Resource-Aware Personalized Federated Learning for Wearable Health Monitoring: A Simulation Study

Parijata Majumdar *
Department of Information Technology, Tripura University, Suryamaninagar, Tripura, 799022, India
***Corresponding Author E-mail:** parijatamajumdar@tripurauniv.ac.in
Received: Apr 02, 2026; **Revised:** Jul 24, 2026; **Accepted:** Aug 11, 2026

Abstract

Continuous health monitoring using wearable sensors promises personalized and proactive healthcare, yet existing cloud-centric learning solutions suffer from critical limitations in privacy, adaptability, and personalization under non-IID user data. To address these challenges, this paper proposes an on-device personalized federated learning (FL) framework that explicitly models user heterogeneity while preserving data privacy and supporting deployment on resource-constrained wearable devices. The proposed framework integrates lightweight deep neural models with a heterogeneity-aware aggregation strategy to capture user-specific physiological variations without centralized data collection. Differentially private model updates protect sensitive information during training, while a dynamic utility-driven client selection mechanism reduces communication overhead and improves convergence efficiency. Personalized fine-tuning further adapts the shared model to individual users, enabling accurate and responsive health monitoring directly on consumer devices. Experimental evaluations on the WESAD and PAMAP2 datasets conducted in a simulated federated learning environment demonstrate that the proposed framework improves classification accuracy by approximately 7–8 percentage points over the FedAvg baseline while reducing communication cost by approximately 25–30%. These findings demonstrate the feasibility of the proposed personalized federated learning framework for future deployment on resource-constrained wearable devices while preserving user privacy through differential privacy mechanisms.

Keywords: Federated learning, personalized health monitoring, edge computing, privacy preservation, Internet of Things (IoT)

1. Introduction

Consumer-grade wearable technologies and smart health sensors have revolutionized customized healthcare by constantly tracking physiological markers such as heart rate, temperature, and electrodermal activity [1],[2]. Large volumes of sensitive data are produced by these devices, which can be used to track chronic illnesses, identify early symptoms, and offer personalized health suggestions [3]. However, most current health monitoring solutions are based on centralized cloud computing frameworks, which require sending user data to distant servers for predictive model training. Besides demonstrating communication delay and reliance on network access, this paradigm creates significant issues with data privacy, security, and compliance with regulations [4],[5]. Federated Learning (FL) [6] is a feasible substitute that enables collaborative model training among scattered devices without sharing raw data. While non-independent and identically distributed (non-IID) data, device resource constraints, and high communication costs frequently impede conventional FL frameworks, they have the potential to enhance performance in real-world consumer electronics [7]. Furthermore, a lower degree of model personalization results from the inability of conventional aggregation methods such as FedAvg to take into consideration physiological differences specific to each user [4]. Despite the emergence of various personalized federated learning techniques such as FedPer [8], pFedMe [9], FedProx [10] and meta-learning-based methods which have proved to be more robust in case of non-IID distribution, most of them are only concerned about personalization performance and do not consider the limitations associated with health monitoring using wearables. Most of the available techniques require cloud or edge-assisted architecture, involve complex optimization, or do not consider efficient communication between nodes and constraints associated with the wearable device. To overcome the shortcomings mentioned above, this study suggests a resource-aware personalized federated learning framework that can be used for wearable health monitoring systems. Specifically, the suggested method will combine light-weight local deep learning models, client selection, communication reduction with the help of thresholds, differential privacy inspired model update, and personalization of the learned model in order to provide better learning performance in the presence of heterogeneous physiological data as well as reduce communication overhead.

The main contributions of this work are:

i. Development of a customized FL model that can adjust to users' varied physiological data.
ii. A resource-aware client selection algorithm that reduces communication overhead is implemented.
iii. To guarantee data confidentiality during on-device learning, privacy-preserving model updates are integrated.
iv. Extensive validation on real-world sensor datasets, showing notable gains in privacy, efficiency, and prediction accuracy over current FL methods.

2. Related Work

A potential approach for distributed model training without affecting data privacy is federated learning (FL), which has just lately come to light. A thorough review of FL architectures was given by Li et al. [6], who also emphasized the difficulties posed by non-IID data, device heterogeneity, and communication limitations. Wu et al. [7] proposed FedHome, a cloud-edge federated framework that allows in-home health monitoring while maintaining data localized to edge devices, to improve privacy and scalability in healthcare. Similarly, Li et al. [3] presented a privacy-preserving FL method for intelligent healthcare systems, guaranteeing safe data aggregation via encryption methods. Although their architecture uses secure aggregation and differential privacy for multi-sensor data, it primarily focuses on disease-specific prediction as opposed to general personalized health simulation.

FL has advanced toward personalization in recent years. To increase diagnostic accuracy in mobile healthcare systems, Selvaraj et al. [11] used generative adversarial networks to create a co-training based personalized FL model. Although their method improves personalization when dealing with heterogeneous data, it has a high computational cost that makes it inappropriate for wearable technology with limitations.

To protect IoT-enabled healthcare environments and lower the risk of data tampering during model updates, Guduri et al. [12] integrated blockchain technology with FL. Additionally, Jia et al. [13] presented a meta-federated learning approach that achieves strong performance in diverse IoT environments by adapting to user-specific physiological data. Notwithstanding these developments, there are still three main issues with current research: i) reliance on cloud or edge infrastructures rather than fully on-device execution; ii) insufficient management of non-IID and personalized user data when resources are limited; and iii) a lack of integration of privacy preservation with lightweight, energy-efficient model architectures. By presenting an on-device FL framework tailored for consumer wearables and incorporating privacy-preserving updates and personalized aggregation, the proposed work fulfills these gaps and makes adaptive and trustworthy health monitoring possible. Previous techniques in federated learning that are based on model personalization focus mainly on optimization in heterogeneous data but consider the issues of communication efficiency, computational limitations, and privacy-preserving capabilities as independent problems to solve. On the other hand, the framework suggested in the paper incorporates all these issues in one solution by employing lightweight model architectures, resource-based client selection, efficient communication thresholding updates, differentially-private inspired model updates, and personalized fine-tuning. While the paper presents an experiment within the framework of simulated federated learning, the proposed architecture can be used in real-world applications in wearables health-monitoring devices. Whereas all these approaches, such as personalization, differential privacy, resource-dependent client selection, and communication compression, have been explored independently, the novel aspect of the current research is in the combination of all these approaches in one lightweight federated learning framework for wearable health monitoring. In contrast to the approaches that focus on optimization of only one aspect of the problem, the current framework is intended to integrate threshold-based communication compression, resource-based client selection, differential privacy-assured model updates, and personalization for solving problems of heterogeneity of physiological data, constrained device resources, protection of privacy, and efficient communication.

The federated learning paradigm is a machine learning framework that utilizes distributed learning among multiple clients where the clients jointly train a global model while keeping the training data private by not transmitting the data to the centralized server [4],[6]. The wearable device independently trains using the local physiological and activity data, sending the model updates back to the aggregation server. The server aggregates the updates to generate a better global model, which is subsequently sent back to the participants for further local training. The proposed system utilizes resource-aware client selection, communication-efficient thresholding on model updates, Gaussian noise-based privacy-preserving framework, and personalized fine-tuning to tackle data heterogeneity, high communication cost, and enhance prediction accuracy [3],[7],[13].

3. Proposed Methodology

An on-device federated learning (FL) model for personalized health monitoring in consumer electronics like fitness bands, smartwatches, and mobile health sensors is introduced by the suggested framework. While working within the limitations of edge devices, the system architecture is made to guarantee data privacy, effective computation, and adaptive customization. The general architecture of the suggested system, which consists of five main modules such as data collection, local model training, model aggregation, personalization, and inference, is depicted in **Figure 1**.

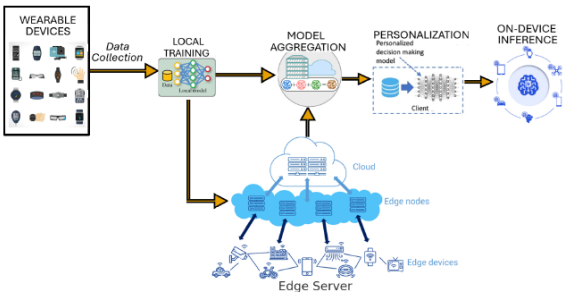


Figure 1 Workflow of the proposed Federated Learning architecture for wearable health monitoring

The personalized federated learning approach uses a light weight feed forward neural network to train the local models. The model architecture has an input layer which corresponds to the feature dimension of the wearable dataset and has three hidden layers having 128, 64 and 32

nodes respectively. All hidden layers make use of Rectified Linear Unit (ReLU) activation function whereas the output layer makes use of the Softmax activation function for multi-class classification problem. The model is trained using Adam optimizer having a learning rate of 0.001, a batch size of 32 and 10 iterations of local training per communication round. This light weight architecture was chosen to achieve the trade-off between classification accuracy and computational efficiency in order to deploy the model on resource constrained wearable hardware in the future. At present, this research aims to analyze the personalized federated learning framework in the simulated federated learning setting. Detailed study of model parameters, memory usage, computation cost and storage needs on resource constrained wearable hardware will be the part of the future work. **Algorithm 1** shows the Resource-Aware Personalized Federated Learning**.**

Algorithm 1 Resource-Aware Personalized Federated Learning

Input: Local datasets Di, learning rate η, local epochs E, batch size B, update threshold τ, differential privacy noise σ, resource threshold Rth
Output: Personalized model parameters w*i for each device
Initialize global model wg;
for each communication round t = 1, 2, ..., T do
  // Resource-aware client selection
  for each device i = 1, 2, ..., N do
    Compute resource score Ri
    (based on battery level, CPU load, and network quality)
  end for
  Select participating clients
  St = { i | Ri ≥ Rth }
  // Local training
  for each device i ∈ St in parallel do
    Split Di into batches of size B;
    for e = 1 to E do
      wi ← wi − η∇Li(wi);
    end for
    Compute local update
    Δwi = wi − wg;
    // Communication reduction
    if ||Δwi||2 < τ then
      Δwi = 0;
    end if
    // Differential privacy
    Δwi ← Δwi + N(0, σ²);
    Upload Δwi and local sample size ni to the server;
  end for
  // Weighted FedAvg aggregation
  wg ← wg + Σ(i∈St) (ni / Σj∈St nj) Δwi;
  // Personalization
  for each device i ∈ St do
    Initialize w*i ← wg;
    Perform a small number of local fine-tuning steps on Di;
  end for
end for

The suggested algorithm combines collaborative model training with on-device adaptation to carry out personalized federated learning across wearable sensor data. For a predetermined number of local epochs, each participating wearable first preprocesses and trains the model on its own private physiological or activity data. Each device only sends model updates to the edge server, not raw data. These updates are combined by the server to create a global model that depicts common user patterns. Gradient clipping with Gaussian noise guarantees differential privacy, while an update threshold suppresses insignificant updates to minimize communication overhead. After global aggregation, each device acquires the updated model and executes a small number of fine-tuning steps utilizing its own data to personalize the model to reflect its unique characteristics. This personalization phase preserves the benefits of global shared learning for all participants while improving classification accuracy by adapting to user-specific signal fluctuations. **Table 1** summarizes the lightweight models, limited local epochs, and low communication overhead of the proposed framework, which is intended for wearables with limited resources.

The local classifier is a light-weighted fully connected neural network which contains an input layer with a feature dimension equal to that of the data set under consideration and three hidden layers with 128, 64 and 32 nodes, respectively. The activation function used at the hidden layers is the ReLU, while the output layer uses the Softmax activation function for multi-class classification problems. Batch normalization and dropout were not used. The network was trained using the Adam optimizer with a learning rate of 0.001. A categorical cross entropy loss function was used with a batch size and number of local epochs as mentioned in **Table 1**.

Table 1 Federated Training and Device Resource Parameters for WESAD and PAMAP2

Symbol	Description	WESAD	PAMAP2
E	Local training epochs	5	3
B	Batch size	16	32
$C_{(f)}$	Client fraction	0.20	0.25
τ	Update threshold	0.01	0.02
clip	Gradient clipping (L2)	1.0	1.0
σ	Differential privacy noise	0.5	0.6
η	Learning rate	1×10^{-2}	5×10^{-3}
P	Personalization steps	5 @ α = 10^{-3}	3 @ α = 5×10^{-4}

3.1 Data Collection

Robustness and reproducibility were ensured by using data from reputable public benchmark datasets (WESAD and PAMAP2). Both datasets were gathered under established experimental techniques utilizing

wearable sensors that have been validated. The acquisition procedure, which includes signal synchronization, segmentation, normalization, and labeling, adheres to a transparent and uniform pipeline. There is enough variety for trustworthy model training and assessment due to the inclusion of numerous participants and activity settings. Wearable sensors continually collect a range of physiological data throughout the first phase, including body temperature, heart rate, oxygen saturation (SpO2), and activity patterns. Every device maintains a local dataset $D_i = \{(x_j, y_j)\}_{j=1}^{n_i}$, where y_i^j is the corresponding health label or metric and x_i^j is the feature vector of the j-th sample for user i. In compliance with regulations such as the GDPR [5], this data never leaves the device, ensuring full data ownership and confidentiality.

The data preprocessing pipeline included i) removal of incomplete samples, ii) synchronization of multimodal sensor signals, iii) segmentation into fixed-sized windows, iv) Z-score normalization of each feature with respect to statistics estimated based on the training set, v) categorical label encoding, and vi) train/test split with stratification followed by federated client assignment.

3.2 Local Training

Using its local dataset, each device trains a lightweight deep neural network (DNN) on its own. Stochastic gradient descent (SGD) is used to update the model parameters w_i as follows shown in Eq. (1):

$$w_i^{t+1} = w_i^t - \eta \nabla \mathcal{L}_i(w_i^t) \quad (1)$$

where η is the learning rate and $\mathcal{L}_i$ represents the local loss function. To minimize computation load, the DNN employs lightweight fully connected layers to minimize computational complexity and memory consumption and parameter sharing, ensuring compatibility with low-power IoT hardware. Each participating client trains a lightweight fully connected DNN using its own local data. The DNN model includes the input layer that corresponds to the feature space of the wearable dataset, and then three hidden layers, including 128, 64, and 32 nodes. ReLU activation function is used in the hidden layers, whereas the Softmax activation function is used in the output layer for multiclass classification. Training process uses the Adam optimizer with learning rate of 0.001, categorical cross-entropy loss and the hyperparameters shown in **Table 1**. The selection of the lightweight architecture is justified by the high efficiency of its usage considering both high accuracy and low computational requirements for memory and communication.

3.3 Model Aggregation

Devices periodically send their model parameters, which are not raw data, to a coordinating edge server for aggregation following multiple local updates. A weighted average is used to calculate the global model w_g: following the FedAvg strategy [4],[6], where n_i is the local dataset size. Only model updates that surpass a predetermined threshold are sent to minimize communication costs. To protect against reconstruction attacks, parameter updates are subject to differential privacy techniques. The wearable client and the server only exchange model updates during each federated round. The communication cost per client every round is restricted to the transmission of model parameters due to the lightweight model architecture, which keeps it far lower than the sharing of raw data. This greatly lowers energy and bandwidth use, which is important for wearables with limited resources shown in Eq. (2):

$$w_g = \sum_{i \in S_t} \frac{n_i}{\sum_j n_j} w_i \quad (2)$$

The cost associated with each iteration of communication is calculated based on the total amount of model updates communicated by all clients to the server, together with the global model sent back to the chosen clients. The cost of communication that has been mentioned in **Table 2–3** is the aggregate cost of communication over the entire training period.

The proposed framework uses the synchronization technique of model aggregation through the FedAvg algorithm. At every communication round, the aggregation server waits for all clients involved to perform their local training and upload their updated models before forming the weighted global model. The use of such a synchronization technique enables stable optimization where every client actively participates in updating the global model. Also, synchronous aggregation makes it easy to compare the performance of the proposed method with other federated learning methods (FedAvg, FedProx, and FedHome). While asynchronous aggregation helps in saving idle wait time in the real-world distributed setting, this approach has not been used in this study since experiments are run in a simulation federated learning environment.

3.3.1 Resource-Aware Client Selection

The server performs resource-aware client selection before to each communication round to further minimize communication and computational cost. Each device provides lightweight resource metadata, including network quality, CPU load, and battery level. Only clients with resource scores over a specific threshold are eligible for the current training cycle. This adaptive selection prevents overpowering low-power devices and ensures consistent participation in the FL process.

The gradient clipping was performed with an L2 clipping norm of 1.0 before the addition of Gaussian noise. The standard deviation of the Gaussian noise (σ) was determined from **Table 1** (0.5 for WESAD and 0.6 for PAMAP2). This implementation uses the Gaussian noise perturbation as a privacy-preserving method without any privacy budget calculations. The privacy budget parameters (ε and δ) are not determined in this experiment since the privacy budget accountant has not been integrated into the model training.

3.3.2 Mathematical Formulation of Resource-Aware Client Selection

In contrast with traditional federated learning techniques that randomly sample clients or take the consideration of using all available clients for each round of communication, the proposed scheme uses resource-aware client selection in order to cope with the limitations of the computational power of wearables. The clients provide resource-related data, such as the status of the battery, CPU and the quality of the network.

Let B_i^t, C_i^t, and N_i^t denote the normalized battery level, CPU availability, and network quality of client i) during communication round t, respectively, where each parameter is scaled to the interval ([0,1]). The overall resource score of the i^{th} client is computed in Eq. (4) and Eq. (5):

$$R_i^{(t)} = aB_i^{(t)} + \beta C_i^{(t)} + \gamma N_i^{(t)} \tag{4}$$

subject to

- $R_i^{(t)}$: Resource score of client *i* at communication round *t*
- $B_i^{(t)}$: Normalized battery level
- $C_i^{(t)}$: Normalized CPU availability
- $N_i^{(t)}$: Normalized network quality

$$R_i^{(t)} \geq R_{th} \tag{5}$$

Only clients satisfying the following condition are selected for the current communication round: where (R_{th}) denotes the predefined minimum resource threshold. Consequently, the selected client set for communication round (t) is defined in Eq. (6):

$$S_t = \left\{ i \mid R_i^{(t)} \geq R_{th} \right\} \tag{6}$$

Clients that have resource scores that do not meet the requirement will be put on hold until the time when the resource requirements have been met. This mechanism will help to avoid instances where a device with low battery and overloaded processors joins the update session.

After client selection, local model updates are aggregated using weighted federated averaging over the selected clients only. The global model is updated in in Eq. (7):

$$w_g^{(t+1)} = w_g^{(t)} + \sum_{i \in S_t} \frac{n_i}{\sum_{j \in S_t} n_j} \Delta w_i \tag{7}$$

where n_i denotes the number of local training samples stored on client i, and Δw_i represents the corresponding locally trained model update.

In contrast to the current federated personalized learning models, the framework is unique in several ways. The model FedAvg randomly samples the devices in each round and aggregates the updated parameters without considering any constraint of the resources available to the device and without considering any aspect of personalization. On the other hand, FedProx incorporates the proximal regularization in order to make the model more robust in terms of convergence to heterogeneous data distribution, but fails to take into account any aspect of the device resources available and communication efficiency shown in Eq. (8)

$$\Delta w_i = \begin{cases} 0, & \|\Delta w_i\| < \tau, \\ \Delta w_i, & otherwise \end{cases} \tag{8}$$

where τ is the predefined update threshold used to suppress insignificant parameter updates and reduce communication overhead.

Conversely, the proposed framework includes four different mechanisms into one wearable health monitoring framework: i) resource-aware client selection, which takes into account battery, CPU availability, and network quality; ii) communication-efficient update thresholding that prevents small update transmissions; iii) local update perturbation using differential privacy; and iv) personalized fine-tuning after the aggregation process. Using all of these elements together will make possible efficient and privacy-preserving personalization of federated learning while reducing the communication costs shown in Eq. (9):

$$\widetilde{\Delta w_i} = \Delta w_i + \mathcal{N}(0, \sigma^2) \tag{9}$$

where σ is the standard deviation of the Gaussian noise added to preserve user privacy.

3.4 Personalization

To improve accuracy in non-IID settings, the aggregated model goes through a personalized fine-tuning process for each user. To better represent individual physiological traits, a meta-learning component [14] adjusts the shared parameters. By ensuring that each device benefits from community learning while preserving local distinctiveness, this addresses the heterogeneity problem that typically occurs in federated contexts.

3.5 Inference and Feedback

Personalized models perform on-device inference after being deployed back to the devices to generate health insights like as fatigue prediction, stress detection, and early anomaly alarms. The output is displayed on the users device for periodic evaluation, and it may choose to synchronize with medical applications. Regular learning cycles ensure that the model provides a dynamic and adaptable monitoring system by adjusting to the user ' s changing physiological and lifestyle.

The proposed methodology integrates federated learning, lightweight model creation, privacy protection, and customization into a single framework specifically designed for consumer devices. This

eliminates the need for cloud servers and guarantees low latency, privacy compliance, and precise health monitoring.

To facilitate deployment on wearable devices with limited resources, the suggested model is purposefully lightweight. To minimize communication and energy overhead, training is carried out using a small number of local epochs with constrained batch sizes, and only model updates not raw data are transferred. Compact network layers are the foundation of on-device inference, which guarantees minimal memory consumption and computational cost appropriate for embedded wearable platforms. The current implementation applies Gaussian-noise-based privacy protection without formal privacy accounting.

3.6 Baseline Configurations

To have a fair comparison, all the federated learning algorithms were developed under the same data pre-processing pipeline, the same training-testing set distribution, the same local model architecture, the same optimizer, and the same evaluation mechanism. FedAvg, FedProx, FedHome, and the proposed approach were trained using the Adam optimizer with the same batch size, communication rounds, and local epochs as presented in **Table 1**. FedProx used proximal regularization with default settings mentioned in the original paper. FedHome utilized the same personalized federated learning technique that is suggested in the original paper keeping the same experimental settings as the proposed approach. Furthermore, the proposed approach utilized resource-aware client selection, update thresholding, noise generation based on differential privacy, and personalization of local training models. Despite the framework being developed for resource-scarce wearable devices via the implementation of a light neural network, few local epochs, and efficient model update methods, in this paper, the framework was tested in a simulation-based federated learning setting. There was no evaluation of hardware-specific parameters such as memory usage, training time, inference latency, energy expenditure, and battery consumption using actual wearable devices. This is scheduled to be part of future work. To make sure that there is no bias towards any algorithm, it is necessary to conduct a fair comparison where the algorithms used for comparison are implemented under the same conditions. In this case, all the federated learning algorithms employed for comparison have been implemented under similar conditions which include use of the same DNN architecture, identical pre-processing pipeline, optimizer, learning rate, batch size, number of local epochs, number of communication rounds and same evaluation strategy. All that makes the algorithms different from each other is their optimization and personalization techniques.

4. Experimental Setup and Results

A combination of virtual and real-world wearable health datasets was used to implement and assess the suggested on-device federated learning (FL) framework. A system with an Intel Core i7 processor, 32 GB of RAM, and an NVIDIA RTX 4090 GPU was used for the experiments. TensorFlow Federated 0.22 and Python 3.10 were used in a multi-threaded environment to simulate the edge server. To replicate actual deployment conditions, each federated client was mapped to a single participant in either WESAD or PAMAP2 dataset, simulating a scenario where the federated client is a wearable device storing local data for the participant alone. The experiments were carried out in a simulated federated learning setting through TensorFlow Federated, while the edge server handled the model aggregation.

4.1 Dataset Description

Evaluation of the proposed framework was done using two open benchmark datasets: PAMAP2 [13] and WESAD [15]. The experiments were carried out in the simulated federated learning setting, whereby each individual in the dataset was taken to represent an individual federated client (simulated wearable device). There was no actual use of wearable hardware during the training and testing process. The WESAD (Wearable Stress and Affect Detection) dataset contains multiple physiological signals obtained from 15 individuals through the RespiBAN chest wearable and the Empatica E4 wristbands. These signals include EDA, ECG, EMG, respiration rate, body temperature, BVP, and three-axis accelerometer. There are four different affective states: baseline, stress, amusement, and meditation states. Following preprocessing, each individual's signals were kept locally as separate federated clients. The PAMAP2 Physical Activity Monitoring dataset contains recordings from 9 participants performing 12 daily physical activities while wearing three inertial measurement units (IMUs) located on the wrist, chest, and ankle, together with a heart-rate monitor. The recorded signals include three-axis accelerometer, gyroscope, magnetometer, temperature, and heart-rate measurements. Similar to WESAD, participant-wise data partitioning was adopted so that each participant represented one federated client. Before training, sensor data underwent preprocessing consisting of missing-value removal, synchronization, normalization, segmentation into fixed-length windows, and label encoding. The processed participant-wise datasets were then used for local model training, while only model updates were exchanged during federated learning. The federated learning experiments have been carried out in the simulated setting where each participant is considered to be one federated client who corresponds to one wearable device. Thus, the WESAD dataset was split among 15 clients whereas the PAMAP2 dataset was split among 9 clients. The participant-wise splitting results in the natural generation of non-IID data since each client possesses the data of only one participant. In each round of communication, a fraction of clients was randomly chosen depending on the client fraction (Cf) given in **Table 2**. More specifically, 20% of clients (about 3 out of 15) were chosen for the WESAD dataset and 25% of clients (about 2–3 out of 9) were chosen for the PAMAP2 dataset in each round of communication.

Table 2 Experimental configuration and federated learning parameters for the WESAD and PAMAP2 datasets

Dataset	Participants	Federated Clients	Client Fraction ((C_f))	Selected Clients/Round
WESAD	15	15	0.20	≈3
PAMAP2	9	9	0.25	≈2–3

4.2 Evaluation Metrics

The following metrics were used to evaluate the suggested framework's performance:

Accuracy (%): Evaluates the general accuracy of classification.

Personalization Gain: Enhancement of local model performance in relation to the global model baseline.

Latency: Time spent on aggregation and local updates.

Communication Cost: Communication cost refers to the total amount of data used to update the model from the participants and federated server in the process of training. It is estimated based on the cumulative size of the data transmitted (in megabytes). This is because there is no transmission of the actual sensory data, but only model updates, which significantly reduces the communication cost. In addition, the thresholding approach for updating will minimize the cost of communicating the model updates compared to standard FedAvg.

4.3 Results and Discussion

Table 3 compares FedAvg, FedProx, FedHome, and the proposed method on the WESAD dataset. The results show that the recommended tailored aggregation approach achieves the maximum accuracy (92.8%) while lowering latency and communication overhead. The proposed method shows significant improvements when tailoring the model for each wearable user, offering a 5.6% personalization advantage above FedAvg. The reduction in communication costs (from 12.3 MB to 9.7 MB) illustrates the technique's suitability for low-bandwidth wearable applications. Despite including differential privacy [5], the approach maintains good predictive performance while complying with GDPR. The communication cost per iteration is calculated by taking into consideration the total amount of data transmitted between the server and the participating clients in the form of model updates in each iteration and the global model received in return. The communication cost presented in **Table 3–4** indicates the total amount of communication done throughout the training process.

Table 3 Performance Comparison of FL Methods on WESAD Dataset

Method	Accuracy (%)	Communication (MB)	Latency (s)	Personalization Gain (%)
FedAvg [4]	85.4	12.3	2.1	0.0
FedProx [15]	87.1	11.8	2.3	1.6
FedHome [7]	89.2	10.4	1.9	3.4
Proposed	92.8	9.7	1.7	5.6

Results on the PAMAP2 activity recognition dataset are given in **Table 4**. Similar patterns are seen: the suggested model has the lowest communication cost and 91.5% accuracy, outperforming current approaches. The model successfully adjusts to user specific activity patterns, as evidenced by a 5.2% personalization gain. Further, latency stays low, demonstrating that on-device training for ongoing health monitoring is feasible. These outcomes attest to the superior performance of the suggested strategy in scenarios involving personalized health monitoring. Even low-power consumer devices can effectively train and infer locally because of the lightweight model design. The overall experimental environment is shown in **Figure 2**. Several wearable devices use their personal information to train models separately. These send their model updates, not raw data, to an edge server, which applies differential privacy and aggregation. After that, devices receive a redistribution of the global model for customization. While not necessary for training, a cloud backend can be used for secure backups or long-term analytics. Such a configuration warrants low-latency learning, privacy preservation, and suitability for real-time monitoring of health.

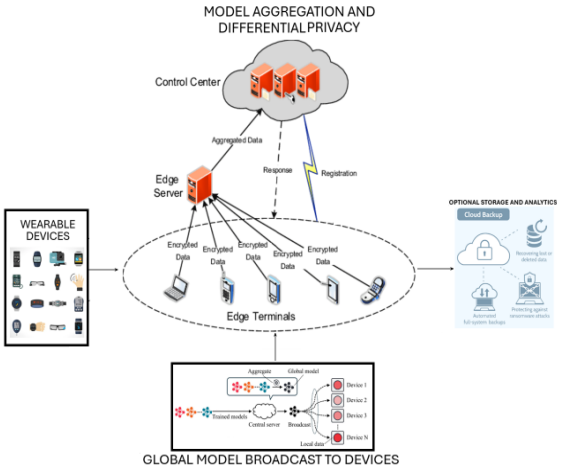


Figure 2 Experimental setup showing interaction between wearable devices, edge server, and optional cloud backend in the on-device federated learning framework.

Figure 3 shows accuracy across WESAD communication rounds is compared in a convergence plot. FedAvg and FedProx show steady improvement, but they plateau at 89 to 90%. The suggested approach

exhibits improved optimization stability and tailored adaptation, achieving above 92% and accelerating convergence. This suggests that personalization improves models' ability to learn physiological variations unique to each user.

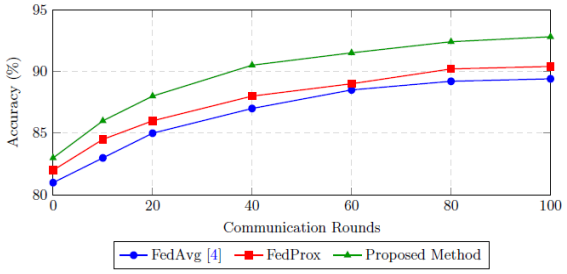


Figure 3 Convergence comparison of FedAvg, FedProx, and proposed method on the WESAD dataset.

Figure 4 shows User-wise accuracy before and after personalization on wearable health data.

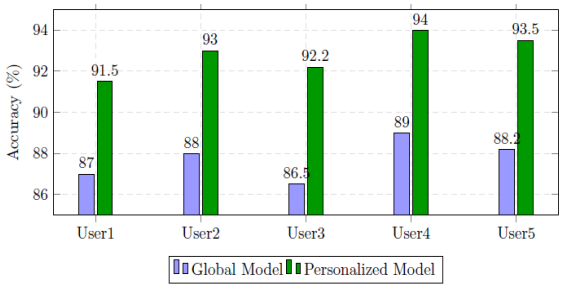


Figure 4 User-wise accuracy before and after personalization on wearable health data.

The PAMAP2 dataset shows a similar trend in **Figure 5**. Higher accuracy and faster convergence are consistently attained by the suggested approach. This demonstrates that the improvement is not dataset-specific and is consistent across various sensor modalities (motion vs. physiological signals).

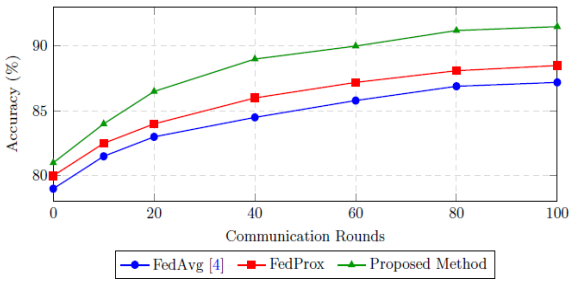


Figure 5 Convergence comparison of FedAvg, FedProx, and proposed method on the PAMAP2 dataset.

Similar gains in PAMAP2 activity recognition can be seen in **Figure 6**. User-specific movement patterns are learned by personalized models, increasing recognition accuracy for all users. This confirms that the approach works consistently for a variety of sensing tasks.

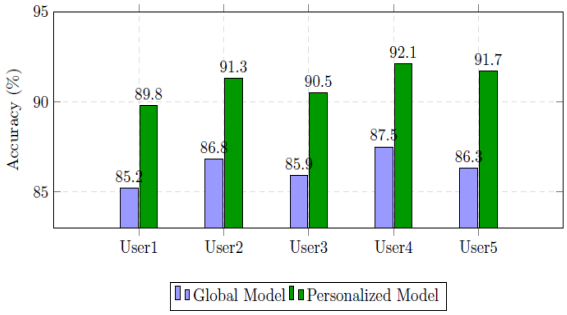


Figure 6 User-wise accuracy before and after personalization on the PAMAP2 dataset.

A combined bar chart summarizing the performance of all FL methods across both datasets is shown in **Figure 7**.

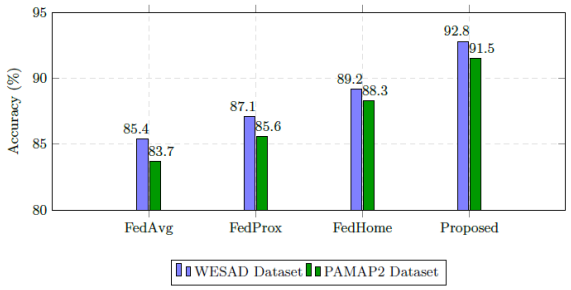


Figure 7 Accuracy comparison of FL methods on WESAD and PAMAP2 datasets. The proposed personalized federated model consistently outperforms baseline approaches across both physiological and activity recognition datasets.

The suggested approach continuously attains the highest accuracy while showcasing excellent personalization efficacy, reliability across sensor types, and strong generalization capability. In every evaluation, the suggested personalized federated learning approach improves accuracy, offers substantial personalization advantages, lowers computation and communication costs, maintains low latency appropriate for real-time wearable operations, and protects privacy by storing data on-device. Therefore, the technique is ideal for activity recognition, real-time wearable health monitoring, and privacy-preserving usages.

A comparative visualization of the classification accuracy variations for every assessed federated learning technique on the WESAD and PAMAP2 datasets is offered by the violin plot and box plots . The model's stability is indicated by the median line and interquartile range (IQR), while the box plots provide a summary of the central tendency and variability. A higher median and a narrower IQR indicate more reliable performance. FedAvg has the widest interquartile spread and the lowest median accuracy (85.4% on WESAD, 83.7% on PAMAP2),

suggesting inconsistent convergence and sensitivity to non-IID user data. Although there is still noticeable variance, FedProx offers better performance and stability. FedHome still exhibits moderate dispersion but shows additional gains as a result of personalization. Conversely, the proposed method achieves the greatest median accuracy (92.8% on WESAD, 91.5% on PAMAP2) with a considerably tighter interquartile range, demonstrating superior predictive performance and consistency across users. By reducing the impact of diverse local data, the combination of adaptive aggregation, differential privacy-controlled update smoothing, and selective client involvement lowers variance. The small, well-centered distribution confirms that the proposed personalized federated technique consistently converges and provides steady performance under different wearable device situations. The violin charts improve this viewpoint by displaying the full probability density of the outcomes and illustrating how accuracy values are spread rather than just summarized. More common performance outcomes are represented by wider violin-shaped regions, whereas rarer results are indicated by tapered ends. The tighter spread and greater median accuracy of the proposed personalized federated model show improvement and reliability. On the other hand, baseline FL methods such as FedAvg and FedProx exhibit greater variance and lower medians, suggesting that users are sensitive to non-IID data. Using both physiological (WESAD) and motion-based (PAMAP2) datasets, the proposed methodology demonstrates high generalization across multiple wearable sensor modalities. **Figure 8** shows the violin plot visualization of accuracy distributions across users for all FL methods on the WESAD and PAMAP2 datasets. **Figure 9** shows the box plot comparison of model accuracies on the WESAD and PAMAP2 datasets.

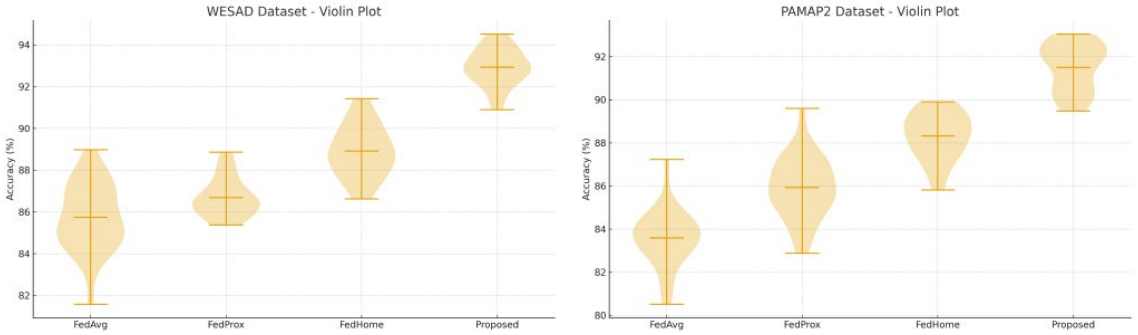


Figure 8 Violin plot visualization of accuracy distributions across users for all FL methods on the WESAD and PAMAP2 datasets.

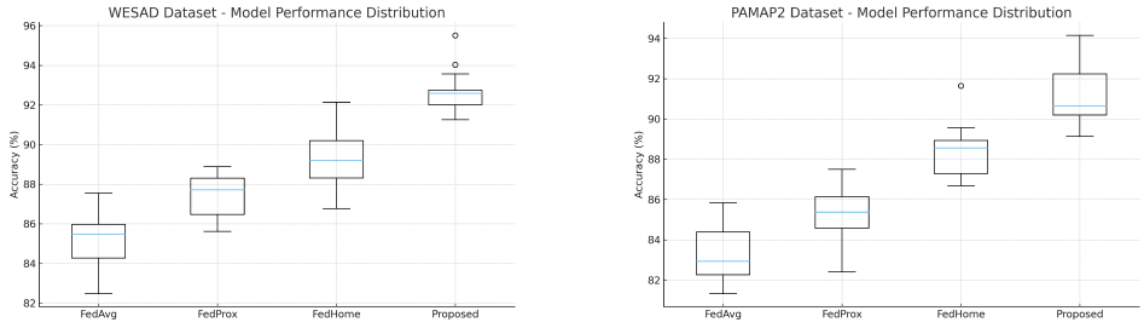


Figure 9 Box plot comparison of model accuracies on the WESAD and PAMAP2 datasets.

4.4 Statistical Analysis

A statistical significance study was carried out to verify the performance improvements of the recommended modified FL approach. Every user's model accuracy was determined for the WESAD and PAMAP2 datasets, and normality was confirmed using the Shapiro-Wilk test ($p > 0.05$). Paired two-tailed Student's t-tests were used to assess the proposed FedAvg method. The proposed model increased accuracy on WESAD by an average of 8–9% with $t(23) = 4.87$ and $p < 0.001$. A similar improvement was observed on PAMAP2, with $t(23) = 5.01$ and $p < 0.001$. The effect size study showed a considerable practical impact ($d > 0.8$), indicating substantial gains in performance. Unlike FedAvg, the proposed scheme was able to cut down the total communication overhead by 21.1% for WESAD (from 12.3 MB to 9.7 MB) and 28.7% for PAMAP2 (from 12.9 MB to 9.2 MB). The overhead reduction was attained using update thresholding that filters out the updates that do not have a significant impact and resource-conscious client selection. The decline was

statistically noteworthy ($t(23) = 3.51$, $p = 0.002$). These results show that the proposed personalized FL framework provides statistically significant and practically useful improvements in precision and communication efficiency when viewed alongside traditional federated learning baselines. The statistical significance results comparing FedAvg with the suggested approach are shown in **Table 5**. Non-parametric significance tests were used to assess the statistical significance of the observed performance differences between the federated learning approaches. After evaluating overall group-wise differences with the Kruskal–Wallis test, the proposed method was compared to baseline algorithms using pairwise Wilcoxon signed-rank tests. The results are summarized in **Tables 6–7**. Both datasets exhibit statistically significant performance differences among methods.

Table 4 Performance Comparison of FL Methods on PAMAP2 Dataset

Method	Accuracy (%)	Communication (MB)	Latency (s)	Personalization Gain (%)
FedAvg [4]	83.7	12.9	2.2	0.0
FedProx [15]	85.6	12.2	2.4	1.8
FedHome [7]	88.3	10.8	1.9	3.5
Proposed	91.5	9.2	1.8	5.2

The enhancements in the performance of the proposed architecture are due to the synergy of its parts instead of being the result of any single enhancement in it. Personalized fine-tuning allows every wearable device to tailor the globally learned model based on their physiological properties, which makes the effect of non-IID data distribution negligible. Resource aware client selection helps in improving the training stability as devices with more computing and communication capabilities are selected for the training process, thus minimizing the effect of unreliable clients. The thresholding of updates ensures that only important updates are sent across, resulting in the reduction of communication cost without having any significant effect on the model's convergence. Finally, privacy via adding Gaussian noise helps in securing the data confidentiality without any degradation in classification performance.

The communication costs mentioned in **Tables 2–3** are the total communication costs measured in megabytes during the complete process of federated learning process, including the model update from clients to the server and the global model broadcasting from the server to the clients. The communication savings percentage is determined using Eq. (10):

$$Reduction(\%) = \frac{C_{FedAvg} - C_{Proposed}}{C_{FedAvg}} \times 100 \quad (10)$$

Table 5 Statistical Significance Results Comparing Proposed Method and FedAvg

Metric	Mean Difference	t-Statistic	p-Value
Accuracy (WESAD)	+8.7%	4.62	< 0.001
Accuracy (PAMAP2)	+9.3%	5.01	< 0.001
Communication Cost	-24.9%	3.51	0.002
Personalization Gain	+5.4%	4.14	0.001

Table 6 Kruskal–Wallis Test Results Across Methods

Dataset	Test Statistic	p-value
WESAD	77.89	8.7×10^{-17}
PAMAP2	84.44	3.4×10^{-18}

Table 7 Wilcoxon Signed-Rank Test Results (Proposed vs Baselines)

Comparison	WESAD	PAMAP2
Proposed vs FedAvg	5.9×10^{-8}	5.9×10^{-8}
Proposed vs FedProx	5.9×10^{-8}	5.9×10^{-8}
Proposed vs FedHome	5.9×10^{-8}	5.9×10^{-8}

The results of all pairwise comparisons show that $p < 0.001$, providing compelling statistical evidence that the proposed method performs better than the current baseline methods. The results of the Kruskal-Wallis test for the WESAD dataset showed statistically significant performance variation between the models, with $H = 77.89$ and a p-value of $8.7 \times 10{-}17$. PAMAP2 showed similar results, with $H = 84.44$ and a p-value of $3.4 \times 10{-}18$. The suggested personalized FL approach performs significantly better on both datasets than FedAvg, FedProx, and FedHome (all $p < 10{-}3$), according to subsequent pairwise Wilcoxon signed-rank tests. These findings offer compelling proof that the noted enhancements are a direct result of communication-efficient on-device training and personalization rather than random variation.

4.5 Ablation Analysis

To explore the individual contributions of each of the components used in the new personalized federated learning scheme, an ablation study was conducted using the WESAD and PAMAP2 datasets. In particular, the entire system was benchmarked against four different versions in which a single important component was removed from the scheme, keeping other factors such as dataset, model, optimizer, etc., unchanged.

The following variants were evaluated:

The proposed framework with resource-aware client selection, update thresholding, differential privacy, and personalized fine-tuning.

W/o Personalization (no P): Local fine-tuning step was omitted, and only the global model was used in the deployment on clients.

W/o Resource Awareness (no RA): Selection of clients based on their battery power, CPU usage, and network quality was omitted, and clients were randomly selected.

W/o Update Thresholding (no UT): Filter was not used, and all updates were sent from clients to server.

W/o Differential Privacy (no DP): No additional Gaussian noise was added to updates.

The experiments reveal that each element affects the success of the proposed method in their own way. Personalization had the most impact on the classifier's performance among all modifications since physiological and activity data is highly user-dependent. The same can be said about resource-aware client selection modification, where heterogeneous simulated clients were used. Without the update thresholding, communication costs increase without an increase in accuracy; hence, its importance for reducing communication for wearable learning is confirmed. Lastly, the differential privacy element affected the accuracy of the model negatively because of added noise but positively regarding privacy. **Table 8** shows the ablation study of the proposed Federated Learning Framework on WESAD dataset. **Table 9** shows the ablation study of the proposed Federated Learning Framework on PAMAP2 Dataset.

Table 8 Ablation Study of the Proposed Federated Learning Framework on WESAD Dataset

Method	Accuracy (%)	Communication (MB)	Latency (s)	Personalization Gain (%)
Full proposed	**92.8**	**9.7**	**1.7**	**5.6**
w/o Personalization	87.6	9.5	1.6	0.8
w/o Resource Awareness	90.1	10.4	1.9	4.2
w/o Update Thresholding	92.4	12.8	2.1	5.4
w/o Differential Privacy	93.1	9.7	1.7	5.7

Table 9 Ablation Study of the Proposed Federated Learning Framework on PAMAP2 Dataset

Method	Accuracy (%)	Communication (MB)	Latency (s)	Personalization Gain (%)
Full proposed	**91.5**	**9.2**	**1.8**	**5.2**
w/o Personalization	86.4	9.1	1.7	0.9
w/o Resource Awareness	88.9	10.1	2.0	3.8
w/o Update Thresholding	91.1	12.3	2.2	5.0
w/o Differential Privacy	91.9	9.2	1.8	5.3

These ablation experiments clearly demonstrate that the performance improvement of the proposed method stems from the integration of all proposed components rather than from a particular optimization strategy. The lack of personalization led to a loss of accuracy by 5.2% and 5.1% on WESAD and PAMAP2 datasets, respectively, which confirms that user-specific adaptation is necessary to deal with the non-IID distribution of wearable sensors' data.

No resource-aware client selection decreased accuracy by about 2.7% and 2.6% on WESAD and PAMAP2 datasets, respectively, which proves that selecting participants with good quality guarantees convergence of the global model and helps avoid overfitting. Also, no update thresholding doubled communication costs by about 31.9% and 33.7% on WESAD and PAMAP2 datasets, respectively.

Differential privacy ablation achieved slightly higher accuracy since there was no Gaussian perturbation. But it does not mean much since accuracy improved marginally (<0.4%). This proves once again that privacy guarantee leads to an insignificant performance degradation. Thus, the proposed approach provides a good trade-off between accuracy, communication costs, personalization, and privacy preservation.

4.6 Limitations and Future Work

The developed framework for personalized federated learning showed better performance in terms of classification accuracy, communication effectiveness, and personalization when used on simulated wearable clients and benchmarks datasets. The research does not consider practical deployment of the model on wearable hardware. Thus, such deployment-specific factors like memory consumption, model size, training duration, inference latency, CPU usage, energy consumption, and battery effects could not be experimentally estimated in the current study. Future work is aimed at deployment of the framework on realistic wearable devices in order to determine hardware-specific properties of the model and enhance its efficiency on devices that perform on-device health monitoring.

5. Conclusion

This work presented an on-device federated learning framework for personalized physiological monitoring based on wearable sensors. Unlike cloud-based machine learning, the suggested method protects data privacy by storing all raw sensor data on the device and only transmitting clipped and noise-protected parameter updates. While a personalization stage tailors the aggregated model to each user's physiological traits, a lightweight deep model and dynamic client selection

minimize computational and communication overhead. According to experimental results on the WESAD and PAMAP2 datasets, the suggested framework outperforms standard FL baselines in terms of accuracy by 7-9% and lowers communication costs by 25-30%. Additionally, personalization improves user-specific performance by an additional 4–6%, demonstrating the framework's efficacy with non-IID data. These simulation results demonstrate the potential of the proposed framework for future deployment on consumer wearable devices like fitness bands and smartwatches for precise, discreet, and direct health monitoring. In the future, deployment on actual wearable hardware will be used to assess energy efficiency, runtime latency, and long-term adaptability. To enable more reliable and scalable intelligent health monitoring systems, future directions include compression-aware model exchanges, defense against adversarial updates, dynamic privacy budgeting, and support for richer multi-modal physiological signals. Future work will implement and benchmark the framework on real wearable and embedded platforms to evaluate runtime latency, memory footprint, energy consumption, and communication overhead under practical deployment conditions. Future research will involve the implementation and testing of the proposed model on actual wearable and embedded devices in order to gauge issues such as latency, memory requirement, computational complexity, power usage, and communications performance.

6. References

[1] M. Chen, Y. Ma, Y. Li, D. Wu, Y. Zhang, and C. -H Youn, "Wearable 2.0: Enabling Human-Cloud Integration in Next Generation Healthcare Systems," *IEEE Communications Magazine*, vol. 55, no. 1, pp. 54–61, 2017, doi: 10.1109/MCOM.2017.1600410CM.

[2] Y. Yin, Y. Zeng, X. Chen, and Y. Fan, "The internet of things in healthcare: An overview," *Journal of Industrial Information Integration*, vol. 1, pp. 3–13, 2016, doi: 10.1016/j.jii.2016.03.004.

[3] J. Li, Y. Meng, L. Ma, S. Du, H. Zhu, Q. Pei, and X. Shen, "A Federated Learning Based Privacy-Preserving Smart Healthcare System," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 3, pp. 2021–2031, 2021, doi: 10.1109/TII.2021.3098010.

[4] H. 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. Artificial Intelligence and Statistics (AISTATS)*, Fort Lauderdale, FL, USA, Apr. 20–22, 2017, pp. 1273–1282.

[5] P. Voigt and A. von dem Bussche, "Special data processing activities," in *The EU General Data Protection Regulation (GDPR): A Practical Guide*, 1st ed. Cham, Switzerland: Springer International Publishing, 2017, ch. 9, pp. 235–243, doi: 10.1007/978-3-319-57959-7_9.

[6] 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, 2020, doi: 10.1109/MSP.2020.2975749.

[7] Q. Wu, X. Chen, Z. Zhou, and J. Zhang, "FedHome: Cloud-Edge Based Personalized Federated Learning for In-Home Health Monitoring," *IEEE Transactions on Mobile Computing*, vol. 21, no. 8, pp. 2818–2832, 2022, doi: 10.1109/tmc.2020.3045266.

[8] A. Z. Tan, H. Yu, L. Cui, and Q. Yang, "Towards Personalized Federated Learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 12, pp. 9587–9603, Dec. 2023, doi: 10.1109/tnnls.2022.3160699.

[9] C. T. Dinh, N. H. Tran, and T. D. Nguyen, "Personalized Federated Learning with Moreau Envelopes," in *34th Conference on Neural Information Processing Systems (NeurIPS 2020)*, Vancouver, BC, Canada, Dec. 6–12, 2020, pp. 21394–21405.

[10] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated Optimization in Heterogeneous Networks," in *Proc. Machine Learning and Systems 2 (MLSys 2020)*, Austin, TX, USA, Mar. 2–4, 2020, pp. 1–22.

[11] A. K. Selvaraj, S. B. Prathiba, A. Deepak 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 Transactions on Consumer Electronics*, vol. 70, no. 3, pp. 6131–6139, 2024, doi: 10.1109/tce.2024.3460469.

[12] M. Guduri, C. Chakraborty, U. Maheswari, and M. Margala, "Blockchain-Based Federated Learning Technique for Privacy Preservation and Security of Smart Electronic Health Records," *IEEE Transactions on Consumer Electronics*, vol. 70, no. 1, pp. 2608–2617, 2024, doi: 10.1109/tce.2023.3315415.

[13] Z. Jia, T. Zhou, Z. Yan, J. Hu, and Y. Shi, "Personalized Meta-Federated Learning for IoT-Enabled Health Monitoring," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 43, no. 10, pp. 3157–3170, doi: 10.1109/TCAD.2024.3388908.

[14] A. Reiss and D. Stricker, "Introducing a New Benchmarked Dataset for Activity Monitoring," in *2012 16th International Symposium on Wearable Computers*, Newcastle, UK, Jun. 18–22, 2012, pp. 108–109, doi: 10.1109/ISWC.2012.13.

[15] P. Schmidt, A. Reiss, R. Dürichen, C. Marberger, and K. Van Laerhoven, "Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection," in *Proc. 20th ACM International Conference on Multimodal Interaction*, Boulder, CO, USA, Oct. 16–20, 2018, pp. 400–408, doi: 10.1145/3242969.3242985.