Fed-HealthGen: A Generative Federated Framework for Privacy-Preserving Personalized Healthcare Using Wearable Consumer Electronics

Jianhui Lv, *Senior Member, IEEE*, Byung-Gyu Kim, *Senior Member, IEEE*, and Keqin Li, *Fellow, IEEE*

***Abstract*— Wearable healthcare consumer electronics generate substantial medical time-series data with significant potential for personalized healthcare applications. However, effectively modeling such data in federated settings presents unique challenges due to pronounced Non-Independent and Identically Distributed (Non-IID) characteristics, privacy concerns, and personalization requirements. This paper proposes Med-FedLSG, a novel federated learning framework that addresses these challenges through three key innovations: a variational temporal representation learning mechanism with explicit disentanglement of shared and personalized features, a conditional temporal generator with physiological constraints, and a two-level federated optimization framework based on knowledge distillation. Extensive experiments on the MIMIC-III clinical database and UCI-HAR dataset demonstrate that Med-FedLSG consistently outperforms existing federated learning methods, achieving 85.47% accuracy on MIMIC-III and 92.28% on UCI-HAR. Furthermore, our framework achieves superior intra-user consistency scores of 0.89 on MIMIC-III and 0.93 on UCI-HAR, demonstrating enhanced stability for long-term medical monitoring. Ablation studies validate the effectiveness of our personalized representation separation mechanism and conditional temporal generator. The proposed approach successfully balances generalization and personalization while maintaining privacy, offering promising solutions for smart wearable health applications.**

***Index Terms*— Federated learning, wearable healthcare, consumer electronics, generative AI, personalized medicine.**

I. Introduction

WEARABLE healthcare consumer electronics have been exploited widely and rapidly in recent years [1], evolving from simple step counters to sophisticated multi-sensor systems capable of continuous monitoring of vital signs, activity patterns, and physiological parameters. These devices generate vast amounts of medical time-series data that hold immense potential for personalized healthcare applications, including early disease detection, treatment optimization, and wellness management [2]. This technological evolution creates unprecedented opportunities to transform healthcare delivery through continuous, personalized monitoring outside traditional clinical settings.

The effective utilization of wearable consumer electronics data faces significant privacy challenges [3], particularly as regulations such as General Data Protection Regulation (GDPR) and Health Insurance Portability and Accountability Act (HIPAA) impose strict requirements on medical data handling [4]. Secure data transmission and identity authentication become critical components in such sensitive healthcare environments [5], [6]. Conventional machine learning approaches typically centralize data collection and processing, creating privacy vulnerabilities and limiting adoption in sensitive healthcare contexts. Federated learning has emerged as a promising paradigm that enables collaborative model training while keeping data decentralized [7]. However, physiological data shows significant heterogeneity across individuals. This variation stems from differences in individual physiology, device types, and usage patterns [8]. Traditional federated learning approaches face substantial challenges with such diverse data. These approaches typically assume data homogeneity across all participating clients [9].

Recent advances in generative artificial intelligence present novel opportunities to address these challenges by enabling synthetic data generation, knowledge transfer, and personalized modeling [10]. By combining generative models with federated learning frameworks, it becomes possible to share knowledge across users without exchanging raw data, while still accounting for individual differences. This approach meets the growing demand for personalized healthcare solutions. Such solutions require tailored interventions based on individual physiological characteristics. They must also account for personal preferences and specific health objectives. Personalization enables more effective health monitoring and management for individual users. Moreover, generative models can help mitigate data imbalance and scarcity issues commonly encountered in medical applications, potentially improving model robustness and clinical utility [11].

Despite these promising advances, several critical challenges remain unresolved in the integration of generative models with federated learning for wearable healthcare consumer electronics. Medical time-series data from wearable devices

Received 21 September 2025; accepted 19 October 2025. Date of publication 23 October 2025; date of current version 8 December 2025. This work was supported by the National Natural Science Foundation of China under Grant 62202247. *(Corresponding author: Jianhui Lv.)*

Jianhui Lv is with the Multi-Modal Data Fusion and Precision Medicine Laboratory, The First Affiliated Hospital of Jinzhou Medical University, Jinzhou 121000, China (e-mail: lvjianhui2012@163.com).

Byung-Gyu Kim is with the Division of Artificial Intelligence Engineering, Sookmyung Women's University, Seoul 04310, Republic of Korea (e-mail: bg.kim@sookmyung.ac.kr).

Keqin Li is with the College of Computer Science, State University of New York, New Paltz, NY 12561 USA (e-mail: lik@newpaltz.edu).

Digital Object Identifier 10.1109/TCE.2025.3624693

presents unique challenges for machine learning models. These data exhibit significant Non-IID (Non-Independent and Identically Distributed) characteristics due to inherent physiological differences among individuals, variations in health conditions, and heterogeneity in device types and wearing locations [12]. Additionally, medical data are highly sensitive, requiring strict privacy protection while still enabling effective knowledge sharing across users [13]. As noted by Wang et al. [14], traditional centralized learning approaches fail to address these privacy concerns, while existing federated learning methods struggle with the personalization required for accurate health monitoring. The fundamental challenge lies in developing a framework that can simultaneously preserve privacy, handle data heterogeneity, and provide personalized health insights from wearable device data. Specifically, existing approaches present significant limitations at three key intersections: (1) the integration of temporal modeling with federated learning for personalized medical applications, where current methods have limited capability to capture complex physiological dynamics while maintaining user privacy; (2) the combination of generative models with privacy preservation for heterogeneous medical data, as existing generative federated approaches demonstrate insufficient effectiveness in handling the unique characteristics of medical time-series; and (3) the balance between knowledge sharing and personal feature protection, where current frameworks lack systematic mechanisms to distinguish and disentangle shared medical knowledge from individual physiological characteristics without compromising privacy.

To address these identified gaps, this research aims to achieve three specific objectives: (1) Design a privacy-preserving federated learning framework for heterogeneous medical time-series data from wearable devices; (2) Develop an explicit mechanism to disentangle shared medical knowledge from personalized physiological characteristics; and (3) Create a conditional generative approach that produces high-quality synthetic medical time-series data for knowledge transfer without raw data exchange. The main contributions of this work are threefold:

1) We propose a variational temporal representation learning mechanism that effectively captures the uncertainty and complexity inherent in medical time-series data in wearable healthcare consumer electronics, while explicitly disentangling shared and personalized features.

2) We develop a conditional temporal generator with physiological constraints that facilitates knowledge transfer without exchanging raw data, thereby preserving privacy while enabling effective learning from heterogeneous medical data sources.

3) We conduct extensive experiments on two significantly heterogeneous datasets, MIMIC-III and UCI-HAR, demonstrating that Med-FedLSG outperforms state-of-the-art federated learning methods.

The remainder of this paper is organized as follows. Section II reviews related work. Section III describes the system architecture and problem formulation. Section IV details the Med-FedLSG method. Section V presents experimental evaluations on two heterogeneous datasets. Section VI concludes with contributions and future directions.

II. Related Work

This section reviews the existing literature across three main areas: traditional deep learning approaches for medical time-series data, federated learning methods for healthcare applications, and generative models for medical data synthesis. We analyze the strengths and limitations of current approaches to highlight the research gaps our work addresses.

A. Traditional Deep Learning for Medical Time-Series

Traditional deep learning approaches for medical time-series data predominantly rely on centralized architectures, requiring data collection at a central server. Hannun et al. [15] demonstrated the effectiveness of deep neural networks for arrhythmia detection from Electrocardiogram (ECG) signals, while Choi et al. [16] utilized recurrent neural networks for disease progression modeling. However, these methods fail to address privacy concerns and require extensive labeled data from each individual, which is often impractical in medical settings. Moreover, as highlighted by Jiang et al. [17], these models typically struggle with the high inter-subject variability inherent in physiological data. Recent advances in large model-driven approaches have shown promise in handling complex multi-sensor healthcare data fusion [8], though these typically require centralized architectures that conflict with privacy requirements. These limitations highlight the need for privacy-preserving approaches that can handle data heterogeneity across individuals.

B. Federated Learning for Healthcare

To address privacy concerns while maintaining collaborative learning capabilities, federated learning has emerged as a promising approach to protect data privacy while enabling collaborative model training. McMahan et al. [18] introduced FedAvg, which allows model training without sharing raw data. Reference [19] demonstrated the potential of federated learning to address data silos in healthcare systems, showing that FL can improve equity and compliance while maintaining privacy, achieving 91.3% accuracy under strict privacy budgets in medical image analysis. Balancing privacy protection and model performance in federated learning environments presents a significant challenge. Wang et al. [20] introduced optimization strategies including dynamic privacy budget allocation and adaptive gradient clipping to improve model convergence and data quality under limited privacy budgets. To enhance decentralization, security, and fairness of federated learning framework, Wang et al. [21] proposed a blockchain-based federated learning model with differential privacy mechanisms. However, federated learning systems remain vulnerable to security threats, as highlighted by Yazdinejad et al. [22], who developed robust privacy-preserving approaches against model poisoning attacks using encrypted gradient evaluation and byzantine-tolerant aggregation. Several variants have been proposed for medical applications, such as FedHealth by Chen et al. [23] for personalized healthcare and FedPer by Arivazhagan et al. [24] for personalized model adaptation.

However, these methods face significant challenges in handling heterogeneous medical data. Li et al. [25] showed that traditional federated learning methods suffer from convergence issues and performance degradation under Non-IID settings. Additionally, as observed by Kulkarni et al. [26], these approaches struggle to effectively transfer knowledge across users with diverse physiological characteristics while maintaining personalization. The integration of FL with consumer IoT devices presents additional challenges, as demonstrated by Namakshenas et al. [27], who proposed quantum-based authentication and homomorphic encryption approaches to address privacy and security concerns in consumer device networks. While these federated approaches address privacy concerns, they still lack effective mechanisms for handling the temporal dynamics and personalization requirements specific to medical time-series data.

C. Generative Models for Medical Data

Recent advances in generative models have shown promise in addressing data heterogeneity. Yoon et al. [28] employed time-series generative adversarial networks for medical data synthesis, while Zhu et al. [29] explored variational approaches for physiological data generation. However, these methods typically operate in centralized settings, compromising privacy. Attempts to integrate generative models with federated learning, such as FedGAN by Rasouli et al. [30], have shown potential but remain limited in their ability to capture the temporal dynamics of medical data and personalize to individual users. Furthermore, as pointed out by Chen et al. [31], existing approaches lack effective mechanisms to distinguish between general medical knowledge that should be shared and personal characteristics that should remain private. These limitations motivate the need for a novel approach that combines the advantages of generative models with federated learning while addressing the specific challenges of medical time-series data.

III. System Model and Problem Formulation

This section describes the system architecture and formulates the mathematical problem underpinning the Med-FedLSG framework, laying the theoretical foundation for subsequent methodological developments.

A. System Architecture

We consider a federated learning system consisting of a central server and N users, as illustrated in Fig. 1. Each user is equipped with various wearable health consumer electronics, such as electrocardiogram monitors, continuous glucose monitors, or sleep trackers, producing medical time-series data. The central server coordinates the learning process without directly accessing users' raw data, thus preserving privacy. The central server maintains a conditional temporal generator that produces high-quality synthetic medical samples, enabling knowledge transfer across users without compromising individual privacy.

The system operates in communication rounds. In each round t, the server selects a subset $\mathcal{A}_t$ from the user set

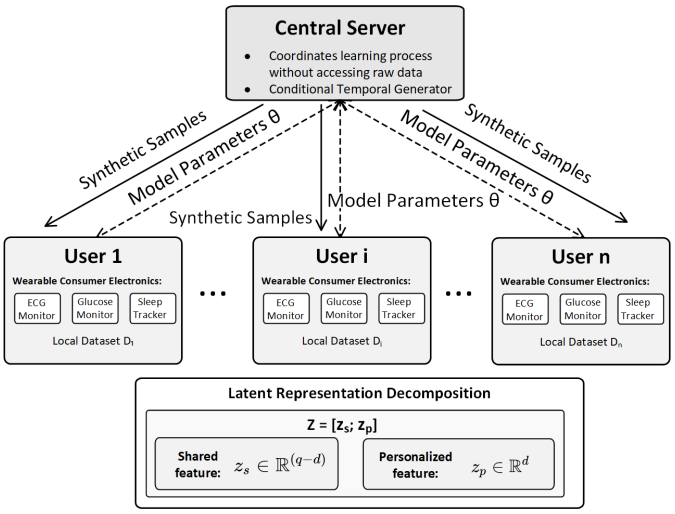


Fig. 1. Architecture of the Med-FedLSG system.

$\{1, 2, \ldots, N\}$ to participate in training. During each round, the server not only aggregates model parameters but also generates synthetic samples using the conditional temporal generator, which are then distributed to selected users to augment their local training data and facilitate cross-user knowledge sharing.

Each user u maintains a local dataset $\mathcal{D}_u = \{(\mathbf{X}_i^T, y_i)\}_{i=1}^{n_u}$, where $\mathbf{X}_i^T \in \mathbb{R}^{p\times\tau}$ represents a physiological time-series consisting of p features over τ time steps, and $y_i \in \mathcal{Y}$ denotes the corresponding health status or risk assessment. These datasets exhibit significant Non-IID characteristics due to (1) physiological differences among individuals; (2) variation in health conditions and diseases; and (3) heterogeneity in device types, wearing locations, and user habits.

B. Problem Formulation

In this subsection, we mathematically formalize the federated learning problem for medical time-series data within the Med-FedLSG framework.

1) Definition of Spaces: We first define the critical spaces of the system:

- $\mathcal{X}^T \subseteq \mathbb{R}^{p\times\tau}$: Space of medical time-series data with p features and τ time steps.
- $\mathcal{Z} \subseteq \mathbb{R}^q$: Latent feature space with dimension q.
- $\mathcal{Y} \subseteq \mathbb{R}$: Output space, representing health status or risk scores.

A key innovation in Med-FedLSG is decomposing the latent space $\mathcal{Z}$ into two complementary subspaces:

- $\mathcal{Z}_s \subseteq \mathbb{R}^{q-d}$: Shared health feature space, capturing common medical knowledge across users.
- $\mathcal{Z}_p \subseteq \mathbb{R}^d$: Personalized feature space, capturing user-specific characteristics.

Thus, the latent representation is expressed as $\mathbf{z} = [\mathbf{z}_s; \mathbf{z}_p]$, where $\mathbf{z}_s \in \mathcal{Z}_s$ and $\mathbf{z}_p \in \mathcal{Z}_p$.

2) Model Parameterization: The Med-FedLSG model parameters are denoted by $\theta := [\theta^f; \theta^c; \theta^p]$, consisting of:

- A temporal encoder $f : \mathcal{X}^T \to \mathcal{Z}$ parameterized by θ^f, mapping time-series data to the latent space.

- A generator $c : \mathcal{Z} \to \mathcal{W}$ parameterized by θ^c, processing latent representations to generate intermediate features.
- A predictor $k : \mathcal{W} \to \mathcal{Y}$ parameterized by θ^p, producing health assessments.

Following the principles of disentangled representation learning, the temporal encoder parameters θ^f are further divided into two parts, $\theta^f = [\theta^f_s; \theta^f_p]$, corresponding to shared and personalized feature extraction, respectively.

3) Task Definition: For each user u, our goal is to find optimal parameters θ_u minimizing the expected loss under the user's data distribution:

$$\min_{\theta_u} \mathbb{E}_{(\mathbf{X}^T, y)\sim\mathcal{D}_u}[\ell(k(f(\mathbf{X}^T; \theta^f_u); \theta^p_u), y)] \tag{1}$$

where $\ell(\cdot, \cdot)$ is a task-specific loss function. However, given the scarcity and imbalance of medical data, optimizing solely on local data typically leads to overfitting and poor generalization. Therefore, Med-FedLSG introduces a server-based generator for knowledge transfer, extending the optimization objective to:

$$\begin{aligned}\min_{\theta_u} \mathcal{J}(\theta_u) := \mathcal{L}_u(\theta_u) \\ + \mathbb{E}_{(Y,\mathbf{Z},\mathbf{W})\sim\mathcal{G}_m}\left[\ell\big(k(\mathbf{W}; \theta^p_u); Y\big)\right] \\ + \lambda_1\mathcal{L}_{KL} + \lambda_2\mathcal{L}_{personal}\end{aligned} \tag{2}$$

where:

- $\mathcal{L}_u(\theta_u)$ is the empirical loss on local data:

$$\mathcal{L}_u(\theta_u) = \frac{1}{|\mathcal{D}_u|}\sum_{(\mathbf{X}^T, y)\in\mathcal{D}_u} \ell(k(f(\mathbf{X}^T; \theta^f_u); \theta^p_u), y)$$

- The second term represents the loss derived from the server's generator, G_m, parameterized by m.
- $(Y, \mathbf{Z}, \mathbf{W}) \sim \mathcal{G}_m$ are sampled as: $Y \sim P(Y)$, $\mathbf{Z} \sim G_m(\mathbf{Z}|Y)$, $\mathbf{W} \sim G_m(\mathbf{W}|\mathbf{Z})$
- $\mathcal{L}_{KL}$ is a KL-divergence regularization term promoting standardized latent distributions.
- $\mathcal{L}_{personal} = \|\mathbf{z}_p - \mathbf{e}_u\|^2$ is a personalization regularization term ensuring consistency of personalized representations with user characteristics.

4) Server Generator Optimization: The server maintains a dual-generator model G_m to produce synthetic samples facilitating knowledge transfer. The optimization objective for the generator is defined as:

$$\begin{aligned}\min_m \mathcal{J}(m) = \mathbb{E}_{Y\sim P(Y)}\Big[\mathbb{E}_{\mathbf{Z}\sim G_m(\mathbf{Z}|Y)}\Big[\\ \mathbb{E}_{\mathbf{W}\sim G_m(\mathbf{W}|\mathbf{Z})}\big[\ell\big(\sigma\big(\tfrac{1}{|\mathcal{A}|}\sum_{u\in\mathcal{A}} g(\mathbf{W}; \theta^p_u)\big), Y\big) \\ + \lambda_{KL}\mathcal{L}_{KL} + \lambda_{physio}\mathcal{L}_{physio}\big]\Big]\Big]\end{aligned} \tag{3}$$

where $g(\mathbf{W}; \theta^p_u)$ are the logits produced by the predictor of user u, $\sigma(\cdot)$ is the softmax function, and $\mathcal{L}_{physio}$ ensures generated data adheres to physiological constraints.

5) Optimization Strategy: The Med-FedLSG framework employs a two-level parameter optimization strategy that distinguishes between shared parameters responsible for capturing common medical knowledge across all users, and personalized parameters that model individual-specific physiological characteristics. Shared parameters enable collaborative learning by aggregating generalizable medical patterns, while personalized parameters remain local to preserve individual privacy and capture user-specific health behaviors. To balance shared knowledge and personalized representations, Med-FedLSG adopts a two-level optimization strategy:

1. Shared parameters θ^s (including shared temporal encoder parameters θ^f_s that extract common physiological patterns across users, and conditional generator parameters θ^c that synthesize generalizable medical knowledge) are aggregated through federated averaging:

$$\theta^s_{global} = \frac{\sum_{u\in\mathcal{A}} w_u\theta^s_u}{\sum_{u\in\mathcal{A}} w_u}$$

2. Personalized parameters θ^p (including personalized temporal encoder parameters θ^f_p that capture user-specific physiological characteristics such as baseline heart rate variability and individual response patterns, and predictor parameters θ^p that adapt to personal health conditions and risk factors) are updated locally without aggregation:

$$\theta^p_{global,u} = \theta^p_u$$

This dual-parameter strategy ensures that common medical knowledge (e.g., general arrhythmia patterns, universal physiological constraints) is shared across all users through θ^s, while individual characteristics (e.g., personal baseline values, individual medication effects, user-specific device calibration) are preserved locally through θ^p. The shared parameters facilitate knowledge transfer and improve generalization, while personalized parameters enable accurate modeling of individual health profiles without compromising privacy.

IV. Med-FedLSG Method

This section details the core components of Med-FedLSG, including variational temporal representation learning, personalized representation disentanglement, and a federated optimization framework based on knowledge distillation. Together, these components form a comprehensive federated learning framework capable of effectively modeling medical time-series data, preserving privacy, and achieving personalized modeling simultaneously.

Specifically, we first employ variational temporal representation learning to capture the complexity and uncertainty inherent in medical data. Next, we propose a dimension selection strategy based on federated principal component analysis (Federated PCA) to explicitly partition shared and personalized representations without exchanging raw data, ensuring privacy protection. Finally, we utilize a conditional temporal generator and a federated optimization framework based on knowledge distillation to facilitate effective cross-user knowledge transfer while maintaining data privacy. The overall architecture of our proposed method is shown in Figure 2.

Med-FedLSG ensures privacy through three key mechanisms: (1) raw medical data never leaves local devices under the federated paradigm, (2) the conditional generator produces synthetic data with statistical dissimilarity (MMD distances of

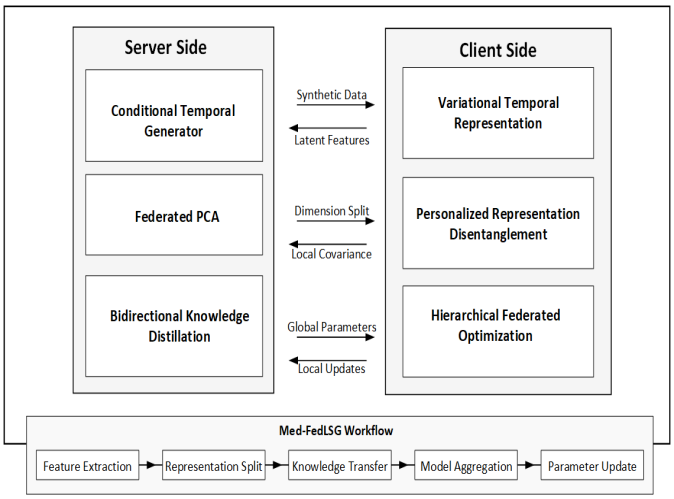


Fig. 2. Overall architecture of the proposed method.

0.138 and 0.121 on our datasets) that reduces direct reconstruction risks, and (3) personalized representations remain local to protect individual characteristics.

A. Variational Temporal Representation Learning

Medical time-series data exhibit complex temporal dynamics and individual differences, necessitating effective representation learning techniques. We adopt variational temporal representation learning to model medical time-series within a probabilistic framework that captures inherent uncertainty.

We encode time-series data using LSTM and employ variational inference to transform deterministic representations into probability distributions:

$$\mathbf{h}_t = \text{LSTM}(\mathbf{x}_t, \mathbf{h}_{t-1}), \quad t = 1, 2, \ldots, \tau \tag{4}$$

where $\mathbf{x}_t \in \mathbb{R}^p$ is the feature vector at time step t. The final representation $\mathbf{h} = \mathbf{h}_\tau$ is mapped to latent space parameters:

$$[\boldsymbol{\mu}, \log \boldsymbol{\sigma}^2] = \text{MLP}(\mathbf{h}) \tag{5}$$

Using the reparameterization trick, we generate the latent representation:

$$\mathbf{z} = \boldsymbol{\mu} + \boldsymbol{\varepsilon} \odot \sqrt{\boldsymbol{\sigma}^2}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(0, \mathbf{I}) \tag{6}$$

KL-divergence regularization ensures latent distribution regularity:

$$\mathcal{L}_{KL} = \frac{1}{2} \sum_{j=1}^{q} (1 + \log(\sigma_j^2) - \mu_j^2 - \sigma_j^2) \tag{7}$$

This probabilistic modeling approach facilitates personalized representation disentanglement and enables synthetic data generation for knowledge distillation in federated learning scenarios.

B. Personalized Representation Disentanglement

Medical data often exhibit significant individual variability due to physiological characteristics, medical history, and lifestyle differences. To accurately model this heterogeneity, we introduce a PCA-based mechanism to explicitly disentangle the latent space into shared and personalized components.

1) Representation Disentanglement: Specifically, we explicitly partition the latent representation $\mathbf{z} \in \mathbb{R}^q$ obtained from the variational encoder into two components:

$$\mathbf{z} = [\mathbf{z}_s; \mathbf{z}_p], \quad \mathbf{z}_s \in \mathbb{R}^{q-d}, \mathbf{z}_p \in \mathbb{R}^d \tag{8}$$

where $\mathbf{z}_s$ captures general health patterns shared across users, while $\mathbf{z}_p$ focuses on user-specific medical patterns.

Two independent fully connected networks extract these representations from the encoder's output parameters ($\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$):

$$\mathbf{z}_s = \text{FC}_{shared}(\boldsymbol{\mu}, \boldsymbol{\sigma}), \quad \mathbf{z}_p = \text{FC}_{personal}(\boldsymbol{\mu}, \boldsymbol{\sigma}) \tag{9}$$

This explicit disentanglement facilitates targeted optimization of both shared and personalized representations in federated learning, enhancing the model's sensitivity and adaptability to individual differences.

2) Dimension Selection Using PCA: To systematically determine dimensions for shared and personalized representations, we propose a PCA-based dimension selection strategy as follows:

1) **Data Preparation**: Extract latent representations $\mathbf{Z} \in \mathbb{R}^{n \times q}$ from selected users.
2) **PCA Computation**: Perform PCA on $\mathbf{Z}$ to obtain eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_q$ and calculate the variance explained:

$$\eta_j = \frac{\lambda_j}{\sum_{k=1}^{q} \lambda_k} \tag{10}$$

3) **Dimension Selection**: Determine the shared dimension based on cumulative variance explained exceeding threshold γ (e.g., $\gamma = 0.9$):

$$q - d = \arg\min_m \left(\sum_{j=1}^{m} \eta_j \geq \gamma \right) \tag{11}$$

4) **Explicit Partitioning**: Set the first $q - d$ principal components as shared and the remaining d as personalized dimensions.

This PCA-based strategy ensures shared dimensions effectively capture common health knowledge, while personalized dimensions reflect unique individual health patterns.

3) Federated PCA Implementation: Conventional PCA requires global data access, which is not feasible under federated settings. Therefore, we employ Federated PCA [32], protecting user privacy and ensuring practical feasibility:

1) **Local Computation**: Each client computes local covariance matrices:

$$\mathbf{C}_u = \frac{1}{n_u - 1} (\mathbf{Z}_u - \bar{\mathbf{Z}}_u)^\top (\mathbf{Z}_u - \bar{\mathbf{Z}}_u) \tag{12}$$

2) **Aggregation**: Server aggregates covariance matrices:

$$\mathbf{C}_{global} = \frac{1}{\sum_u n_u} \sum_u n_u \mathbf{C}_u \tag{13}$$

3) **Dimension Partitioning**: Server performs eigendecomposition on $\mathbf{C}_{global}$ and computes cumulative variance to determine dimension split.

4) **Broadcast Results**: The dimension partitioning results are sent back to clients for constructing their local representation networks.

This federated PCA method enables accurate dimension selection without exchanging raw data, maintaining privacy compliance within the federated framework.

4) Regularization for Personalized Representation: To further stabilize personalized representations and maintain consistency across sessions, we introduce personalized embeddings $\mathbf{e}_u$:

$$\mathcal{L}_{personal} = \|\mathbf{z}_p - \mathbf{e}_u\|^2 \tag{14}$$

The embeddings are updated via exponential moving average (EMA):

$$\mathbf{e}_u \leftarrow (1-\alpha)\mathbf{e}_u + \alpha \mathbf{z}_p \tag{15}$$

This approach smooths individual-specific medical features over time, preventing overfitting to single-session data.

C. Conditional Temporal Generator

In federated learning, effectively sharing knowledge while preserving privacy remains a core challenge. We design a conditional temporal generator to produce high-quality personalized synthetic samples, facilitating knowledge transfer without exchanging raw data.

1) Conditional Variational Generation: The conditional temporal generator takes latent representations $\mathbf{z}$ and health labels y as inputs to generate corresponding feature representations:

$$\mathbf{w} = c(\mathbf{z}, y) = c([\mathbf{z}_s; \mathbf{z}_p], y) \tag{16}$$

We employ a conditional variational autoencoder (CVAE) structure, allowing precise control of the generation process based on health status y:

$$\begin{aligned} \mathbf{h}_{cond} &= \text{MLP}([\mathbf{z}; \text{Embedding}(y)]) \\ \mathbf{w} &= \text{Activation}(\mathbf{h}_{cond}) \end{aligned} \tag{17}$$

This conditional approach allows generating diverse and targeted synthetic samples corresponding to different health states.

2) Temporal Reconstruction and Constraints: To reconstruct realistic medical time-series data, we utilize an LSTM-based decoder to transform the feature representation $\mathbf{w}$ into sequential data:

$$\begin{aligned} \mathbf{h}_0 &= \text{MLP}_{init}(\mathbf{w}) \\ \mathbf{h}_t &= \text{LSTM}(\mathbf{h}_{t-1}, \mathbf{w}), \quad t = 1, 2, \ldots, \tau \\ \mathbf{x}'_t &= \text{MLP}_{output}(\mathbf{h}_t), \quad t = 1, 2, \ldots, \tau \end{aligned} \tag{18}$$

We introduce physiological constraints as regularization to ensure the generated data is medically realistic:

$$\begin{aligned} \mathcal{L}_{physio} = &\sum_{t=1}^{\tau}\sum_{j=1}^{p}\left(\max(0, x'_{t,j} - u_j)^2 + \max(0, l_j - x'_{t,j})^2\right) \\ &+ \lambda_{smooth}\sum_{t=2}^{\tau}\|\mathbf{x}'_t - \mathbf{x}'_{t-1}\|^2 \end{aligned} \tag{19}$$

The first term ensures generated values remain within physiologically plausible bounds $[l_j, u_j]$, and the second promotes smooth temporal variations.

This conditional temporal generator facilitates personalized synthetic time-series data creation, allowing cross-user knowledge sharing without exchanging raw data, addressing challenges of data heterogeneity and privacy protection in federated learning.

D. Federated Optimization Framework Based on Knowledge Distillation

Building upon the variational temporal representation learning, personalized representation disentanglement, and conditional temporal generator, we propose a federated optimization framework utilizing knowledge distillation to coordinate the optimization between the server and clients.

This federated optimization design inherently provides privacy preservation. Med-FedLSG ensures privacy through three key mechanisms: (1) raw medical data never leaves local devices under the federated paradigm, (2) the conditional generator produces synthetic data with statistical dissimilarity (MMD distances of 0.138 and 0.121 on our datasets) that reduces direct reconstruction risks, and (3) personalized representations remain local to protect individual characteristics.

1) Bidirectional Knowledge Distillation: Med-FedLSG employs bidirectional knowledge distillation, involving global distillation from the server to clients and personalized knowledge integration from clients to the server:

1. Global knowledge distillation: The server shares integrated knowledge with clients via generated synthetic samples:

$$\mathcal{L}_{global\_distill} = \mathbb{E}_{Y,\mathbf{Z},\mathbf{W}\sim G_m}[\ell(k(\mathbf{W}; \theta_u^p); Y)] \tag{20}$$

2. Personalized knowledge integration: Clients communicate personalized knowledge back to the server through model predictions:

$$\mathcal{L}_{personal\_distill} = D_{KL}\left[\sigma\left(\frac{g_u(\mathbf{W})}{\tau}\right) \| \sigma\left(\frac{g_{server}(\mathbf{W})}{\tau}\right)\right] \tag{21}$$

where g_u and g_{server} denote the client and server predictors respectively, and τ is the temperature parameter.

This mechanism allows effective knowledge exchange without raw data transmission while preserving personalized modeling capability.

2) Hierarchical Federated Optimization Objective: For each user u, we define a hierarchical optimization objective:

$$\begin{aligned} \min_{\theta_u} \mathcal{J}(\theta_u) = &\mathcal{L}_{empirical} + \lambda_1\mathcal{L}_{global\_distill} + \lambda_2\mathcal{L}_{KL} \\ &+ \lambda_3\mathcal{L}_{personal} + \lambda_4\mathcal{L}_{personal\_distill} \end{aligned} \tag{22}$$

where the empirical loss on local data is:

$$\mathcal{L}_{empirical} = \frac{1}{|\mathcal{D}_u|}\sum_{(\mathbf{X}^T, y)\in\mathcal{D}_u} \ell(k(f(\mathbf{X}^T; \theta_u^f); \theta_u^p), y) \tag{23}$$

This hierarchical design balances local data fitting, global knowledge transfer, and personalized representation learning.

3) Two-Level Model Aggregation: To balance shared and personalized knowledge, we introduce a two-level model aggregation strategy:

1. Aggregation of shared parameters (θ^s, including shared encoder parameters θ_s^f and generator parameters θ^c):

$$\theta_{global}^s = \frac{1}{|\mathcal{A}|} \sum_{u \in \mathcal{A}} \theta_u^s \tag{24}$$

2. Preservation of personalized parameters (θ^p, including personalized encoder parameters θ_p^f and predictor parameters θ^p):

$$\theta_{global,u}^p = \theta_u^p \tag{25}$$

During client updates, parameters are updated as follows:

$$\theta_u^s \leftarrow \theta_{global}^s, \quad \theta_u^p \leftarrow \theta_u^p \text{ (unchanged)} \tag{26}$$

This strategy effectively handles medical data heterogeneity by retaining personalized parameters while sharing general medical knowledge.

E. Algorithm Workflow

The complete Med-FedLSG algorithm consists of initialization and iterative federated training phases, detailed in Algorithm 1.

In summary, Med-FedLSG achieves personalized federated learning for medical time-series data through: (1) variational temporal representation learning, (2) explicit disentanglement of shared and personalized representations, (3) bidirectional knowledge distillation, and (4) two-level model aggregation. These components form an integrated framework effectively addressing data heterogeneity, privacy protection, and personalized modeling in medical applications.

Algorithm 1 Med-FedLSG (with Federated PCA-Based Dimension Partitioning)

Input: User set $\mathcal{U}$, total rounds T, local epochs E
Output: Global model parameters and personalized parameters for each user
// Initialization
Initialize model parameters $\theta = [\theta^f; \theta^c; \theta^p]$, generator m, user embeddings $\mathbf{E}$
Perform federated PCA: each user uploads local covariance $\mathbf{C}_u$, server computes global covariance $\mathbf{C}_{global}$ and determines shared/personal dimensions
Broadcast dimension partitioning; each user initializes FC_{shared} and $\text{FC}_{personal}$
// Federated Training
for round $t = 1$ to T **do**
  Server selects user subset $\mathcal{A}_t$ and broadcasts θ, m, and $P(Y)$
  for each user $u \in \mathcal{A}_t$ in parallel **do**
    Initialize local shared parameters from global model
    Generate pseudo samples using generator m
    for epoch $e = 1$ to E **do**
      Train local model with real and pseudo data to minimize composite loss $\mathcal{J}(\theta_u)$
      Update user embedding $\mathbf{e}_u$
    end for
    Upload updated θ_u and label statistics to server
  end for
  Server aggregates shared parameters and updates label distribution $P(Y)$
  if $t \bmod T_{update} = 0$ **then**
    Update generator m using collected pseudo data feedback
  end if
end for
Return Final global parameters θ_{global} and personalized parameters $\{\theta_u^p\}_{u \in \mathcal{U}}$

V. Experimental Evaluation and Analysis

In this section, we systematically evaluate the effectiveness and advantages of the proposed Med-FedLSG framework in federated learning for medical time series. We first present the experimental setup and parameter configuration, then provide comparative results with existing methods, followed by detailed ablation analysis and discussion.

A. Experimental Setup

1) Datasets Description and Preprocessing: We conducted comprehensive evaluations using two representative medical time-series datasets. From the MIMIC-III (Medical Information Mart for Intensive Care III) clinical database, physiological monitoring data from 5,000 ICU patients were extracted from the original 40,000 patient records. Key physiological indicators, including heart rate, systolic/diastolic blood pressure, respiratory rate, body temperature, and oxygen saturation, were selected to construct continuous 48-hour time series samples. We defined a clinically relevant binary prediction task: predicting the mortality risk of patients within a 48-hour period. The UCI-HAR (University of California Irvine Human Activity Recognition) dataset consists of accelerometer and gyroscope data collected from smartphones worn by 30 subjects at a sampling rate of 50Hz. The raw three-axis acceleration and angular velocity data were segmented into windows of 2.56 seconds (128 data points) for classifying six common activities: walking, ascending stairs, descending stairs, sitting, standing, and lying.

To realistically simulate data heterogeneity encountered in medical scenarios, we employed non-IID data partitioning. For the MIMIC-III dataset, we generated 50 virtual users grouped by age categories (< 40 years, 40-60 years, > 60 years), introducing feature distribution shifts across groups and label distribution shifts by assigning a higher proportion of high-risk samples to the elderly group. Each user had approximately 100 time series samples. For the UCI-HAR dataset, we maintained the original 30 subjects as independent users. Each user primarily retained 80% of the data for 2-4 main activity types, with only 20% for remaining activities. Additionally, random noise with a standard deviation of 0-5% was added to simulate

inter-device variability, creating a significantly heterogeneous data environment.

2) Implementation Details: The core architecture of Med-FedLSG includes a temporal encoder, generator network, and predictor network. The temporal encoder consists of three layers of bidirectional LSTM (256 hidden units each), integrated with an attention mechanism to identify critical time points, and batch normalization layers to enhance training stability. The latent space dimension was set to 128, with the ratio between shared and personalized representations dynamically determined via federated PCA, initialized at 8:2. The generator network is structured as a three-layer MLP (256-128-64), incorporating residual connections to mitigate gradient vanishing issues in deep networks and Leaky ReLU activation to handle anomalies in medical data effectively. The predictor network comprises a two-layer MLP (128-64-output dimension), complemented by a dropout rate of 0.3 to reduce overfitting.

During training, we utilized the Adam optimizer (initial learning rate of 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.999$), along with weight decay (1e-5) for enhanced generalization. The learning rate followed a stepwise decay strategy, decreasing by 10% every 10 communication rounds, with a minimum of 0.0001. Batch sizes were dynamically adjusted based on dataset scale: 64 for MIMIC-III and 32 for UCI-HAR. Each experimental scenario was repeated five times, and results were reported as mean performance with standard deviation, ensuring statistical reliability.

TABLE I

PERFORMANCE COMPARISON OF DIFFERENT METHODS ON THE MIMIC-III DATASET (MEAN ± STD)

Method	Accuracy (%)	F1 Score	AUC-ROC
FedAvg	75.83 ± 1.42	0.732 ± 0.021	0.806 ± 0.018
FedProx	77.64 ± 1.31	0.758 ± 0.019	0.821 ± 0.017
pFedMe	80.27 ± 1.18	0.785 ± 0.016	0.845 ± 0.014
FedBN	79.56 ± 1.23	0.778 ± 0.018	0.839 ± 0.015
FedMCSA	83.36 ± 0.97	0.817 ± 0.014	0.869 ± 0.012
FedPHP	83.94 ± 0.95	0.823 ± 0.014	0.875 ± 0.012
pFedCK	83.67 ± 0.97	0.821 ± 0.015	0.872 ± 0.013
CFG-SHD	84.21 ± 0.93	0.826 ± 0.013	0.878 ± 0.011
Ref. [33]	81.92 ± 1.05	0.803 ± 0.015	0.857 ± 0.013
Med-FedLSG	**85.47** ± 0.89	**0.843** ± 0.012	**0.891** ± 0.010

TABLE II

PERFORMANCE COMPARISON OF DIFFERENT METHODS ON THE UCI-HAR DATASET (MEAN ± STD)

Method	Accuracy (%)	F1 Score	AUC-ROC
FedAvg	85.32 ± 1.24	0.831 ± 0.019	0.917 ± 0.016
FedProx	86.85 ± 1.15	0.852 ± 0.017	0.926 ± 0.015
pFedMe	88.47 ± 1.06	0.873 ± 0.015	0.938 ± 0.013
FedBN	87.95 ± 1.09	0.865 ± 0.016	0.933 ± 0.014
FedMCSA	90.38 ± 0.92	0.895 ± 0.013	0.951 ± 0.011
FedPHP	90.85 ± 0.88	0.901 ± 0.012	0.956 ± 0.010
pFedCK	90.62 ± 0.90	0.898 ± 0.013	0.954 ± 0.011
CFG-SHD	91.15 ± 0.87	0.905 ± 0.012	0.959 ± 0.010
Ref. [33]	89.26 ± 0.98	0.882 ± 0.014	0.943 ± 0.012
Med-FedLSG	**92.28** ± 0.83	**0.917** ± 0.011	**0.967** ± 0.009

B. Experimental Evaluations

We compared Med-FedLSG with several representative federated learning algorithms, including both general federated learning approaches and those specifically designed for healthcare applications: FedAvg [18] as the standard baseline; FedProx [25] with proximal regularization; pFedMe [34] for personalized federated meta-learning; FedBN [35], retaining local batch normalization layers; FedMCSA [36], using a self-attention mechanism to handle non-IID data. FedPHP [37], which employs inherited private models with temporal ensembling of historical personalized models for enhanced knowledge transfer; and pFedCK [38], which combines clustering-based client selection with knowledge distillation between personalized and interactive models. Additionally, to ensure strong relevance to the healthcare domain, we included two specialized methods: CFG-SHD [11], a co-training-based personalized federated learning approach with generative adversarial networks specifically designed for smart healthcare diagnosis; and a differential privacy-enhanced federated learning framework [33] developed for medical image data that balances privacy protection and diagnostic accuracy in clinical applications. To ensure fair evaluation, we retain the original CFG-SHD framework and minimally adapt the convolutional layers to one-dimensional operations for compatibility with sequential data. Similarly, for the federated learning framework with differential privacy, we preserve the overall structure and minimally replace the input module with a 1D-CNN suitable for time-series inputs. For FedPHP and pFedCK, we implement their core personalization mechanisms while adapting the model architectures to be compatible with our medical time-series tasks.

Firstly, we conducted systematic comparative experiments on these two significantly heterogeneous datasets. Tables I and II summarize the final testing performance of all compared methods across accuracy, F1 score, and AUC-ROC metrics. The results clearly demonstrate that Med-FedLSG significantly outperforms existing federated learning methods in all evaluation metrics. Specifically, on the MIMIC-III dataset, Med-FedLSG achieves an accuracy of 85.47%, an F1 score of 0.843, and an AUC-ROC of 0.891, surpassing the second-best method CFG-SHD by approximately 1.26%, 0.015, and 0.013 respectively. The advanced personalized methods FedPHP and pFedCK achieve competitive performance with 83.94% and 83.67% accuracy respectively, demonstrating the effectiveness of their personalization mechanisms, yet still falling short of our comprehensive approach that combines variational representation learning, explicit feature disentanglement, and conditional temporal generation. Similarly, on the UCI-HAR dataset, Med-FedLSG obtains the highest accuracy of 92.28%, F1 score of 0.917, and AUC-ROC of 0.967. These improvements validate the effectiveness of integrating latent space personalization and adaptive representation learning within the proposed framework, particularly in addressing data heterogeneity inherent in medical applications.

In medical applications, the stability and consistency of model predictions are of critical importance. To evaluate these aspects, we introduce the metric of *Intra-User Consistency* (IUC), which quantifies how consistent a model's predictions

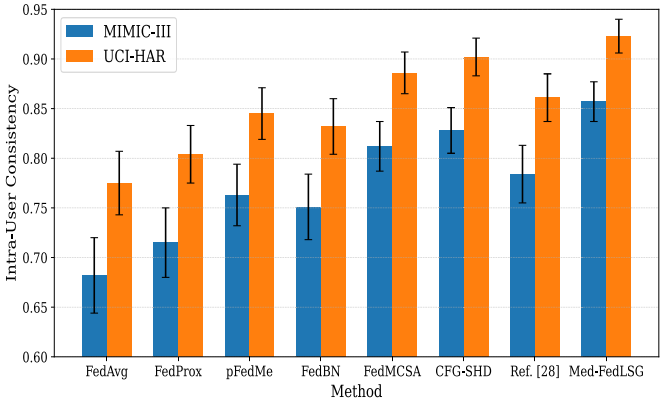


Fig. 3. Intra-user consistency comparison across methods.

TABLE III
PERSONALIZED DIMENSIONAL RATIO (%) OF DIFFERENT METHODS

MIMIC-III Dataset	
Method	**Personalized Dim. Ratio (%)**
Med-FedLSG-NP	0
Med-FedLSG-RS	30 (fixed)
Med-FedLSG-Full	24.6 (adaptive)
UCI-HAR Dataset	
Method	**Personalized Dim. Ratio (%)**
Med-FedLSG-NP	0
Med-FedLSG-RS	30 (fixed)
Med-FedLSG-Full	27.3 (adaptive)

are for the same user across different time windows. The results of this evaluation are presented in Figure 3. A higher IUC score indicates that the model consistently captures the long-term health characteristics of individual users, rather than reacting to short-term variations. This is particularly important in long-term medical monitoring scenarios, such as chronic disease management, where stable and reliable predictions are crucial for effective clinical decision-making.

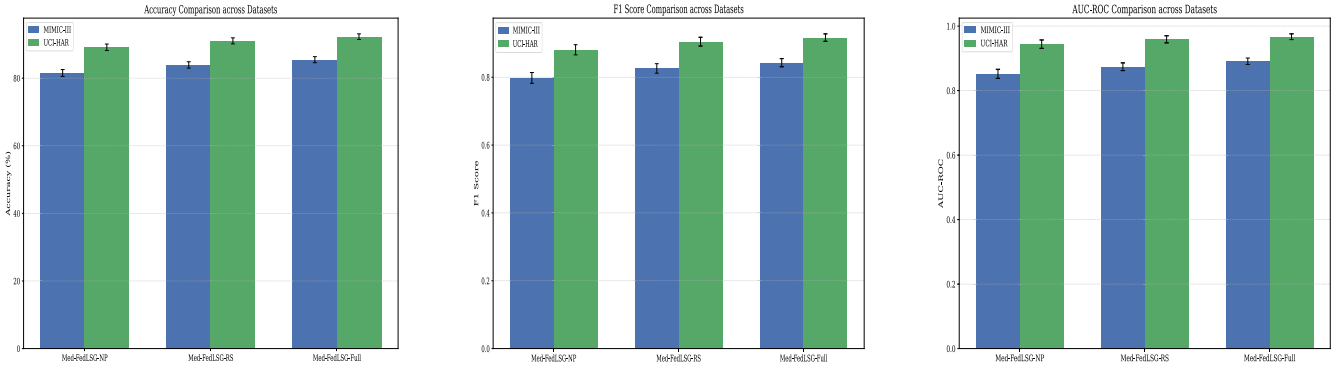
As illustrated in Figure 3, baseline methods such as FedAvg and FedProx exhibit relatively lower consistency, suggesting limited capability in capturing stable user-specific patterns over time. In contrast, methods incorporating personalization mechanisms, such as pFedMe and FedBN, demonstrate notable gains. FedMCSA further enhances consistency by leveraging latent representations or attention mechanisms. CFG-SHD achieves the second-highest consistency score, primarily due to its effective integration of co-training and GAN-based synthetic data generation that better captures individual physiological characteristics. Med-FedLSG achieves the highest consistency across both datasets, highlighting its ability to generate stable and coherent predictions for individual users. This result indicates that our method not only maintains strong predictive performance but also more accurately captures users' long-term stable health characteristics, thereby offering more reliable decision support for clinical applications.

Next, we conducted a series of ablation experiments. One of the core innovations of Med-FedLSG lies in its personalized representation separation mechanism, which is based on federated PCA. To validate its effectiveness, we conducted systematic ablation studies by comparing three model variants: (1) Med-FedLSG-Full, the complete version of our model with adaptive dimension allocation guided by federated PCA; (2) Med-FedLSG-NP, where the separation mechanism is removed and all latent dimensions are treated as shared; and (3) Med-FedLSG-RS, which replaces the PCA-based strategy with a fixed-ratio random separation.

Table III presents the proportion of latent dimensions allocated to personalized representations in each model variant. As we can see, Med-FedLSG-NP assigns no dimensions to personalization, while Med-FedLSG-RS enforces a fixed allocation of 30%. In contrast, Med-FedLSG-Full employs an adaptive strategy based on federated PCA, resulting in slightly lower but data-driven ratios of 24.6% and 27.3% on the MIMIC-III and UCI-HAR datasets, respectively. The personalized dimensional ratios in Table III serve as the underlying configuration for the model variants evaluated in Figure 4. This figure shows the performance of the three variants on the MIMIC-III and UCI-HAR datasets. The results clearly demonstrate that the personalized representation separation mechanism significantly contributes to performance improvement. On the MIMIC-III dataset, Med-FedLSG-Full outperforms Med-FedLSG-NP by a substantial margin in accuracy, suggesting that explicitly distinguishing between shared and personalized features helps better capture the heterogeneity of medical data. Additionally, Med-FedLSG-Full achieves higher accuracy than Med-FedLSG-RS, validating the advantage of adaptive dimension selection guided by federated PCA. A similar trend is observed on the UCI-HAR dataset, indicating the generalizability of the mechanism across different types of medical time-series data.

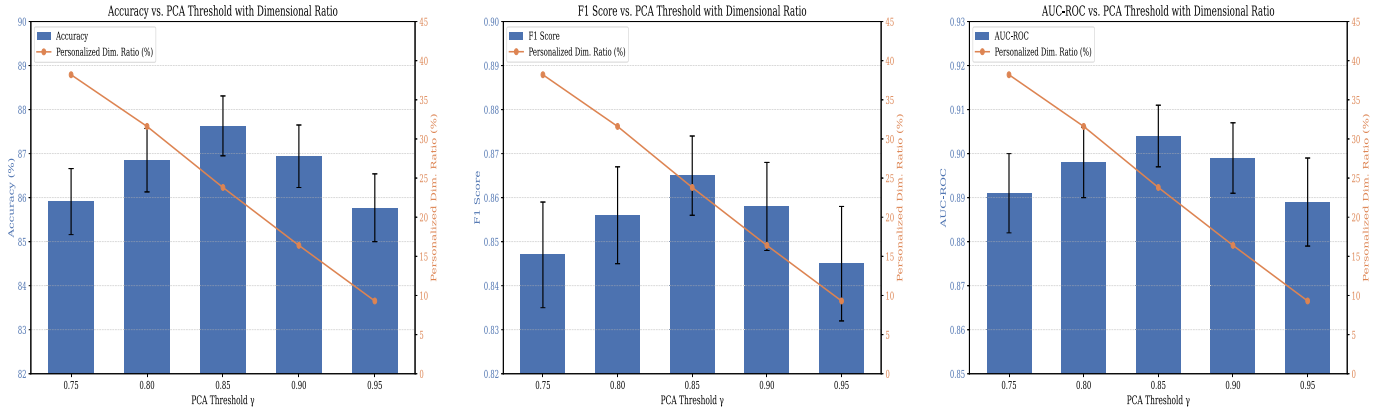
The PCA threshold γ is a critical hyperparameter that determines the ratio between shared and personalized representation dimensions. Figure 5 presents the effects of varying γ on performance and dimensionality allocation. At $\gamma = 0.85$, the model achieves optimal performance with a personalized dimension ratio of 23.8%, yielding the highest accuracy of 87.63% on the MIMIC-III dataset. Lower values of γ (e.g., 0.75) assign a larger proportion of dimensions to personalized components, which enhances individual expressiveness but weakens knowledge sharing across users, slightly reducing performance. Conversely, higher values (e.g., 0.95) overly restrict personalized capacity, leading to underfitting of user-specific characteristics. These results validate the effectiveness of our adaptive dimension allocation strategy in balancing generalization and personalization.

To further examine the semantic structure captured by the personalized representations, we conduct a clustering analysis on user representations from the MIMIC-III dataset. Table IV summarizes the clustering performance across various demographic and clinical subgroups, including age brackets, gender, and illness severity levels. Notably, the clustering accuracy ranges from 84.96% to 89.24%, indicating that the personalized representation space effectively preserves meaningful user distinctions. Among all subgroups, the elderly population (> 60 years) and patients with severe conditions achieve



(a) Accuracy Comparison (b) F1 Score Comparison (c) AUC-ROC Comparison

Fig. 4. Performance comparison of different personalization strategies on MIMIC-III and UCI-HAR datasets. Error bars indicate standard deviation across multiple runs.



(a) Accuracy vs. PCA Threshold (b) F1 Score vs. PCA Threshold (c) AUC-ROC vs. PCA Threshold

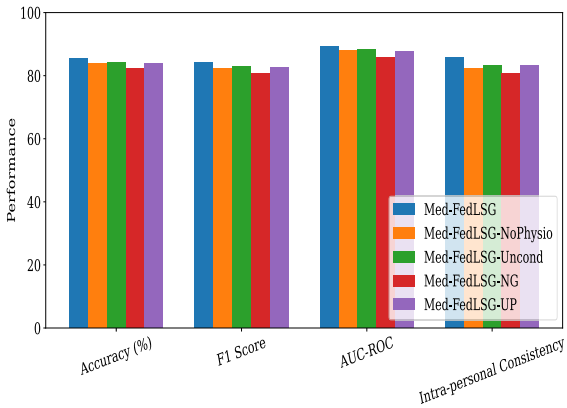
Fig. 5. Impact of PCA threshold γ on dimension allocation and performance (MIMIC-III dataset).

TABLE IV
CLUSTERING RESULTS OF PERSONALIZED REPRESENTATIONS BY USER GROUP (MIMIC-III DATASET)

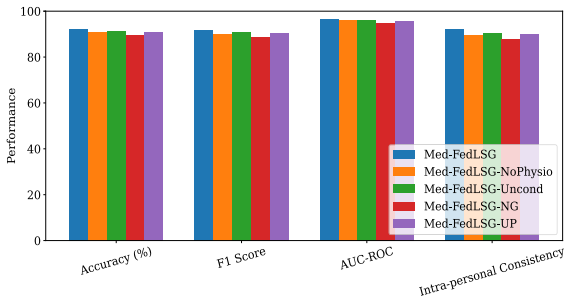
User Group	Clustering Accuracy (%)	Silhouette Score	Inter-Cluster Dist.	Intra-Cluster Dist.
Age $<$ 40	87.35	0.713	3.85	1.42
Age 40–60	85.92	0.682	3.64	1.57
Age $>$ 60	88.41	0.726	3.97	1.35
Gender: Male	86.73	0.695	3.72	1.49
Gender: Female	87.05	0.704	3.78	1.45
Severity: Mild	84.96	0.673	3.58	1.62
Severity: Moderate	86.37	0.688	3.69	1.53
Severity: Severe	89.24	0.741	4.05	1.28

the highest clustering accuracies of 88.41% and 89.24%, respectively. These groups also exhibit superior silhouette coefficients (0.726 and 0.741), larger inter-cluster distances, and smaller intra-cluster distances, suggesting higher internal consistency and clearer separation from other groups. Such results demonstrate that the proposed Med-FedLSG framework effectively captures salient user-specific characteristics, particularly for populations with more pronounced medical heterogeneity.

We further evaluate the effectiveness of the conditional temporal generator, a key component introduced in the Med-FedLSG framework to facilitate knowledge transfer in federated learning by generating high-quality pseudo samples. The evaluation is conducted from three perspectives: comparison of model variants, analysis of the impact of pseudo sample ratios, and assessment of the quality of generated data. We first compare five model variants to isolate the contributions of different generator components: the complete Med-FedLSG, Med-FedLSG-NG (with the generator module removed), Med-FedLSG-UP (utilizing an unconditioned generator), Med-FedLSG-NoPhysio (removing physiological constraints by setting $\lambda_{physio} = 0$), and Med-FedLSG-Uncond (removing the conditional mechanism where the generator operates without health label y as input).

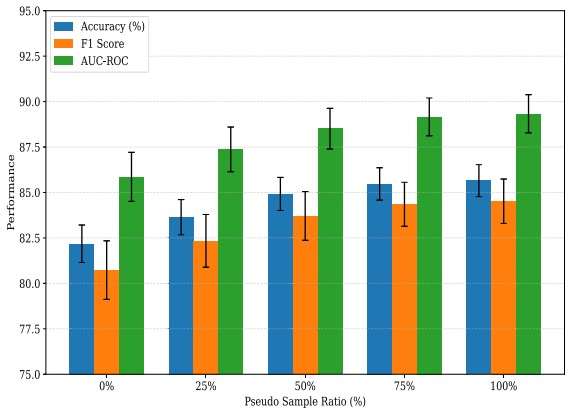


(a) Performance Comparison on MIMIC-III

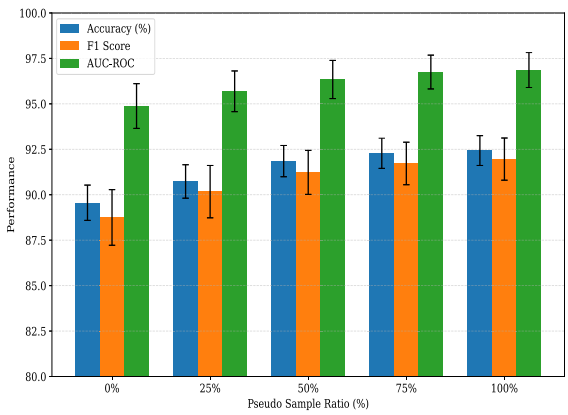


(b) Performance Comparison on UCI-HAR

Fig. 6. Comparison of different med-fedlsg variants on two datasets across four evaluation metrics.



(a) Performance Comparison on MIMIC-III



(b) Performance Comparison on UCI-HAR

Fig. 7. Impact of pseudo sample ratio on performance of two datasets.

As shown in Figure 6, each generator component significantly influences model performance. On the MIMIC-III dataset, Med-FedLSG achieves an improvement of 3.29 percentage points in accuracy over Med-FedLSG-NG, along with increases in F1 score, AUC-ROC, and intra-user consistency. The physiological constraints contribute 1.53 percentage points in accuracy (85.47% vs. 83.94% for Med-FedLSG-NoPhysio), while the conditional mechanism adds 1.18 percentage points (85.47% vs. 84.29% for Med-FedLSG-Uncond). Compared to the unconditioned variant (Med-FedLSG-UP), the complete model also yields notable gains, confirming the importance of conditioning in temporal generation. A similar trend is observed on the UCI-HAR dataset, where Med-FedLSG improves accuracy by 2.72 percentage points compared to its no-generator variant, with physiological constraints contributing 1.26 percentage points and the conditional mechanism adding 0.81 percentage points. These results suggest that the conditional temporal generator effectively mitigates non-IID challenges in medical federated learning by enabling privacy-preserving cross-user knowledge transfer, with both physiological constraints and conditional mechanisms playing essential roles.

To identify an optimal strategy for incorporating pseudo samples during training, we analyze the effect of varying pseudo sample ratios. As shown in Figure 7, model performance improves as the proportion of pseudo samples increases, but begins to plateau around 75%. On the MIMIC-III dataset, increasing the pseudo sample ratio from 0% to 75% improves accuracy by 3.29%, while the gain from 75% to 100% is only 0.18%. Similar trends are observed for other evaluation metrics. This indicates that a moderate amount of pseudo data effectively enhances model performance, while excessive reliance may lead to distributional drift. Based on this observation, we adopt a 75% pseudo sample ratio as the default configuration in Med-FedLSG to balance accuracy and computational efficiency.

To comprehensively evaluate the quality of generated samples, we analyze both their statistical properties and utility in downstream tasks, as presented in Tables V and VI. For the MIMIC-III dataset, the relative differences between generated and real samples in terms of mean, standard deviation, and autocorrelation are 5.83%, 8.27%, and 7.12%, respectively, with a maximum mean discrepancy (MMD) of 0.138. The UCI-HAR dataset shows slightly lower variation, with relative differences ranging from 4.87% to 7.25%. These results suggest that the generator is capable of simulating realistic medical time-series data. Notably, the utility of generated data in downstream tasks is high, with only minor AUC-ROC differences of 2.47% (MIMIC-III) and 1.65% (UCI-HAR) compared to real samples. In addition, the physiological constraint satisfaction rates reach 94.37% and 96.85% on MIMIC-III and UCI-HAR, respectively, indicating that the generated sequences adhere to realistic physiological patterns. The radar chart in Figure 8 illustrates the relative differences across multiple evaluation metrics, further validating the fidelity of synthetic data across both datasets.

To analyze the scalability of Med-FedLSG, we examine the theoretical communication and computational complexity. The communication cost per round scales linearly with the number

TABLE V
QUALITY EVALUATION OF GENERATED SAMPLES (MIMIC-III DATASET)

Evaluation Metric	Generated Samples	Real Samples	Relative Difference (%)
Mean Difference	0.127 ± 0.018	–	5.83
Standard Deviation Difference	0.185 ± 0.022	–	8.27
Autocorrelation Difference	0.154 ± 0.019	–	7.12
MMD Distance	0.138 ± 0.014	–	6.41
Downstream Task Effectiveness (AUC-ROC)	0.869 ± 0.013	0.891 ± 0.010	2.47
Physiological Constraint Satisfaction Rate (%)	94.37 ± 1.25	100	5.63

TABLE VI
QUALITY EVALUATION OF GENERATED SAMPLES (UCI-HAR DATASET)

Evaluation Metric	Generated Samples	Real Samples	Relative Difference (%)
Mean Difference	0.109 ± 0.015	–	4.87
Standard Deviation Difference	0.163 ± 0.019	–	7.25
Autocorrelation Difference	0.132 ± 0.016	–	5.94
MMD Distance	0.121 ± 0.012	–	5.36
Downstream Task Effectiveness (AUC-ROC)	0.951 ± 0.011	0.967 ± 0.009	1.65
Physiological Constraint Satisfaction Rate (%)	96.85 ± 1.08	100	3.15

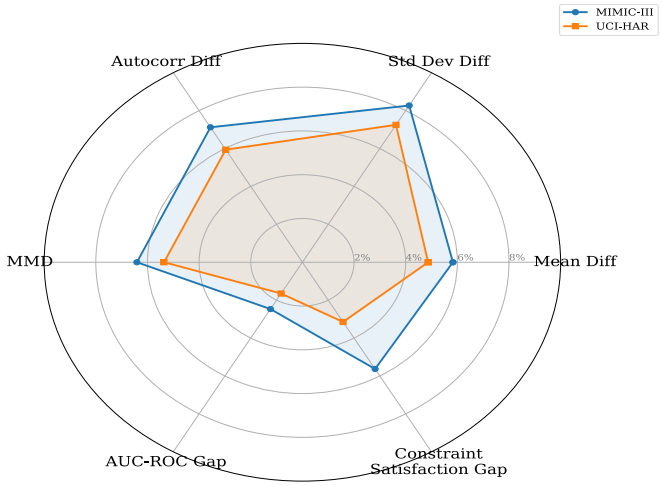


Fig. 8. Relative difference between generated and real samples.

of participating clients, requiring O(K) parameter transmissions for K selected clients. Given that only shared parameters (1.6M) are transmitted while personalized parameters remain local, the framework maintains constant per-client communication overhead regardless of total federation size. The server's aggregation complexity is O(K), making it computationally feasible for large-scale deployments.

To validate this theoretical analysis, we conduct additional simulations with up to 200 virtual clients on MIMIC-III by further partitioning the existing data. Results show that Med-FedLSG maintains stable convergence with accuracy degradation of only 1.2% (84.23% vs 85.47%) when scaling from 50 to 200 clients, while requiring 15% more communication rounds to achieve convergence. The framework demonstrates robust performance even with client participation rates as low as 10% (20 out of 200 clients per round).

To evaluate the robustness of Med-FedLSG against system failures, we simulate random client dropout scenarios where 10-30% of selected clients fail to participate in each communication round. Results show that the framework maintains stable performance with only 2.1% accuracy degradation under 15% dropout rate, demonstrating resilience to client unavailability commonly encountered in real-world deployments. Med-FedLSG incorporates several design elements that provide inherent robustness: (1) the two-level parameter aggregation prevents individual malicious clients from completely compromising the global model since personalized parameters remain local, (2) the federated PCA-based dimension allocation reduces the impact of outlier representations, and (3) the conditional generator's physiological constraints help filter unrealistic synthetic data that might result from corrupted inputs.

To assess the practical feasibility of Med-FedLSG for deployment on resource-constrained devices, we analyze the computational and communication overhead. The total model parameters of Med-FedLSG is approximately 2.1M, which is comparable to other personalized federated learning methods such as pFedMe (1.8M) and CFG-SHD (2.3M). The communication overhead per round is dominated by shared parameter transmission (approximately 1.6M parameters), while personalized parameters remain local. Memory requirements during training peak at approximately 45MB on MIMIC-III and 32MB on UCI-HAR, which falls within the capabilities of modern mobile devices and edge computing platforms such as smartphones and IoT healthcare devices.

VI. CONCLUSION

In this paper, we proposed Med-FedLSG, a novel federated learning framework for medical time-series data from wearable health consumer electronics. Our framework effectively addresses three critical challenges in federated learning for medical applications: data heterogeneity, privacy preservation, and knowledge transfer. The key innovations include a variational temporal representation learning mechanism with explicit disentanglement of shared and personalized features, and a conditional temporal generator that facilitates knowledge

transfer while maintaining privacy. Experimental results on the MIMIC-III clinical database and UCI-HAR dataset demonstrated that Med-FedLSG consistently outperforms existing federated learning methods across multiple evaluation metrics, including accuracy, F1 score, and AUC-ROC. Our ablation studies validated the effectiveness of the personalized representation separation mechanism and the conditional temporal generator. The adaptive dimension allocation strategy guided by federated PCA significantly improved model performance by balancing generalization and personalization capabilities.

While our framework demonstrates promising performance and theoretical scalability, several important limitations point to critical directions for future research. First, comprehensive evaluation in large-scale federated settings with thousands of clients and real-world deployment studies on wearable devices remain essential for validating practical feasibility. Second, formal privacy analysis including differential privacy guarantees and resistance to adversarial attacks such as model poisoning and membership inference requires further investigation to ensure robust security for sensitive medical applications.

REFERENCES

[1] C. Wang et al., "Trustworthy health monitoring based on distributed wearable electronics with edge intelligence," *IEEE Trans. Consum. Electron.*, vol. 70, no. 1, pp. 2333–2341, Feb. 2024.

[2] J. M. Peake, G. Kerr, and J. P. Sullivan, "A critical review of consumer wearables, mobile applications, and equipment for providing biofeedback, monitoring stress, and sleep in physically active populations," *Frontiers Physiol.*, vol. 9, p. 743, Jun. 2018.

[3] X. Wang, J. Lyu, J. D. Peter, and B.-G. Kim, "Privacy-preserving AI framework for 6G-enabled consumer electronics," *IEEE Trans. Consum. Electron.*, vol. 70, no. 1, pp. 3940–3950, Feb. 2024.

[4] W. J. Gordon, A. Landman, H. Zhang, and D. W. Bates, "Beyond validation: Getting health apps into clinical practice," *npj Digit. Med.*, vol. 3, no. 1, p. 14, Feb. 2020.

[5] M. Wang, J. Wu, T. Zhang, J. Wu, and G. Li, "Effective identity authentication based on multiattribute centers for secure government data sharing," *Tsinghua Sci. Technol.*, vol. 29, no. 3, pp. 736–752, Jun. 2024.

[6] X. Lv, S. Rani, S. Manimurugan, A. Slowik, and Y. Feng, "Quantum-inspired sensitive data measurement and secure transmission in 5G-enabled healthcare systems," *Tsinghua Sci. Technol.*, vol. 30, no. 1, pp. 456–478, Feb. 2025.

[7] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Trans. Intell. Syst. Technol. (TIST)*, vol. 10, no. 2, pp. 1–19, 2019.

[8] J. Lv, B.-G. Kim, B. D. Parameshachari, A. Slowik, and K. Li, "Large model-driven hyperscale healthcare data fusion analysis in complex multi-sensors," *Inf. Fusion*, vol. 115, Mar. 2025, Art. no. 102780.

[9] K. Hsieh, A. Phanishayee, O. Mutlu, and P. Gibbons, "The non-iid data quagmire of decentralized machine learning," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 4387–4398.

[10] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 53–65, Jan. 2018.

[11] A. K. Selvaraj, S. B. Prathiba, A. D. Kumar, R. Dhanalakshmi, T. R. Gadekallu, and G. Srivastava, "Co-training-based personalized federated learning with generative adversarial networks for enhanced mobile smart healthcare diagnosis," *IEEE Trans. Consum. Electron.*, vol. 70, no. 3, pp. 6131–6139, Aug. 2024.

[12] Y. Su, Q. Chi, B. Shen, B. Zhang, J. Zhao, and L. Chen, "Client selection for mobile edge computing-assisted hierarchical federated learning with non-independent and identically distributed data," in *Proc. 4th Int. Conf. Mach. Learn. Comput. Appl.*, Oct. 2023, pp. 473–478.

[13] P. Kairouz et al., "Advances and open problems in federated learning," *Found. Trends Mach. Learn.*, vol. 14, nos. 1–2, pp. 1–210, 2021.

[14] H. Wang, M. Yurochkin, Y. Sun, D. Papailiopoulos, and Y. Khazaeni, "Federated learning with matched averaging," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Apr. 2020, pp. 1–14.

[15] A. Y. Hannun et al., "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Med.*, vol. 25, no. 1, pp. 65–69, Jan. 2019.

[16] E. Choi, M. T. Bahadori, A. Schuetz, W. F. Stewart, and J. Sun, "Doctor AI: Predicting clinical events via recurrent neural networks," in *Proc. Mach. Learn. Healthcare Conf.*, Dec. 2015, pp. 301–318.

[17] A. İ. Tekkeşin, "Artificial intelligence in healthcare: Past, present and future," *Anatolian J. Cardiol.*, vol. 2, no. 4, pp. 1–12, 2019.

[18] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. 20th Int. Conf. Artif. Intell. Statist.*, 2017, pp. 1273–1282.

[19] A. Yazdinejad and J. D. Kong, "Breaking interprovincial data silos: How federated learning can unlock Canada's public health potential," *Available SSRN*, vol. 5, pp. 1–27, May 2025.

[20] X. Wang et al., "Generative adversarial privacy for multimedia analytics across the IoT-edge continuum," *IEEE Trans. Cloud Comput.*, vol. 12, no. 4, pp. 1260–1272, Oct. 2024.

[21] X. Wang, A. Shankar, K. Li, B. D. Parameshachari, and J. Lv, "Blockchain-enabled decentralized edge intelligence for trustworthy 6G consumer electronics," *IEEE Trans. Consum. Electron.*, vol. 70, no. 1, pp. 1214–1225, Feb. 2024.

[22] A. Yazdinejad, A. Dehghantanha, H. Karimipour, G. Srivastava, and R. M. Parizi, "A robust privacy-preserving federated learning model against model poisoning attacks," *IEEE Trans. Inf. Forensics Security*, vol. 19, pp. 6693–6708, 2024.

[23] Y. Chen, X. Qin, J. Wang, C. Yu, and W. Gao, "FedHealth: A federated transfer learning framework for wearable healthcare," *IEEE Intell. Syst.*, vol. 35, no. 4, pp. 83–93, Jul. 2020.

[24] M. G. Arivazhagan, V. Aggarwal, A. K. Singh, and S. Choudhary, "Federated learning with personalization layers," 2019, *arXiv:1912.00818*.

[25] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," *Proc. Mach. Learn. Syst.*, vol. 2, pp. 429–450, Mar. 2018.

[26] V. Kulkarni, M. Kulkarni, and A. Pant, "Survey of personalization techniques for federated learning," in *Proc. 4th World Conf. Smart Trends Syst., Security Sustain. (WorldS)*, 2020, pp. 794–797.

[27] D. Namakshenas, A. Yazdinejad, A. Dehghantanha, and G. Srivastava, "Federated quantum-based privacy-preserving threat detection model for consumer Internet of Things," *IEEE Trans. Consum. Electron.*, vol. 70, no. 3, pp. 5829–5838, Aug. 2024.

[28] J. Yoon, D. Jarrett, and M. V. der Schaar, "Time-series generative adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 5508–5518.

[29] F. Zhu, F. Ye, Y. Fu, Q. Liu, and B. Shen, "Electrocardiogram generation with a bidirectional LSTM-CNN generative adversarial network," *Sci. Rep.*, vol. 9, no. 1, p. 6734, May 2019.

[30] M. Rasouli, T. Sun, and R. Rajagopal, "FedGAN: Federated generative adversarial networks for distributed data," 2020, *arXiv:2006.07228*.

[31] M. Chen, N. Shlezinger, H. V. Poor, Y. C. Eldar, and S. Cui, "Communication-efficient federated learning," *Proc. Nat. Acad. Sci. USA*, vol. 118, no. 17, 2021, Art. no. 2024789118.

[32] A. Grammenos, R. Mendoza-Smith, J. Crowcroft, and C. Mascolo, "Federated principal component analysis," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, Dec. 2019, pp. 6453–6464.

[33] M. H. Mehmood, M. I. Khan, and A. Ibrahim, "Balancing privacy and accuracy: Federated learning with differential privacy for medical image data," in *Proc. 7th Int. Conf. Data Sci. Inf. Technol. (DSIT)*, Dec. 2024, pp. 1–6.

[34] C. T. Dinh, N. H. Tran, and T. D. Nguyen, "Personalized federated learning with Moreau envelopes," in *Proc. NIPS*, Dec. 2020, pp. 21394–21405.

[35] X. Li, M. Jiang, X. Zhang, M. Kamp, and Q. Dou, "FedBN: Federated learning on non-IID features via local batch normalization," 2021, *arXiv:2102.07623*.

[36] Q. Guo, Y. Qi, S. Qi, D. Wu, and Q. Li, "FedMCSA: Personalized federated learning via model components self-attention," *Neurocomputing*, vol. 560, Dec. 2023, Art. no. 126831.

[37] X.-C. Li, D.-C. Zhan, Y. Shao, B. Li, and S. Song, "FedPHP: Federated personalization with inherited private models," in *Proc. Eur. Conf. ECML PKDD*, Bilbao, Spain. Cham, Switzerland: Springer, Sep. 2021, pp. 587–602.

[38] J. Zhang and Y. Shi, "A personalized federated learning method based on clustering and knowledge distillation," *Electronics*, vol. 13, no. 5, p. 857, Feb. 2024.