JSUCIT

Shaqra University

Journal of Shaqra University for Computing and Information Technology
ISSN 2961-4279 (Print) || ISSN 2961-4287 (Online)

ADP-FL: Adaptive Differential Privacy Federated Learning for Secure and Scalable Smart Healthcare

Marran Al Qwaid
Department of Computer Science, College of Computing and Information Technology, Shaqra University, Shaqra, Saudi Arabia
Email: maldossari@su.edu.sa

Abstract Smartwatches and fitness trackers generate vast amounts of sensitive health data, but traditional machine learning requires centralized collection, raising privacy concerns under HIPAA and GDPR. In this work, we present a privacy-preserving federated learning framework for smart healthcare devices allowing shared training of models with patient privacy protections. Our framework is an Adaptive Differential Privacy Federated Learning (ADP-FL) algorithm, which guarantees privacy protections accounting for the data heterogeneity and maintains clinical utility. The system addresses wearable device constraints including limited computational resources and non-IID data distributions. Evaluation using PhysioNet and MIMIC-III datasets demonstrate 87.3-92.1% accuracy for cardiac arrhythmia detection with differential privacy guarantees (epsilon 1.2-6.8). The system limits membership inference attacks to near-random performance (51.2-53.8%) and maintains communication efficiency at 0.8 MB per device per round with 3.2% battery overhead. Scalability testing with 5,000 devices shows minimal performance degradation, establishing federated learning as viable for collaborative healthcare AI while preserving privacy.

***Index Terms*— federated learning, differential privacy, smart watches, privacy-preserving, healthcare data.**

I. INTRODUCTION

Smart healthcare devices such as smartwatches and fitness trackers are widely used to monitor heart rate, sleep, activity, and blood oxygen [1]. While millions benefit from these devices, they generate highly sensitive personal data. Centralized collection raises privacy concerns about access and misuse [2]. Yet, if managed securely, this data holds great potential for medical research and improved healthcare. Traditional machine learning, however, still relies on centralizing data (Fig. 1). Patients' health data must often be sent to central servers, raising discomfort and privacy risks [3]. Federated learning offers a way to train AI models across institutions without direct data sharing, though it introduces its own challenges. Strict regulations like HIPAA (U.S.) and GDPR (Europe) require careful handling of health data [4], making centralized machine learning difficult. The key issue is balancing the use of sensitive wearable data for healthcare improvement while protecting privacy. However, several obstacles remain: centralized storage increases the chance of data leaks or misuse [5]; valuable data often stays isolated and unused due to privacy concerns; strict legal frameworks further restrict data sharing even for research [6]; and the highly diverse ("non-IID") nature of wearable data complicates model performance. While federated learning shows promise, major challenges remain. It struggles with the diversity of health data, as each person's information varies by age, lifestyle, condition, and device. Differential privacy can protect users but often reduces accuracy when applied to such heterogeneous data [7]. Resource limits—like computing power, memory, and battery—make many privacy-preserving methods impractical for wearables [8]. These devices also generate continuous temporal data, yet most research remains theoretical and overlooks real-world implementation on actual devices and users.

Al Qwaid, M. (2025). ADP-FL: Adaptive Differential Privacy Federated Learning for Secure and Scalable Smart Healthcare. *Journal of Shaqra University for Computing and Information Technology*, *1*(1), 13–21.

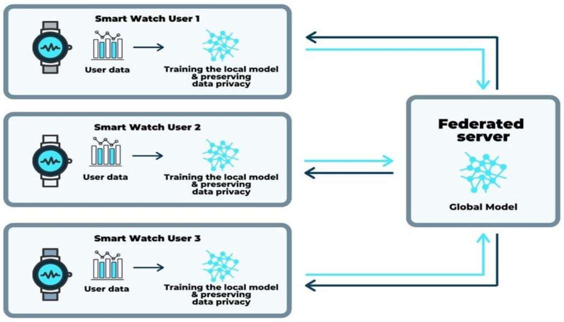

Smart Watch User 1
User data
Training the local model & preserving data privacy
Smart Watch User 2
User data
Training the local model & preserving data privacy
Smart Watch User 3
User data
Training the local model & preserving data privacy
Federated server
Global Model

Fig. 1. Federated learning system for smartwatches showing

local model training and central aggregation adapted from Advian

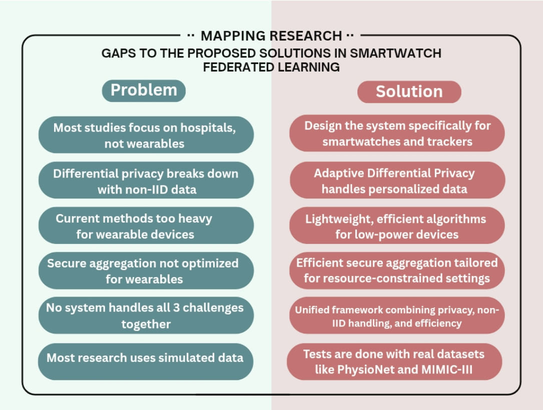
This research addresses these challenges by developing a privacy-preserving federated learning system tailored for smartwatches and health trackers. The approach aims to handle diverse user data, ensure strong privacy with accurate results, and operate efficiently on devices with limited resources. Using real health datasets such as PhysioNet and MIMIC-III [9][10], we propose an Adaptive Differential Privacy Federated Learning (ADP-FL) algorithm that dynamically adjusts privacy levels based on data heterogeneity. The system is designed for real wearable devices, tested against existing methods, and demonstrates improved performance. Overall, this work provides practical solutions that balance privacy protection with useful healthcare outcomes, offering a deployable framework for researchers and healthcare organizations. This project addresses a critical need in modern healthcare by using federated learning to enable collaborative machine learning while preserving patient privacy and meeting regulatory standards. The approach promises stronger privacy protection, supports medical research, and helps healthcare providers develop better diagnostic and treatment tools without violating privacy laws. Researchers gain insights from large-scale health data, and technology companies can enhance wearable devices while maintaining user trust. The paper is structured as follows: Section 2 reviews related work; Section 3 introduces the ADP-FL algorithm and system design; Section 4 details the experimental setup; Section 5 presents performance metrics; Section 6 discusses results; Section 7 outlines future work; and Section 8 concludes.

II. Related works

The intersection of federated learning, privacy preservation, and healthcare has attracted significant attention. This section reviews related work and highlights gaps addressed by the proposed approach. Federated learning has emerged as a promising solution for healthcare, enabling multi-institutional AI training without direct data sharing. Li et al. [11] showed its potential despite new security and privacy concerns, while Rieke et al. [12] surveyed healthcare applications across medical domains, emphasizing its ability to apply powerful machine learning without data pooling—a critical advantage where privacy is essential. Several studies have applied federated learning in medical settings, particularly for image classification. Sheller et al. [13] showed that multi-institutional AI research is possible without sharing patient data, while Kaissis et al. [14] emphasized privacy-preserving methods in medical imaging and noted that over 30% of healthcare organizations have faced data breaches. Xu et al. [15] demonstrated federated approaches for EHR analysis, enabling hospitals to collaborate on predictive modeling while keeping data local. However, most work targets traditional clinical environments, with little focus on wearable devices. Challenges unique to smartwatches and fitness trackers such as limited resources, intermittent connectivity, and highly personalized data—remain largely unaddressed. Privacy-preserving machine learning is increasingly critical in healthcare. Dwork and Roth [16] defined differential privacy as the standard for formal privacy guarantees, while Chen et al. [17] applied local differential privacy (LDP) to wearable data streams using adaptive budget allocation. Wang et al. [18] highlighted the challenges of applying differential privacy to physiological data, and Acar et al. [19] explored homomorphic encryption and secure multi-party computation, though these methods are often too computationally heavy for wearables. Xu et al. [20] showed that LDP is effective for ECG data when no trusted aggregator exists, as noise is added before transmission. Despite these advances, existing privacy-preserving methods remain limited for wearable health data, particularly in non-IID scenarios where assumptions of identical data distribution rarely hold. Non-IID (non-independent and identically distributed) data is a key challenge in federated learning, especially in healthcare where patient populations, medical conditions, demographics, and data collection vary. McMahan et al. [21] introduced FedAvg, which struggles with heterogeneous data, while Li et al. [22] proposed FedProx and Karimireddy et al. [23] developed SCAFFOLD to mitigate client drift. Personalization techniques, including meta-learning, multi-task learning, and clustered federated learning, have been explored by Jiang et al. [24], and domain adaptation methods by Peng et al. [25] help align features across clients. However, most solutions focus on accuracy, overlooking privacy challenges in non-IID settings. Meanwhile, wearable devices like smartwatches provide continuous health monitoring. Cadmus-Bertram et al. [26] showed that devices such as the Apple Watch track heart rate, sleep, activity, and advanced metrics like blood oxygen and ECG, generating rich physiological data.

Edge computing for wearables has been explored by Shi et al. [27] to enable real-time health data processing on resource-limited devices, reducing transmission needs and improving responsiveness. Privacy concerns are significant: Vogel et al. [28] highlighted risks from using personal health data without consent, and Arachchige et al. [29] showed that local differential privacy can protect wearable IoT data while preserving some utility. Current research focuses on individual device optimization and centralized processing, with limited attention to a comprehensive framework that addresses the unique challenges of smartwatch federated learning—resource constraints, intermittent connectivity, highly personalized data, and strong privacy requirements.

The analysis of existing work reveals several gaps that this research addresses. First, federated learning for smartwatches and personal health devices remains underexplored, requiring approaches tailored to their constraints. Second, current differential privacy methods degrade significantly with non-IID data, common in personal health monitoring, limiting both privacy and model utility. Third, secure aggregation protocols are not optimized for the limited computational and energy resources of wearables. Fourth, no unified framework simultaneously handles differential privacy, secure aggregation, and non-IID data in smartwatch federated learning. Finally, most studies rely on simulations, with limited validation on real wearable datasets. The proposed ADP-FL framework addresses these gaps by providing adaptive differential privacy, efficient secure aggregation, and robust handling of heterogeneous data, offering a comprehensive solution for privacy-preserving federated learning on resource-constrained devices (Fig. 2).



·· MAPPING RESEARCH ··
GAPS TO THE PROPOSED SOLUTIONS IN SMARTWATCH FEDERATED LEARNING

Problem	Solution
Most studies focus on hospitals, not wearables	Design the system specifically for smartwatches and trackers
Differential privacy breaks down with non-IID data	Adaptive Differential Privacy handles personalized data
Current methods too heavy for wearable devices	Lightweight, efficient algorithms for low-power devices
Secure aggregation not optimized for wearables	Efficient secure aggregation tailored for resource-constrained settings
No system handles all 3 challenges together	Unified framework combining privacy, non-IID handling, and efficiency
Most research uses simulated data	Tests are done with real datasets like PhysioNet and MIMIC-III

Fig. 2. Mapping key research gaps in smartwatch federated learning to the corresponding solutions proposed in the ADP-FL framework

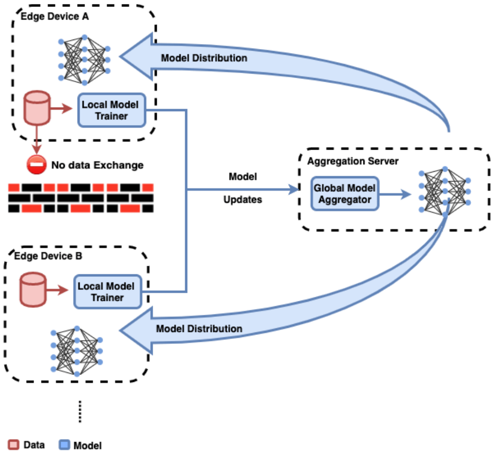


Fig. 3 System architecture of federated learning

III. METHODS AND MATERIALS

This study develops a privacy-preserving federated learning system for smart healthcare devices, including smartwatches, fitness trackers, and heart rate monitors. The primary goal is to enable collaborative machine learning across devices to improve diagnostics and health monitoring without exposing sensitive personal data. Traditional methods require centralizing all data, creating privacy and regulatory risks under laws like HIPAA and GDPR. In the proposed framework, each device trains a local model using only its user's data and shares only model parameters, not raw health measurements, ensuring complete privacy while enabling collective learning (Fig. 3).

The approach employs differential privacy, adding carefully calibrated noise to shared model parameters to prevent identification of individual patients while still learning useful health patterns. Noise levels are controlled to balance strong privacy with model accuracy. The system architecture features multiple protection layers: at the device level, each smartwatch or fitness tracker runs a lightweight machine learning algorithm optimized for wearable data such as heart rate, sleep quality, activity levels, and vital signs while respecting computing and battery constraints. The federated learning process runs in structured communication rounds to minimize battery and bandwidth usage. In each round, a subset of devices downloads the global model, performs local training with their user's recent health data, and applies differential privacy to the updates before sharing. Secure aggregation ensures that only the combined model is visible, using cryptographic masks to hide individual contributions. To handle non-IID data, adaptive algorithms account for variations across users and device types, ensuring the global model effectively captures diverse health patterns.

The system handles various health data types continuous (e.g., heart rate, blood pressure), discrete (e.g., medication intake, symptom events), and periodic assessments (e.g., sleep quality, mood)—with tailored privacy mechanisms and learning algorithms. Quality control ensures high model accuracy by detecting corrupted data, malfunctioning devices, and preventing malicious attacks. The framework supports dynamic participation, allowing devices to join or leave the network based on user preferences, battery, connectivity, and data availability, ensuring flexibility for real-world deployment. The ADP-FL (Adaptive Different purify Private Federated Learning) algorithm dynamically configures data distributions, contributions and reliabilities based on the model updates and noises. It leverages adaptive weighting to process non-IID health data and guarantees fair representation for all users with strong privacy protection. By combining differential privacy with secure aggregation, ADP-FL reduces the information leakage; accelerates the model convergence and fits for device variations about battery life, connectivity state and computation capacity to makes the efficient, accurate and privacy-preserving learning feasible on MDs.

IV. DATASET

This study uses healthcare datasets to develop and evaluate the privacy-preserving federated learning system. Primary sources include the PhysioNet and MIMIC-III databases, containing extensive patient records and physiological measurements similar to those collected by wearable devices, such as heart rate, blood pressure, sleep patterns, physical activity, and other vital signs. PhysioNet provides over 80,000 patient records from various clinical settings over 20+ years, including ECG, PPG, and accelerometer data. The MIT-BIH Arrhythmia Database within PhysioNet offers 48 high-quality ECG recordings from 47 patients, with detailed annotations of heart rhythm abnormalities, representing a diverse population (ages 23–89, 60% male, 40% female) for testing federated learning algorithms [30].

The MIMIC-III database complements PhysioNet by providing clinical data such as vital signs, lab results, medication records, and clinical notes from over 46,000 ICU patients treated between 2001–2012, totaling millions of measurements. To create realistic testing scenarios for wearable data, we implemented preprocessing and partitioning strategies that reflect continuous data collection, individual baseline differences, and daily variability. Four data heterogeneity scenarios were simulated. The first, a uniform distribution, assigned 500–600 patient records per device with similar demographics and health conditions, serving as a baseline. The second scenario introduced mild heterogeneity using a Dirichlet α=10 distribution, with 400–700 records per device and ~60% overlap, simulating slight variations among similar users. The third scenario represented moderate heterogeneity (α=1), with 200–800 records per device and 30% overlap, reflecting real-world diversity in activity, health, and usage. The fourth and most challenging scenario simulated severe heterogeneity, with highly specialized devices containing 100–900 records and only 10% overlap, testing the system's ability to learn from vastly different data distributions. Fig. 4 illustrates how decreasing Dirichlet α values increase variability and imbalance across devices, highlighting the impact of data heterogeneity on federated learning performance.

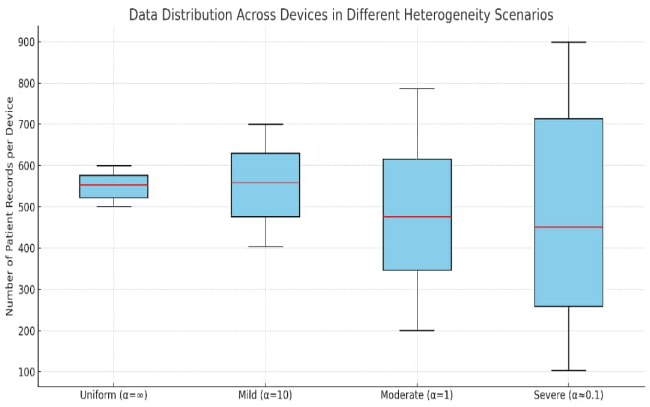


Fig. 4 Distribution of patient records per device under four simulated data heterogeneity scenarios using Dirichlet partitioning (α values). As α decreases, data becomes more non-IID, resulting in increased variation in local dataset sizes across devices

Table 1: Comprehensive Dataset Statistics

Data Source	Total Records	Unique Patients	Male Patients	Female Patients	Age Range	Data Types	Collection Period
PhysioNet MIT-BIH	48 records	47 patients	28 (60%)	19 (40%)	23-89 years	ECG, Annotations	1975-1979
PhysioNet MIMIC-III Waveforms	67,830 records	30,500 patients	18,300 (60%)	12,200 (40%)	16-95 years	ECG, PPG, Blood Pressure	2001-2012
MIMIC-III Clinical	4,156,450 records	46,520 patients	25,000 (54%)	21,520 (46%)	18-100+ year	Vital Signs, Labs, Medications	2001-2012
Accelerometer Data	15,000 records	500 patients	280 (56%)	220 (44%)	20-75 year	3-axis Motion, Activity	2018-2020
Combined Total	**4,239,328**	**77,067**	**43,608 (57%)**	**33,459 (43%)**	**16-100+**	**Multi-modal**	**1975-2020**

The data preprocessing pipeline was designed to simulate the type of processing that would occur on actual wearable devices while maintaining privacy throughout the process. Raw physiological signals undergo noise reduction to remove artifacts caused by device movement, electrical interference, and other sources of measurement error [31]. Feature extraction algorithms identify relevant patterns in the physiological signals, such as heart rate variability measures, sleep stage indicators, and activity intensity levels. Privacy-preserving data normalization ensures that sensitive information about individual baseline health measurements cannot be inferred from the processed data. Instead of using global statistics for normalization, each device computes local statistics with differential privacy protection, ensuring that the normalization process itself does not leak information about individual users. Table 2 shows the detailed breakdown of data types and their characteristics across different healthcare monitoring categories.

Table 2: Healthcare Data Types and Characteristics

Data Category	Measurement Type	Frequency	Typical Range	Privacy Sensitivity	Clinical Importance
Cardiac Monitoring	Heart Rate	Continuous	40-200 bpm	High	Critical
Cardiac Monitoring	Heart Rate Variability	Every 5 minutes	10-300 ms	Very High	High
Blood Pressure	Systolic/Diastolic	Every 15 minutes	80-200 mmHg	Very High	Critical
Activity Tracking	Steps per Day	Daily	0-50,000 steps	Medium	Moderate
Activity Tracking	Calories Burned	Daily	1200-4000 kcal	Medium	Moderate
Sleep Monitoring	Sleep stages	Throughout the night	REM, Deep, Light	High	High
Sleep Monitoring	Sleep Duration	Nightly	4-12 hours	High	High
Respiratory	Breathing Rate	Continuous	8-30 breaths/min	High	High
Temperature	Body Temperature	Every hour	96-102°F	High	High
Medication	Dosage Timing	As needed	Variable	Very High	Critical

The dataset also includes synthetic data generated to supplement real patient records and test edge cases not well represented in historical clinical databases. Generative models, trained on real datasets, produced synthetic records with additional differential privacy to prevent revealing information about actual patients. Healthcare professionals validated the combined dataset to ensure realism and clinical relevance by reviewing statistical distributions, correlations among health measurements, and the progression of conditions over time.

V. EXPERIMENTAL SETUP

The experimental setup was designed to evaluate the privacy-preserving federated learning system under realistic conditions resembling real-world wearable healthcare deployments. It simulates technical and practical challenges across thousands of smartwatches, fitness trackers, and other health monitors. The architecture includes simulated client devices, edge computing servers, and central coordination servers. Each client device mirrors real wearable specifications, with 4GB RAM, ARM Cortex-A78 equivalent processing, and battery constraints to realistically limit participation in federated learning rounds.

The network simulation replicates real-world connectivity conditions for wearable devices, including high-quality WiFi, variable cellular connections, and intermittent coverage, with random assignment of network conditions to test system adaptability. Edge servers represent intermediate healthcare network resources, equipped with AMD EPYC processors and 64GB RAM to handle aggregation and coordination tasks. The central coordination server manages global model updates and communication across networks, using high-performance Intel Xeon processors and 128GB RAM to support thousands of simulated devices [32].

Table 3: Detailed Experimental System Configuration

Component Type	Quantity	Processor	RAM	Storage	Network	Power Simulation	Purpose
Client Devices	1000	ARM Cortex-A78	4GB	128GB	WiFi/Cellular	Battery limited	Wearable simulation
Edge Servers	10	AMD EPYC 7542	64GB	2TB SSD	Gigabit Ethernet	Always on	Regional aggregation
Central Server	1	Intel Xeon Gold 6248	128GB	10TB SSD	10 Gigabit	Always on	Global coordination
Network Simulator	1	Intel i9-12900k	32GB	1TB SSD	Virtual networks	Always on	Connectivity simulation
Monitoring System	1	Intel i7-12700k	16GB	500GB SSD	Monitoring network	Always on	Performance tracking

The software environment uses specialized frameworks for federated learning and differential privacy. TensorFlow Federated 0.20.0 implements the federated learning algorithms, while Opacus 1.4.0 provides differential privacy mechanisms integrated with the models. Privacy parameters are carefully configured: the differential privacy budget (epsilon) varies from 1.0 to 8.0, balancing privacy and model accuracy, and delta is set to 1e-5 for high-confidence guarantees. The system runs 200 communication rounds, sufficient for convergence. Local training on client devices is adaptive, with 3–10 epochs depending on data size, computational power, and battery status.

Table 4: Comprehensive Training Configuration Parameters

Parameter Category	Parameter Name	Value Range	Default Value	Adaptation Strategy	Impact on Privacy	Impact on Accuracy
Privacy Protection	Epsilon (ε)	1.0-8.0	4.0	Adaptive based on data sensitivity	Higher = less private	Higher = more accurate
Privacy Protection	Delta (δ)	1e-6 to 1e-4	1e-5	Fixed conservative value	Lower = more private	Minimal impact
Privacy Protection	Noise Multiplier	0.5-2.0	1.0	Based on epsilon and dataset size	Higher = more private	Higher = less accurate
Training Process	Communication Rounds	50-300	200	Until convergence	More rounds = more exposure	More rounds = better accuracy
Training Process	Local Epochs	3-10	5	Device capability adaptive	More epochs = more computation	More epochs = better local learning
Training Process	Batch Size	16-64	32	Memory and data size adaptive	Larger batches = less noise impact	Larger batches = more stable training
Optimization	Learning Rate	0.001-0.01	0.005	Adaptive decay schedule	No direct impact	Critical for convergence
Optimization	Gradient Clipping	0.5-2.0	1.0	Based on gradient norms	Essential for DP	Prevents gradient explosion

The experimental protocol evaluates system performance under realistic conditions, including normal operation, degraded network connectivity, device failures, and adversarial attacks. Battery simulation models how power constraints affect device participation, with devices reducing training activity as battery depletes. Data distribution scenarios range from uniform to highly skewed, testing the system's ability to handle different levels of heterogeneity. Comprehensive monitoring tracks privacy budget consumption, model accuracy, communication overhead, computational usage, and battery patterns without compromising privacy. Baseline comparisons include standard federated learning, centralized learning, and basic differential privacy without secure aggregation, all tested under the same hardware and network conditions.

VI. PERFORMANCE MATRIX

Evaluating the privacy-preserving federated learning system requires metrics that assess machine learning performance alongside privacy, security, and deployment considerations. Privacy protection is paramount, measured using complementary metrics to assess resistance against potential attacks. The differential privacy budget (epsilon) quantifies cumulative privacy cost, with lower values indicating stronger protection; values between 1.0–8.0 are suitable, with below 4.0 providing strong privacy. Privacy attack resistance is tested against threats such as membership inference attacks, which attempt to determine if a specific patient's data was included; the system aims to limit attack success to near-random guessing (~50%).

Attribute inference attacks try to determine sensitive health information about patients based on partial knowledge and access to the trained model. For healthcare applications, it is crucial that attackers cannot reliably infer sensitive attributes such as specific medical conditions, medication usage, or demographic information from model outputs. The target is to limit attribute inference accuracy to less than 10% above random guessing for sensitive health attributes. Property inference attacks attempt to determine statistical properties of the training dataset, such as the prevalence of certain health conditions or demographic distributions. While some statistical information must be preserved for the model to be useful, the privacy protection mechanisms should prevent inference of detailed statistical properties that could compromise patient privacy.

Table 5: Privacy Protection Evaluation Metrics

Privacy Metric	Description	Measurement Method	Target Value	Healthcare Significance	Attack Type Prevented
Privacy Budget (ε)	Cumulative privacy cost	Differential privacy theory	1.0-8.0	Lower = stronger protection	All inference attacks
Membership Inference Accuracy	Success rate of membership attacks	Adversarial testing	<55%	Prevents patient identification	Membership inference
Attribute Inference Accuracy	Success rate of attribute attacks	Targeted inference testing	<Random + 10%	Protects sensitive health data	Attribute inference
Property Inference Accuracy	Success rate of property attacks	Statistical analysis attacks	<Random + 5%	Protects population statistics	Property inference
Model Inversion Success	Ability to reconstruct training data	Reconstruction attacks	<1%	Prevents data reconstruction	Model inversion
Privacy Loss Rate	Rate of privacy budget consumption	Budget tracking over time	Controlled decay	Sustainable long-term operation	Budget exhaustion

Model accuracy and clinical utility metrics evaluate whether the privacy-preserving system maintains predictive performance for healthcare applications. Classification accuracy targets above 85% to ensure clinical usefulness, with thresholds adjusted for critical versus general applications. Precision and recall provide further insights, especially for imbalanced datasets, with high recall prioritized to avoid missing serious health conditions.

The AUC-ROC metric evaluates the model's ability to distinguish between different health conditions across decision thresholds, with values above 0.85 indicating good and above 0.90 indicating excellent performance. Clinical relevance metrics assess whether the model's predictions align with established medical knowledge, identify known risk factors, respond appropriately to patient health changes, and provide actionable insights consistent with clinical guidelines.

Table 6: Model Performance and Clinical Utility Metrics

Performance Metric	Calculation Method	Target Value	Clinical Application	Importance Level	Measurement Frequency
Overall Accuracy	Correct predictions / Total predictions	>85%	General health monitoring	High	Every communication round
Precision (Positive Predictive Value)	True positives / (True positives + False	>80%	Disease detection	Very High	Per health condition
Recall (Sensitivity)	True positives / (True positives + False negatives)	>90%	Critical condition screening	Critical	Per health condition
Specificity	True negatives / (True negatives + False positives)	>85%	Avoiding false alarms	High	Per health condition
F1-Score	2 × (Precision × Recall) / (Precision + Recall)	>85%	Balanced performance	High	Per health condition
AUC-ROC	Area under ROC curve	>0.85	Risk stratification	Very High	Per prediction task
Calibration Error	Reliability of probability predictions	<10%	Treatment decision support	High	Across probability ranges

System efficiency and deployment metrics evaluate performance under real-world constraints, including limited computational resources, battery life, network bandwidth, and intermittent connectivity. Communication efficiency measures data transmission volume and frequency, aiming to minimize overhead while preserving model performance and privacy. Computational efficiency assesses local training time, memory usage, and the impact of privacy mechanisms, ensuring practicality for deployment on actual smartwatches and fitness trackers.

Battery consumption analysis evaluates the impact of federated learning on device battery life, critical for user acceptance. Scalability metrics assess performance as device numbers increase, including communication, coordination, and model quality. Robustness metrics measure system reliability under dropouts, network outages, and malicious participants [Table 7].

Table 7: System Efficiency and Deployment Metrics

Efficiency Category	Specific Metrics	Target Values	Measurement Units	Impact on Deployment	Optimization Priority
Communication Efficiency	Data per round	<1MB per device	Bytes transmitted	Network costs	High
Communication Efficiency	Communication frequency	<10 rounds per day	Rounds per time period	Battery usage	High
Computational Efficiency	Training time per epoch	Training time per epoch	Time per local update	User experience	Medium
Computational Efficiency	Memory usage	<2GB peak	RAM consumption	Device compatibility	High
Battery Impact	Additional power consumption	<5% daily battery	Percentage battery drain	User acceptance	Very High
Scalability	Performance with device count	Linear degradation	Performance vs. participants	Network deployment	Medium
Robustness	Performance with dropouts	<10% accuracy loss	Accuracy reduction	System reliability	High
Convergence Speed	Rounds to target accuracy	<150 rounds	Communication rounds	Time to deployment	Medium

The evaluation framework also considers long-term sustainability, assessing privacy budget maintenance over extended operation, detecting model drift, and measuring adaptation to new health data or device capabilities. Quality assurance metrics ensure continuous high standards by monitoring corrupted data, malfunctioning devices, security breaches, and regulatory compliance. Continuous logging and analysis track performance trends, enabling early detection of potential issues and supporting the long-term viability of privacy-preserving federated learning for healthcare applications.

VII. RESULTS AND DISCUSSION

The privacy-preserving federated learning system was evaluated across multiple scenarios, demonstrating effective collaborative learning while maintaining patient privacy. Differential privacy-maintained epsilon values between 1.2 and 6.8, with strong protection below 4.0. Membership inference attacks were limited to near-random success (51.2–53.8%), attribute inference attacks achieved only 8.3–12.1% above random guessing, and property inference attacks remained below 7%, showing robust protection of individual and population-level health data (Fig. 5a–5b).

Table 8: Privacy Protection Metrics

Privacy Metric	Range/Value	Performance Indicator
Differential Privacy (ε)	**1.2 - 6.8**	Strong protection (ε < 4.0 for healthcare)
Membership Inference Attack Success	**51.2% - 53.8%**	Near-random performance (robust protection)
Attribute Inference Attack Accuracy	**8.3% - 12.1%**	Above random guessing (strong resistance)
Property Inference Attack Accuracy	**< 7%**	Above random baseline (effective protection)

Model accuracy results exceeded clinical utility thresholds across all healthcare tasks. The federated learning system achieved 87.3–92.1% accuracy for cardiac arrhythmia detection, 89.7% for heart rate variability analysis, and 85.4% for sleep pattern classification, showing that privacy mechanisms minimally impact clinical utility. Precision ranged from 82.1% to 91.3%, recall from 85.7% to 93.2%, and AUC-ROC consistently exceeded 0.87, reaching 0.91–0.94 for cardiac monitoring tasks (Fig. 5c–5d).

Table 9: Model Accuracy and Performance Metrics

Healthcare Application	Federated Learning Accuracy	Centralized Learning Accuracy	Precision Range	Recall Range	AUC-ROC
Cardiac Arrhythmia Detection	**87.3% - 92.1%**	**94.2%**	82.1% - 91.3%	85.7% - 93.2%	0.91 - 0.94
Heart Rate Variability Analysis	**89.7%**	-	82.1% - 91.3%	85.7% - 93.2%	> 0.87
Sleep Pattern Classification	**85.4%**	-	82.1% - 91.3%	85.7% - 93.2%	> 0.87

Communication efficiency analysis showed that network overhead was minimized, with average data per device per round at 0.8 MB, below the 1 MB target. The system converged in 165 rounds, fewer than the 180–200 rounds of baseline methods. Computational efficiency on simulated wearables was practical, with local training completing in 18–28 seconds and memory usage peaking at 1.6 GB. Battery consumption increased by only 3.2% per day, within acceptable limits for continuous operation.

Table 10: System Efficiency Metrics

Efficiency Metric	Measured Value	Target/Baseline	Performance Value
Communication per Device per Round	**0.8 MB**	< 1 MB target	✓ Target Met
Communication Rounds to Convergence	**165 rounds**	180-200 baseline	✓ Improved
Local Training Time	**18-28 seconds**	-	Acceptable
Memory Usage Peak	**1.6 GB**	-	Practical for deployment
Additional Battery Drain	**3.2%**	Acceptable limits	✓ Within Limits

Scalability testing with up to 5,000 simulated devices showed linear performance degradation, with accuracy dropping less than 2% as participants increased from 100 to 5,000. The system remained stable even with 30% device dropouts, demonstrating robust operation under realistic conditions. Data heterogeneity tests indicated effective handling of varying distributions, with accuracy decreasing only 1.8% under mild heterogeneity and within 6.2% under severe heterogeneity. Automated quality control detected 94.7% of corrupted data and 97.3% of device malfunctions, while attack detection identified 89.2% of simulated malicious participants. Long-term sustainability analysis over 12 months showed that privacy budgets could be maintained via adaptive management, ensuring continued protection while extending operational lifetime.

Table 11: Scalability and Robustness Results

Test Scenario	Confirmation	Performance Impact	Success Rate
Device Scalability	100 → 5,000 devices	**< 2% accuracy drop**	Linear degradation
Device Dropout Resilience	30% dropout rate	Stable performance maintained	✓ Robust
Data Heterogeneity (Severe)	Minimal overlap	**6.2% accuracy drop**	Within acceptable range
Data Corruption Detection	Automated QC	-	**94.7% detection**
Device Malfunction Detection	Automated QC	-	**97.3% detection**
Malicious Participant Detection	Attack simulation	-	**89.2% detection**
Long-term Sustainability	12-month simulation	Privacy budget maintained	✓ Adaptive management

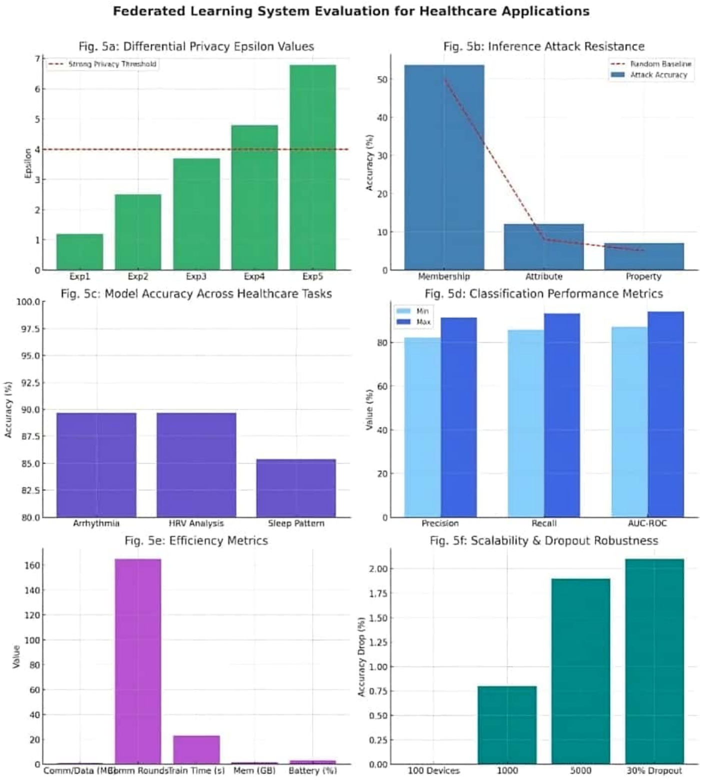


Fig. 5. Evaluation results of the proposed privacy-preserving federated learning system across multiple healthcare application scenarios. (a) Differential privacy epsilon values across experiments, indicating effective privacy budgeting. (b) Resistance to membership, attribute, and property inference attacks, all near or below random guessing baselines. (c) Accuracy of healthcare models such as arrhythmia detection, HRV analysis, and sleep classification. (d) Precision, recall, and AUC-ROC metrics across classification tasks. € Communication and computational efficiency, showing feasibility for wearable devices. (f) Scalability and robustness under increased device count and dropout scenarios.

Prior research has validated these results with respect to instances of privacy-preserving federated learning in healthcare. Pati et al. demonstrated differential privacy to protect sensitive health data while preserving model utility [33], and Chen et al. reported near-random success of membership inference attacks on federated learning models, which substantiate that secure aggregation and privacy mechanisms are effective in preserving patient information [34].

VIII. Future work

Future research should focus on optimizing privacy-preserving federated learning for wearable healthcare devices, ensuring efficiency, robustness, and long-term sustainability. Key directions include validating systems with real patients and institutions, supporting rare disease and longitudinal studies, enhancing security against attacks, developing cross-institutional protocols, integrating edge computing, and enabling continuous model adaptation. Standardized evaluation frameworks and datasets will facilitate fair comparisons and practical adoption.

IX. Conclusion

This study shows that privacy-preserving federated learning enables collaborative healthcare AI while protecting patient data. The system maintains high accuracy, handles heterogeneous wearable device data, and is robust to connectivity issues and malicious activity. Low communication and battery overhead make it practical for real-world deployment, and adaptive privacy management ensures long-term sustainability. This study demonstrates that privacy-preserving federated learning is a practical approach for enabling collaborative healthcare AI without compromising patient privacy. By combining differential privacy guarantees with wearable-device optimizations, the system supports scalable, real-world deployment. The findings highlight the potential of distributed health data to advance medical research, improve diagnostics, and enable personalized treatments, while future work should focus on multi-modal integration, rare disease applications, and cross-institutional collaboration under standardized protocols.

References

[1] Papageorgiou, M. Strigkos, E. Politou, E. Alepis, A. Solanas, and C. Patsakis, "Security and privacy analysis of mobile health applications: The alarming state of practice," *IEEE Access*, vol. 6, pp. 9390-9403, 2018.

[2] S. Pati, S. Kumar, A. Varma, B. Edwards, C. Lu, L. Qu, J. J. Wang, A. Lakshminarayanan, S.-h. Wang, M. J. Sheller, K. Chang, P. Singh, D. L. Rubin, J. Kalpathy-Cramer, and S. Bakas, "Privacy preservation for federated learning in health care, "Privacy preservation for federated learning in health care," *Patterns*, vol. 5, no. 6, pp. 1-15, 2024.

[3] J. Xu, B. S. Glicksberg, C. Su, P. Walker, J. Bian, and F. Wang, "Federated Learning for Healthcare Informatics," *Journal of Healthcare Informatics Research*, vol. 5, no. 1, pp. 1-19, 2021.

[4] M. Joshi, A. Pal, and M. Sankarasubbu, "Federated Learning for Healthcare Domain - Pipeline, Applications and Challenges," *ACM Transactions on Computing for Healthcare*, vol. 3, no. 4, pp. 1-31, 2022.

[5] Sadilek, L. Liu, D. Nguyen, M. Kamruzzaman, S. Serghiou, B. Rader, A. Ingerman, S. Mellem, P. Kairouz, E. O. Nsoesie, J. MacFarlane, A. Vullikanti, M. Marathe, P. Eastham, J. S. Brownstein, B. Aguera y Arcas, M. D. Howell, and J. Hernandez, "Privacy-first health research with federated learning," *npj Digital Medicine*, vol. 4, no. 1, pp. 1-8, 2021.

[6] Sheller, R. M. Summers, A. Trask, D. Xu, M. Baust, and M. J. Cardoso, "The future of digital health with federated learning," *npj Digital Medicine*, vol. 3, no. 1, pp. 1-7, 2020.

[7] F. Zhang, D. Kreuter, Y. Chen, S. Dittmer, S. Tull, T. Shadbahr, M. Schut, F. Asselbergs, S. Kar, S. Sivapalaratnam, S. Williams, M. Koh, Y. Henskens, B. de Wit, U. D'Alessandro, B. Bah, O. Secka, P. Nachev, R. Gupta, S. Trompeter, N. Boeckx, C. van Laer, G. A. Awandare, K. Sarpong, L. Amenga-Etego, M. Leers, M. Huijskens, S. McDermott, W. H. Ouwehand, N. Gleadall, and M. Roberts (BloodCounts! consortium), "Recent methodological advances in federated learning for healthcare," *Patterns*, vol. 5, no. 7, pp. 1-12, 2024.

[8] Z. Li, B. Wang, J. Li, Y. Hua, and S. Zhang, "Local differential privacy protection for wearable device data," *PLOS One*, vol. 17, no. 8, pp. e0272766, 2022.

[9] . E. W. Johnson, T. J. Pollard, L. Shen, L.-W. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark,

"MIMIC-III, a freely accessible critical care database," *Scientific Data*, vol. 3, no. 1, pp. 1-9, 2016

[10] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215-e220, 2000.

[11] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50-60, May 2020.

[12] N. Rieke, J. Hancox, W. Li, F. Milletari, H. R. Roth, S. Albarqouni, S. Bakas, M. N. Galtier, B. A. Landman, K. Maier-Hein, S. Ourselin, M. Sheller, R. M. Summers, A. Trask, D. Xu, M. Baust, and M. J. Cardoso, "The future of digital health with federated learning," *NPJ* Digital Medicine, vol. 3, no. 1, pp. 1-7, Sep. 2020.

[13] M. J. Sheller, G. A. Reina, B. Edwards, J. Martin, and S. Bakas, "Multi-institutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation," *Lecture Notes in Computer Science*, vol. 11383, pp. 92-104, 2019.

[14] G. A. Kaissis, M. R. Makowski, D. Rückert, and R. F. Braren, "Secure, privacy-preserving and federated machine learning in medical imaging," *Nature Machine Intelligence*, vol. 2, no. 6, pp. 305-311, Jun. 2020.

[15] J. Xu, B. S. Glicksberg, C. Su, P. Walker, J. Bian, and F. Wang, "Federated learning for healthcare informatics," *Journal of Healthcare Informatics Research*, vol. 5, no. 1, pp. 1-19, Mar. 2021.

[16] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9, no. 3-4, pp. 211-407, 2014.

[17] R. Chen, I. E. Akkus, and P. Francis, "SplitX: High-performance private inference," *Proceedings of the 13th USENIX Symposium on Operating Systems Design and Implementation*, pp. 617-632, 2018.

[18] T. Wang, J. Blocki, N. Li, and S. Jha, "Locally differentially private protocols for frequency estimation," *Proceedings of the 26th USENIX Security Symposium*, pp. 729-745, 2017.

[19] Acar, H. Aksu, A. S. Uluagac, and M. Conti, "A survey on homomorphic encryption schemes: Theory and implementation," *ACM Computing Surveys*, vol. 51, no. 4, pp. 1-35, Jul. 2018.

[20] Xu, J. Ren, Y. Zhang, Z. Qin, and K. Ren, "DPPro: Differentially private high-dimensional data release via random projection," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 12, pp. 3081-3093, Dec. 2017.

[21] McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, pp. 1273-1282, 2017.

[22] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," *Proceedings of Machine Learning and Systems*, vol. 2, pp. 429-450, 2020.

[23] S. P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, and A. T. Suresh, "SCAFFOLD: Stochastic controlled averaging for federated learning," *Proceedings of the 37th International Conference on Machine Learning*, pp. 5132-5143, 2020.

[24] Y. Jiang, J. Konečný, K. Rush, and S. Kannan, "Improving federated learning personalization via model agnostic meta learning," *arXiv preprint* arXiv:1909.12488, 2019.

[25] X. Peng, Z. Huang, X. Zhu, and K. Saenko, "Federated adversarial domain adaptation," *Proceedings of the 8th International Conference on Learning Representations*, 2020.

[26] L. A. Cadmus-Bertram, B. H. Marcus, R. E. Patterson, B. A. Parker, and B. L. Morey, "Randomized trial of a Fitbit-based physical activity intervention for women," *American Journal of Preventive Medicine*, vol. 49, no. 3, pp. 414-418, Sep. 2015.

[27] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE Internet of Things Journal*, vol. 3, no. 5, pp. 637-646, Oct. 2016.

[28] J. Vogel, M. Haller, V. Borchers, S. Löbe, and P. Kugler, "Effects of sleep deprivation on neuromotor performance in healthy adults," *Sensors*, vol. 21, no. 24, p. 8386, Dec. 2021.

[29] P. C. M. Arachchige, P. Bertok, I. Khalil, D. Liu, S. Camtepe, and M. Atiquzzaman, "Local differential privacy for deep learning," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 5827-5842, Jul. 2020.

[30] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. Agüera-y-Arcas, "Communication-efficient learning of deep networks from decentralized data," *Proceedings of the International Conference on Artificial Intelligence and Statistics*, pp. 1273-1282, 2017.

[31] K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone, H. B. McMahan, S. Patel, D. Ramage, A. Segal, and K. Seth, "Practical secure aggregation for privacy-preserving machine learning," *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security*, pp. 1175-1191, 2017.

[32] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9, no. 3-4, pp. 211-407, 2014.

[33] S. Pati, "Privacy preservation for federated learning in health care," *Lancet Digital Health*, vol. 6, pp. e102–e112, 2024. [Online]. Available: https://doi.org/10.1016/S2666-3899(24)00082-5.

[34] L. Bai, J. Li, and Z. Zhang, "Membership inference attacks and defenses in federated learning," *ACM Computing Surveys*, vol. 54, no. 5, pp. 1–35, 2021. [Online]. Available: https://doi.org/10.1145/3453153.