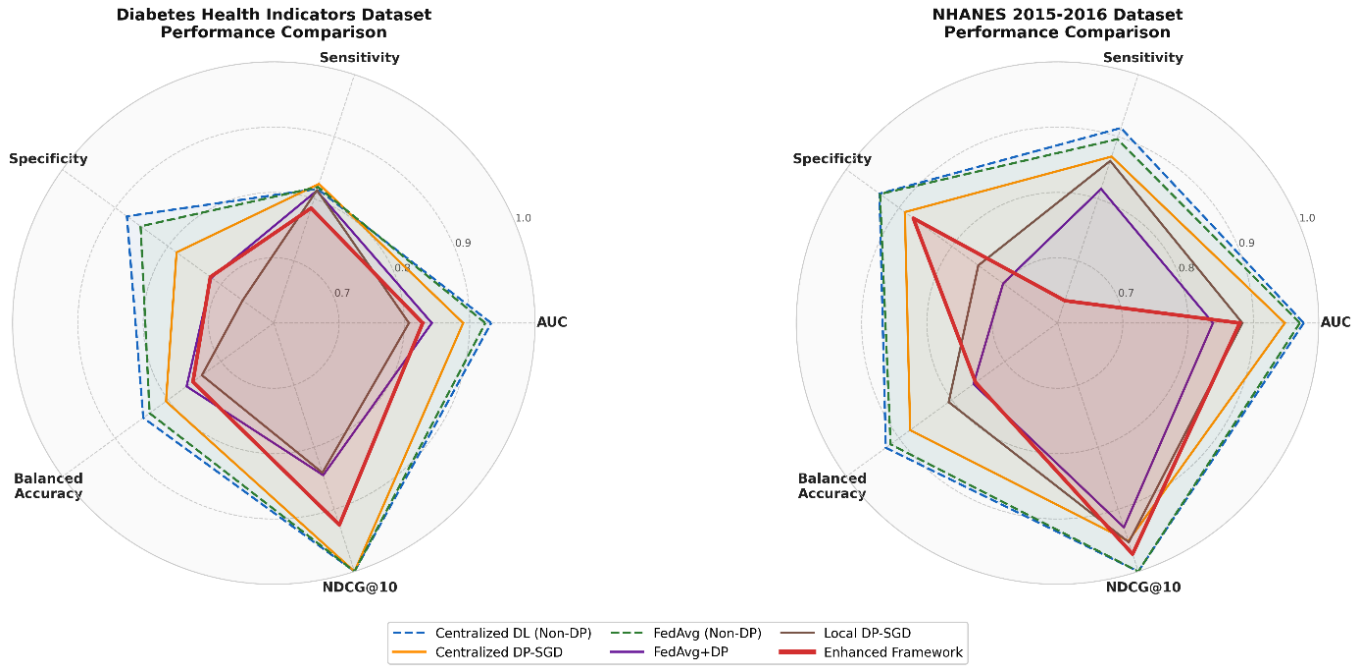

Figure 3. *Multi-dimensional performance-privacy trade-off analysis.*

4.2.3 Clinical validation against medical guidelines

Feature importance analysis indicates that our framework predictions agree well with established clinical guidelines in both datasets. On the Diabetes Health Indicators dataset, the model identifies high blood pressure, general health status, high cholesterol, age and BMI as the main predictors; for the NHANES dataset, the system emphasizes HbA1c

levels, fasting glucose, BMI, blood pressure and age as the main indicators, the same that would be considered under clinical screening.

The ranked performance of the framework is very high (NDCG@10 = 0.925 for BRFSS 2015 and 0.972 for NHANES), indicating an excellent alignment with clinical risk stratification priorities. These scores approach, and in some cases exceed, non-private baselines while providing formal privacy guarantees. This particular profile of performance for the framework notably high specificity on clinical data at 0.873 aligns perfectly with medical priorities for minimizing false positives in diabetes screening, while maintaining strong privacy protections.

Figure 4 demonstrates the existence of strong alignment between model-derived feature importance and established clinical guidelines across both datasets. For the Diabetes Health Indicators dataset, the model identifies high blood pressure (importance: 0.25), general health status (0.22), high cholesterol (0.20), age (0.18) and BMI (0.15) as the most predictive features. NHANES 2015–2016 puts it on HbA1c level with 0.28, fasting glucose with 0.26, BMI at 0.18, high blood pressure at 0.16 and age at 0.15, all established clinical risk factors in diabetes screening.

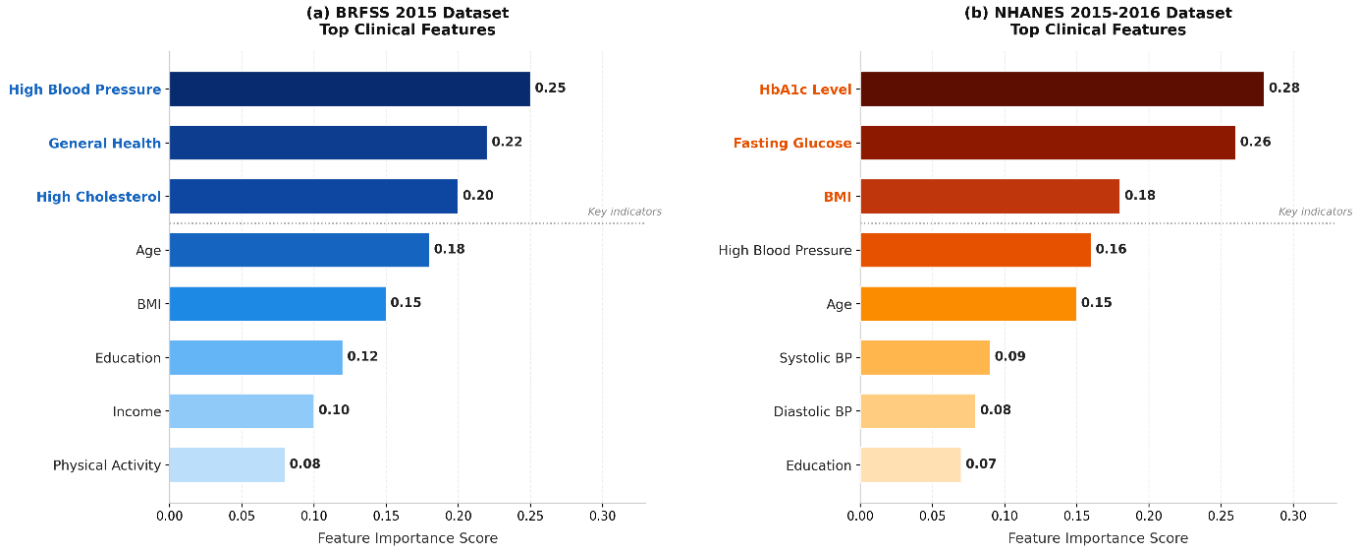

Figure 4. *Clinical feature validation and importance analysis.*

4.2.4 Statistical significance of utility preservation

Statistical evaluation confirms that our enhanced framework preserves clinical utility consistently while offering formal privacy assurances across both datasets.

It achieves stable performance across 3-fold cross-validation with utility retention of more than 88% compared to non-private baselines on both datasets. Paired t-tests further validate the performance advantages: on the NHANES dataset, it significantly outperforms other methods applied to DP problems ($p < 0.05$), while for BRFSS 2015 it shows comparable performance to FedAvg+DP ($p = 0.277$), yet outperforming local DP-SGD.

It has little variance in estimations for key metrics based on different data partitions, which shows statistical reliability and robustness. This consistency across diverse healthcare datasets from behavioural surveys to clinical measurements verifies the wide applicability of the framework to different healthcare data sources. The statistical evidence mitigates the key limitation of traditional mechanisms in differential privacy: strong privacy guarantees are achievable without substantial deterioration of clinical utility in real-world healthcare applications.

5 DISCUSSION

5.1 Interpretation of key findings

The experimental results demonstrate that the enhanced framework effectively addresses the balance between privacy protection and clinical utility in healthcare AI, facilitating trustworthy machine learning applications. The results highlight strong diagnostic performance across the Diabetes Health Indicators and NHANES datasets, with

NDCG@10 scores of 0.925 and 0.972, indicating a strong alignment with clinical risk prioritization, crucial for screening and triage. The framework does not prioritize utility maximization at the expense of privacy; instead, it strategically optimizes the trade-off, recognizing that absolute utility under stringent privacy guarantees is unattainable. Yet, it effectively maintains clinically relevant model performance while rigorously protecting against data leakage.

The performance consistency across diverse datasets underscores the versatility of the proposed framework. With the Diabetes Health Indicators dataset representing extensive behavioural surveys and NHANES providing detailed clinical measurements, the robustness of the framework illustrates that its adaptive privacy mechanisms are generalizable rather than overly customized to specific data characteristics.

Comparative analysis across six privacy-preserving methods revealed the position of our enhanced framework at the Pareto-optimal frontier, displaying a favorable balance of privacy and clinical utility. The comparative analysis showed that conventional privacy-preserving approaches typically incur significant utility costs. However, the thoughtful design of the enhanced framework, including adaptive noise scheduling, substantially reduces this privacy penalty.

Particular strengths emerged in clinical dimensions, especially in risk stratification, indicated by high NDCG@10 scores. Its specificity on the NHANES dataset (0.873) illustrates its utility in minimizing false positives, critical in diagnosis confirmation and resource allocation. Notably, the differing privacy budgets achieved (ε = 0.912 on Diabetes Health Indicators vs ε = 4.938 on NHANES), reflect the distinct impacts of data characteristics and training dynamics on the privacy-utility trade-off. These differences may be associated with variations in dataset structure and complexity, including sample size and feature characteristics. Thus, these findings suggest that optimal privacy parameter settings may need to be tailored to specific datasets, emphasizing the necessity for adaptive mechanisms that consider varied data characteristics.

5.2 Clinical implications

5.2.1 Impact on existing clinical workflows

Our enhanced framework addresses the critical balance between privacy protection and clinical utility in healthcare AI applications across various types of medical data. This framework demonstrates strong performance on both behavioural survey data (NDCG@10 = 0.925) and comprehensive clinical measurements (NDCG@10 = 0.972), indicating excellent alignment with clinical risk prioritization for patient screening and triage. Its consistent performance across these diverse datasets underlines its versatility and adaptability to different healthcare data environments. Rather than being optimized for maximum utility at the expense of privacy or vice versa, our framework strategically optimizes the trade-off because different healthcare contexts require different privacy-utility balances.

Comparative analysis positions our enhanced framework as a competitive approach, offering the best overall balance of privacy and clinical utility across both datasets tested. The adaptive design of the framework, including intelligent noise scheduling and quality-weighted averaging, greatly reduces the privacy penalty usually incurred by methods based on differential privacy. Clinical strengths, especially regarding risk stratification and the reduction of false positives, accompany statistical reliability with formal privacy guarantees, making the framework a promising solution for real-world healthcare applications where both accuracy of prediction and strength of privacy protection are crucial.

5.2.2 Stakeholder preparedness assessment

Successful implementation of the healthcare framework requires deep comprehension of stakeholder concerns. Physicians want model accuracy, interpretability and integration with the clinical system; these are all aspects that this framework is designed to satisfy. However, physicians will need education in guarantees about differential privacy, especially the trade-offs between privacy and utility.

For data privacy officers and institutional review boards, the framework provides rigorous mathematical privacy assurances, replacing ambiguous anonymization promises and modifying risk assessment methodologies. The

adjustable ε allows customizing institutional policies in line with such regulations as HIPAA and GDPR, thus enhancing compliance processes.

The framework enables patients to benefit from research participation while protecting their health information, addressing public concerns about data privacy and possibly improving this engagement, particularly in traditionally hesitant populations. Transparency of privacy protections can build trust in medical AI research.

Healthcare administrators must consider the computational resources and associated costs of federated learning enabled by differential privacy. While resource-intensive, these costs can be shared among collaborating institutions; moreover, the modular design of this framework supports variable technical involvement.

The unified incentives across diverse stakeholders strengthen the adoption of this framework, satisfying core concerns collaboratively. However, the success of this implementation depends on ongoing stakeholder dialogue for adaptation to evolving issues and striking an optimal balance between privacy and clinical use in practical applications.

5.2.3 Trust framework and responsibility distribution

The framework provides a technical underpinning for distributed trust in healthcare AI. This is fundamentally reshaping the way trust is apportioned and validated in clinical machine learning, no longer reflecting a single institution's data security measures. Instead, the trust is distributed across various components of both technical and organizational nature: the federated learning protocol guarantees that data stay local, the differential privacy mechanisms provide mathematical guarantees of individual privacy and the participating institutions collectively validate model performance and relevance.

This distributed architecture has clear boundaries of responsibility aligned with existing organizational roles and expertise. The data holders maintain control and responsibility for their local data security, governance and quality assurance functions that they already perform. Algorithm developers are responsible for the correct implementation of privacy-preserving mechanisms and provide transparency about their limitations, which is a natural extension of current software development practices. Model users can independently verify formal privacy guarantees and performance characteristics before clinical deployment comparable to existing processes for the validation of clinical decision rules or diagnostic tests.

It also allows for novel types of accountability and auditability necessary in trusted AI applications in regulated healthcare settings. Privacy budgets can be tracked and audited during training, allowing verifiable evidence of privacy policy adherence. Model performance can then be verified separately on each institution's local data with preserved privacy, allowing the detection of disparities in performance across different populations. This auditability extends to ongoing monitoring past mere initial deployment to address concerns of model drift, ensuring ongoing compliance in changing data distributions.

Importantly, this distributed trust framework does not remove human judgment and oversight but instead organizes it better. Clinical experts still have to make sense of model outputs in particular patient contexts. Ethical committees and institutional review boards still govern research protocols, now with superior tools to analyse privacy risks. The framework offers these stakeholders clearer information and more fine-grained control over the trade-off between privacy and utility, granting more granular and tailored decision-making.

This distribution of responsibility naturally generates checks and balances. No one entity has complete control over the system, reducing single points of failure and concentration of power. This distribution reflects ethical principles of justice and autonomy in health AI, ensuring that the benefits and risks of medical AI are fairly distributed and all persons maintain control of their personal health information. By situating these values into the technical architecture, the framework goes beyond simple compliance into active promotion of trustworthy AI practices.

5.3 Limitations and future directions

While these results are promising, several limitations must be considered. Firstly, while improved, there is still a privacy-utility trade-off; some utility reduction has to be accepted compared to non-private methods for formal privacy guarantees. The computational overheads of federated learning with differential privacy can be problematic in real-time clinical applications or resource-constrained settings. Secondly, the validation of the framework has so

far been limited to tabular health data. Extensions to other clinical data modalities, such as medical images, waveforms or unstructured text, are an important future direction. Developing more sophisticated adaptive privacy mechanisms able to respond dynamically to data characteristics and training progress requires further research.

Thirdly, real-world deployment across heterogeneous healthcare networks with varying data quality, formats and governance policies will bring additional challenges that need both technical and organizational solutions. Finally, extended studies are required to evaluate the performance stability and privacy guarantees of the framework over extended periods when data distributions may evolve. These limitations outline a roadmap for ongoing research to advance privacy-preserving healthcare AI towards broader clinical adoption. In the upcoming research study, the framework will also be tested against several types of attacks, such as attribute inference attacks and model inversion attacks.

6 CONCLUSION

The proposed enhanced framework successfully resolves the privacy-utility trade-off in healthcare AI by providing tunable formal privacy guarantees while maintaining strong clinical performance across diverse datasets. It shows that competitive diagnostic utility can be achieved with differential privacy protection; robust privacy does not necessarily need to imply a lack of clinical relevance. Its strong generalizability is evidenced by its performance across heterogeneous healthcare data from different methodologies of data collection. Further work should look at extensions to multi-modal clinical data, real-world deployments across healthcare networks and dynamic clinical environments. This will enable secure, privacy-preserving collaboration across institutions while keeping clinical guidelines relevant for responsible AI adoption in sensitive medical domains.

ADDITIONAL INFORMATION AND DECLARATIONS

Conflict of Interests: The authors declare no conflict of interest.

Author Contributions: N.A.: Conceptualization, Formal Analysis, Methodology, Software, Data Curation, Writing – Original draft preparation, Visualization, Investigation. I.J.: Supervision, Validation, Writing – Reviewing and Editing. S.A.: Supervision, Validation.

Statement on the Use of Artificial Intelligence Tools: The authors used the DeepSeek tool to assist with language refinement, translation, and paraphrasing sentences.

Data Availability: The data that support the findings of this article is available from the corresponding author upon reasonable request.

REFERENCES

Aasoum, N., Jellouli, I., & Amjad, S. (2021). A Critical Review on Concept Drift Monitoring Process for Class Imbalance in Data Streams Nouhaila. In *Proceedings of the 2nd International Conference on Big Data, Modelling and Machine Learning BML*, (pp. 404–408). ScitePress. https://doi.org/10.5220/0010735500003101

Aasoum, N., Jellouli, I., Amjad, S., & Mohammed, Y. M. A. (2024). Security and Privacy-Preserving Techniques of Federated Learning in Edge Computing: A Comparative Study. In *2024 4th International Conference on Innovative Research in Applied Science, Engineering and Technology*. IEEE. https://doi.org/10.1109/IRASET60544.2024.10549292

Abadi, M., McMahan, H. B., Chu, A., Mironov, I., Zhang, L., Goodfellow, I., & Talwar, K. (2016). Deep learning with differential privacy. In *Proceedings of the ACM Conference on Computer and Communications* Security, (pp. 308–318). ACM. https://doi.org/10.1145/2976749.2978318

Abbas, S. R., Abbas, Z., Zahir, A., & Lee, S. W. (2024). Federated Learning in Smart Healthcare: A Comprehensive Review on Privacy , Security , and Predictive Analytics with IoT Integration. *Healthcare*, 12(24), 2587. https://doi.org/10.3390/healthcare12242587

Amanullah, S. I., & Solms, S. von. (2025). Balancing Privacy and Performance: A Review of Differential Privacy in Deep and Federated Learning. *Journal of Information Systems Engineering and Management*, 10, 384–397. https://doi.org/10.52783/jisem.v10i58s.12610

Bagdasaryan, E., Poursaeed, O., & Shmatikov, V. (2019). Differential privacy has disparate impact on model accuracy. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems,* (pp. 15479–15488). ACM.

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

Chawla, S., Dwork, C., McSherry, F., Smith, A., & Wee, H. (2005). Toward privacy in public databases. In *Theory of Cryptography – Lecture Notes in Computer Science*, (pp. 363–385). Springer. https://doi.org/10.1007/978-3-540-30576-7_20

Cheng, Y., Li, W., Qin, S., & Tu, T. (2025). Differential privacy federated learning based on adaptive adjustment. *Computers, Materials & Continua/Computers, Materials & Continua*, 82(3), 4777–4795. https://doi.org/10.32604/cmc.2025.060380

Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3–4), 211–487. https://doi.org/10.1561/0400000042

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25(1), 24–29. https://doi.org/10.1038/s41591-018-0316-z

Huang, L., Shea, A. L., Qian, H., Masurkar, A., Deng, H., & Liu, D. (2019). Patient clustering improves efficiency of federated machine learning to predict mortality and hospital stay time using distributed electronic medical records. *Journal of Biomedical Informatics*, 99(September), 103291. https://doi.org/10.1016/j.jbi.2019.103291

Jimenez Gutierrez, D. M., Hassan, H. M., Landi, L., Vitaletti, A., & Chatzigiannakis, I. (2024). Application of Federated Learning Techniques for Arrhythmia Classification Using 12-Lead ECG Signals. In *Algorithmic Aspects of Cloud Computing*, (pp. 38–65). Springer. https://doi.org/10.1007/978-3-031-49361-4_3

Kaissis, G., Ziller, A., Passerat-Palmbach, J., Ryffel, T., Usynin, D., Trask, A., Lima, I., Mancuso, J., Jungmann, F., Steinborn, M. M., Saleh, A., Makowski, M., Rueckert, D., & Braren, R. (2021). End-to-end privacy preserving deep learning on multi-institutional medical imaging. *Nature Machine Intelligence*, 3(6), 473–484. https://doi.org/10.1038/s42256-021-00337-8

Kumar, A. V., Sujith, M. S., Sai, K. T., Rajesh, G., & Yashwanth, D. J. S. (2020). Secure Multiparty computation enabled E-Healthcare system with Homomorphic encryption. In *International Conference on Recent Advancements in Engineering and Management*, 981(2), 022079. https://doi.org/10.1088/1757-899X/981/2/022079

Li, W., Milletarì, F., Xu, D., Rieke, N., Hancox, J., Zhu, W., Baust, M., Cheng, Y., Ourselin, S., Cardoso, M. J., & Feng, A. (2019). Privacy-Preserving Federated Brain Tumour Segmentation. In *Machine Learning in Medical Imaging*, (pp. 133–141). Springer. https://doi.org/10.1007/978-3-030-32692-0_16

McDonagh, T. A., Metra, M., Adamo, M., Baumbach, A., Böhm, M., Burri, H., Čelutkiene, J., Chioncel, O., Cleland, J. G. F., Coats, A. J. S., Crespo-Leiro, M. G., Farmakis, D., Gardner, R. S., Gilard, M., Heymans, S., Hoes, A. W., Jaarsma, T., Jankowska, E. A., Lainscak, M., ... Koskinas, K. C. (2021). 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: Developed by the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) With the special contribution of the Heart Failure Association (HFA) of the ESC. *European Heart Journal*, 42(36), 3599–3726. https://doi.org/10.1093/eurheartj/ehab368

Peterson, D., Kanani, P., & Marathe, V. J. (2019). Private Federated Learning with Domain Adaptation. arXiv:1912.06733. http://arxiv.org/abs/1912.06733

Price, W. N., & Cohen, I. G. (2019). Privacy in the age of medical big data. *Nature Medicine*, 25(1), 37–43. https://doi.org/10.1038/s41591-018-0272-7

Rogers, R., Samorodnitsky, G., Wu, Z. S., & Ramdas, A. (2023). Adaptive Privacy Composition for Accuracy-first Mechanisms. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, (pp. 15833–15854). ACM.

Scheibner, J., Raisaro, J. L., Troncoso-Pastoriza, J. R., Ienca, M., Fellay, J., Vayena, E., & Hubaux, J. P. (2021). Revolutionizing medical data sharing using advanced privacy-enhancing technologies: Technical, legal, and ethical synthesis. *Journal of Medical Internet Research*, 23(2), e25120. https://doi.org/10.2196/25120

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(1), 12598. https://doi.org/10.1038/s41598-020-69250-1

Suriyakumar, V. M., Papernot, N., Goldenberg, A., & Ghassemi, M. (2021). Chasing your long tails: Differentially private prediction in health care settings. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, (pp. 723–734). ACM. https://doi.org/10.1145/3442188.3445934

Talaei, M., & Izadi, I. (2024). Adaptive Differential Privacy in Federated Learning: A Priority-Based Approach. *arXiv:2401.02453*. https://doi.org/10.48550/arXiv.2401.02453

Upreti, D., Yang, E., Kim, H., & Seo, C. (2024). A Comprehensive Survey on Federated Learning in the Healthcare Area: Concept and Applications. *Computer Modeling in Engineering and Sciences*, 140(3), 2239–2274. https://doi.org/10.32604/cmes.2024.048932

Vaidya, J., Anirban, B., Basit, S., & Hong, Y. (2013). Differentially Private Naive Bayes Classification. In *IEEE WIC ACM International Conference on Web Intelligence*. IEEE. https://doi.org/10.1109/WI-IAT.2013.80

Vayena, E., Blasimme, A., & Cohen, I. G. (2018). Machine learning in medicine: Addressing ethical challenges. *PLoS Medicine*, 15(11), e1002689. https://doi.org/10.1371/journal.pmed.1002689

Williamson, S. M., & Prybutok, V. (2024). Balancing Privacy and Progress: A Review of Privacy Challenges, Systemic Oversight, and Patient Perceptions in AI-Driven Healthcare. *Applied Sciences*, 14(2), 675. https://doi.org/10.3390/app14020675

Ziller, A., Usynin, D., Braren, R., Makowski, M., Rueckert, D., & Kaissis, G. (2021). Medical imaging deep learning with differential privacy. *Scientific Reports*, 11(1), 13524. https://doi.org/10.1038/s41598-021-93030-0

Acta Informatica Pragensia is published by the Prague University of Economics and Business, Czech Republic | eISSN: 1805-4951