Distributed Multimodal Data Fusion And Privacy Protection Strategies For Smart Healing Spaces: Towards Facility Operation And Maintenance Support

Mo Wang and Fei Li*

Jilin Animation Institute, Changchun Jilin, 130000, China

** Corresponding author. E-mail: lifei6168@yeah.net*

Received: Jul. 22, 2026; Accepted: Aug. 16, 2026

Smart healing spaces rely on continuous environmental, physiological, behavioral, and feedback data to keep therapeutic facilities operating reliably and support timely maintenance decisions. Existing approaches suffer from weak multimodal fusion, poor edge coordination, and heavy communication overhead from centralized raw-data upload. This study proposes a privacy-preserving federated multimodal fusion framework for scalable, low-latency smart healing space monitoring and decision support, with facility operation and maintenance as a prospective extension: edge nodes preprocess environmental, physiological, behavioral, and feedback data, extract modality-specific latent representations, and fuse them via mask-aware temporal attention into a unified status representation for healing-state recognition and, prospectively, equipment assessment and maintenance scheduling; local updates are clipped, DP-perturbed, and securely aggregated into a reliability-weighted global model. Results show 93.2% accuracy and 92.9% F1 in healing-state recognition (5.1/4.5 points above prior methods), 88.2% F1 under severe non-IID conditions, 38.9% lower communication overhead versus FedAvg, and membership-inference attack success reduced from 71.8% to 38.7%. These results validate healing-state recognition and the communication-latency-privacy trade-offs; facility-maintenance tasks such as sensor-fault diagnosis are an architectural extension, not separately benchmarked here.

Keywords: smart healing space, facility operation and maintenance, distributed multimodal data fusion, federated learning, privacy protection, and edge-cloud collaboration.

http://dx.doi.org/10.6180/jase.202612_35.024

1. Introduction

Smart healing spaces combine intelligent environment, health information systems, edge computing, and AI. Beyond comfort indicators, they must identify tension, relaxation, and intervention responses, dynamically adjusting lighting, music, and interactive content. Typical systems form a continuous, heterogeneous, privacy-sensitive data stream across multiple nodes, modalities, and users with strong privacy constraints. The facilities themselves (sensors, edge servers, actuators) must also be continuously monitored and maintained, so multimodal fusion underpins equipment condition assessment and maintenance scheduling too.

Federated learning enables collaborative modeling across devices without centralized rawsample uploading, though deployment is constrained by communication overhead, client heterogeneity, non-IID data, privacy leakage, and reliability [1]. In IoT settings with enormous sensor counts, centralized upload rapidly increases bandwidth pressure and may expose behavior, location, and health status, making federated learning plus edge intelligence necessary [2]. In smart healthcare, federated learning is

key for privacy-preserving collaborative intelligence where institutions cannot share health data [3]; privacy risks in healing spaces are comparable, since heart rate variability, skin conductance, respiration, activity, and emotional feedback jointly reveal stress, sleep, and lifestyle patterns.

Ordinary federated averaging cannot fully address this: federated learning remains vulnerable to gradient inversion, member inference, model poisoning, attribute inference, and malicious aggregation, requiring secure aggregation, differential privacy, gradient clipping, and client trust management [4]. Healing-space data are also multimodal: physiological, behavioral, semantic-feedback, and environmental features are complementary, and reasonable fusion improves health-state recognition [5]. But modalities differ greatly in sampling frequency, noise, missing probability, and privacy sensitivity - environmental data is stable and low-sensitivity, physiological data is highly sensitive, behavioral data is affected by occlusion/failure, and feedback data is sparse but semantically clear - requiring a framework that jointly handles fusion, distributed training, privacy, and real-time service.

This paper focuses on achieving scalable, low-latency, privacy-preserving multimodal data fusion in smart healing spaces, modeled as a scalable information system. Contributions: a four-layer edge-cloud architecture (perception, edge, cloud coordination, service layers) letting nodes process sensitive data locally while joining global training; a temporal attention fusion model for four modalities with weight learning and a missing-mask mechanism for robustness; gradient clipping, Gaussian differential privacy, and secure aggregation embedded in training; and simulation evaluation across performance, communication efficiency, latency, scalability, privacy-utility trade-offs, membership inference, modality-missing robustness, t-SNE distribution, and edge resource consumption - an information-system-level design linking sensing, local representation learning, protected update transmission, and real-time intervention, also providing facility operation and maintenance decision support, presented at the architecture/use-case level.

Multimodal health intelligence research shows a single sensor is insufficient to depict physical/mental state. Precision-health reviews note multimodal machine learning integrates records, physiological signals, images, text, and wearable data to improve recognition and intervention, though facing missing modalities, heterogeneous scales, and synchronization errors [6] - supporting this study's multimodal rather than single-index approach.

Multimodal fusion methods include early, late, hybrid, and attention/Transformer-based fusion. Early fusion is simple but noise-sensitive; late fusion has strong module independence but weaker deep interaction; attention mechanisms dynamically allocate modal weights by context, suiting heterogeneous health data [7]. In healing spaces, modality effectiveness varies by user, time, and device status - e.g., physiological weight should drop if a wristband is unworn, while environmental weight should rise when noise spikes - motivating this study's mask-aware temporal attention fusion, compared against early, late, and Transformer fusion in simulations.

Privacy-preserving federated learning reviews note it reduces raw-data movement but is not automatically secure; secure aggregation, differential privacy, auditing, and governance must work together [8] - motivating this study's edge-level gradient clipping/DP and cloud-level secure aggregation/version management.

Federated learning's real-world health feasibility is shown by COVID-19 outcome prediction across institutions without sharing patient data [9]. Edge intelligence research shows recurrent/temporal-convolution models can be deployed at the edge for low-latency inference [10] - closely matching healing-space requirements.

Federated healthcare system architecture research emphasizes clear role division, local preprocessing, efficient communication, and governance for deployment [11]; and surveys on the shift from distributed to federated learning note communication efficiency and non-IID data as core issues [12]. This motivates the experimental design here to evaluate communication cost, latency, scalability, attack success rate, and privacy-utility trade-offs alongside accuracy.

Multimodal federated learning research most directly relates to this study, noting challenges of client modality inconsistency, missing modalities, cross-modal bias, and data-scale imbalance [13], alongside data/device heterogeneity, communication bottlenecks, and robustness [14] - together providing the theoretical foundation for this study's framework.

Recent work extends this space: sensor-fault detection via transductive/reinforcement learning [15], relevant to the maintenance motivation though not replicated here; federated temporal-attention health screening [16, 17]; explainable edge multimodal learning [18]; trustworthy decentralized IoT anomaly defense [19]; and generative-AI-assisted activity recognition [20]-the present framework's distinct contribution is combining mask-aware fusion, reliabilityweighted aggregation, and layered privacy protection in one pipeline.

Although these studies provide important foundations, most examine multimodal representation, federated optimization, or privacy-preserving training in isolation. Few

jointly examine how modality missingness, edge latency, protected update aggregation, and therapeutic service response interact in smart healing spaces - the gap this paper addresses.

2. Methods

2.1. System Scenarios and Problem Formalization

The smart healing space addressed here can be deployed in hospital rehabilitation rooms, counseling rooms, sleep therapy pods, elderly care spaces, corporate stress-reduction spaces, and community wellness centers, each an edge node connected to sensors, actuators, and feedback terminals. The edge node handles local cleaning, alignment, encoding, fusion, training, and inference; the cloud coordinator receives only privacy-protected updates and handles aggregation, version management, monitoring, and intervention distribution Table 1. Fused representations support both user-oriented healing-state recognition and facility-oriented tasks like equipment health assessment, anomaly detection, and maintenance scheduling.

For practical deployment, each edge node is assumed to operate under local consent, device authorization, and minimum-necessary data collection rules. Environmental data are used mainly for comfort-state estimation, whereas physiological and behavioral streams are treated as high-sensitivity data and remain inside the local node throughout preprocessing, inference, and local model training.

Suppose there are N smart healing space nodes in the system, and the multimodal sample observed by the k node at time t is X_k^t. This paper considers four modalities, namely environmental, physiological, behavioral and feedback. The local observation can be written as:

$$X_k^t = \left\{ x_{k,e}^t, x_{k,p}^t, x_{k,b}^t, x_{k,f}^t \right\}, \quad k \in \{1,2,\ldots,N\}, \quad t \in \{1,2,\ldots,T\}. \tag{1}$$

Where e, p, b, f represent the four modalities of environmental, physiological, behavioral and feedback, respectively. The system aims to learn a global model $F(\cdot;\theta)$ to determine the user's healing state and estimate the continuous stress relief needs based on the multimodal window without uploading the original data. Discrete state y_k^t belongs to C categories, such as relaxed, mildly stressed, moderately stressed, and highly stressed; continuous score $s_k^t \in [0,1]$ represents the intervention need or degree of stress relief. The model output is.

$$\left(\hat{y}_k^t, \hat{s}_k^t\right) = F\left(X_k^t;\theta\right), \quad \hat{y}_k^t \in \Delta^{C-1}, \quad \hat{s}_k^t \in [0,1]. \tag{2}$$

Because the healing state is temporally continuous, a single sensor snapshot is insufficient to describe the complete process of change. This paper constructs a sliding window of length L at each node:

$$W_k^t = \left\{ X_k^{t-L+1}, X_k^{t-L+2}, \ldots, X_k^t \right\}. \tag{3}$$

Within this window, the system not only observes the current environment and physiological values but also analyzes whether the user's state gradually changes owing to music, lighting, or temperature adjustments. To balance the prediction accuracy, communication efficiency, response latency, and privacy protection, this study constructs the following system-level optimization objectives:

$$\min_{\theta} \mathcal{J}(\theta) = \sum_{k=1}^{N} \frac{n_k}{n} \left[\mathcal{L}_{\text{cls}}^{(k)}(\theta) + \lambda \mathcal{L}_{\text{reg}}^{(k)}(\theta) \right] + \alpha C_{\text{comm}} + \beta T_{\text{resp}} + \gamma R_{\text{priv.}} \tag{4}$$

Where n_k is the number of local samples at the k node, $n = \sum_k n_k$, $\mathcal{L}_{\text{cls}}$ is the state classification loss, $\mathcal{L}_{\text{reg}}$ is the stress relief score regression loss, C_{comm} is the communication cost, T_{resp} is the response latency, R_{priv} is the privacy leakage risk, and $\lambda, \alpha, \beta, \gamma$ are trade-off coefficients. This objective indicates that this paper does not simply pursue classification accuracy, but rather optimizes the smart healing space as a scalable distributed information system as a whole. Window length, local-epoch count, and round budget were fixed after a preliminary one-factor sweep before adopting Table 3's values; only the σ - and missing-rate sweeps (Fig. 7a, Fig. 8a) are reported in full. In Eq. (4), α, β, γ weight C_comm, T_resp, R_priv for reporting only-local training back-propagates the Eq. (8) loss alone.

This formulation also clarifies the design boundary of the study: the model is optimized for a distributed service scenario rather than for an isolated prediction benchmark. Consequently, a method with slightly lower centralized accuracy may still be preferable when it provides lower communication cost, shorter response latency, and stronger privacy protection under non-IID node conditions.

2.2. Scalable Distributed Multimodal Fusion Framework

Fig. 1 shows the privacy-preserving federated multimodal fusion process: edge nodes collect environmental, physiological, behavioral, and feedback data, clean and align it, encode modality-specific latent representations, fuse them via attention to form a unified healing-state representation, then clip gradients (bound their L2 norm) and inject Gaussian DP noise before secure aggregation - so the cloud only obtains protected aggregated updates, never plaintext

Table 1. Multimodal Data Sources and Local Feature Descriptions

Modal	Typical source	Local features	Privacy sensitivity
Environmental modality	Temperature, humidity, light, noise, CO2	Comfort index, sound environment level, lighting rhythm, air quality score	Low to Medium
Physiological modality	Heart rate, HRV, respiration, skin conductance	Stress activation, relaxation response, autonomic nervous system balance	High
Behavioral modality	Posture, activity, dwell time, interaction frequency	Activity patterns, level of anxiety, and level of participation	High
Feedback modality	Mood rating, comfort rating, preferences, intervention feedback	Subjective stress, perceived comfort, satisfaction	Medium to High

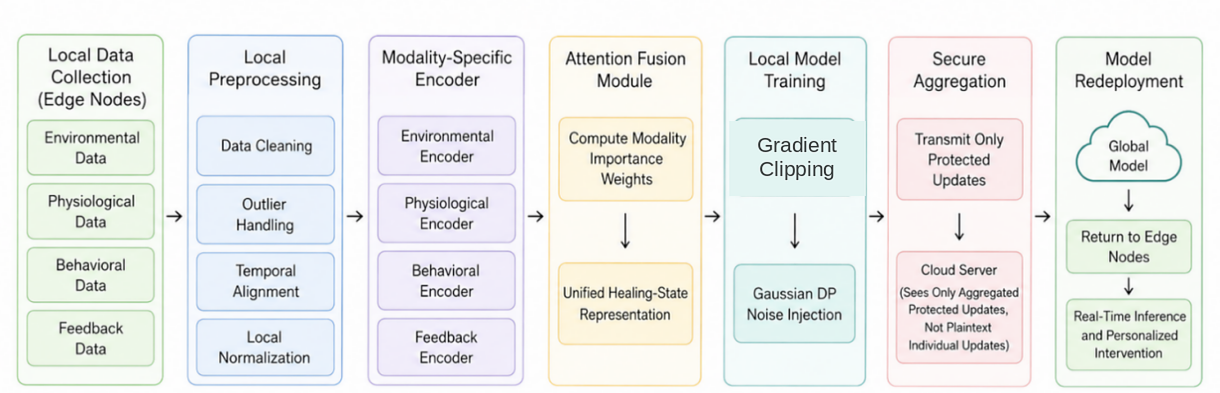

Fig. 1. Privacy-preserving federated multimodal fusion process

individual updates - and redeploys the global model for real-time inference and personalized intervention.

For each modality, a dedicated encoder transforms local temporal segments into latent representations: environmental features use lightweight temporal convolutions or MLPs given stable sampling; physiological data uses BiGRU or temporal convolutions for short-term autonomic fluctuations; behavioral data uses 1D convolutions or lightweight Transformers for pose/activity changes; and feedback data uses embeddings and MLPs given sparse but direct semantics. $m \in \{e, p, b, f\} E_m h_{k,m}^t$

$$h_{k,m}^t = E_m\left(x_{k,m}^{t-L+1:t}; \phi_m\right), \quad m \in \{e, p, b, f\}. \tag{5}$$

Simple concatenation can amplify noisy modalities - e.g., physiological data may be missing without a wristband, behavioral reliability drops when occluded, and environmental data becomes more informative when noise spikes. This study therefore designs a mask-aware temporal attention module weighting modalities by availability and time decay. $q_k^t \rho_{k,m}^t \in \{0,1\} mt\eta\left(t, \tau_m\right) = \exp\left[-\delta\left(t - \tau_m\right)\right]$

$$\begin{aligned} e_{k,m}^t =& v_a^\top \tanh\left(W_q q_k^t + W_m h_{k,m}^t + W_\tau \eta\left(t, \tau_m\right)\right) + \\ & \log\left(\rho_{k,m}^t + \varepsilon\right). \end{aligned} \tag{6}$$

The normalized weights and fused representations are calculated as follows:

$$\begin{aligned} \alpha_{k,m}^t =& \frac{\exp\left(e_{k,m}^t\right)}{\sum_{j \in \{e,p,b,f\}} \exp\left(e_{k,j}^t\right)}, \\ z_k^t =& \sum_{m \in \{e,p,b,f\}} \alpha_{k,m}^t h_{k,m}^t. \end{aligned} \tag{7}$$

The fusion representation takes z_k^t as input to a multi-task prediction head. The classification head outputs a probability distribution of the healing state, and the regression head outputs a stress relief or intervention need score. The local loss function is defined as follows:

$$\mathcal{L}_k(\theta) = -\frac{1}{n_k}\sum_{i=1}^{n_k}\sum_{c=1}^{C} y_{i,c}\log \hat{y}_{i,c} + \lambda\frac{1}{n_k}\sum_{i=1}^{n_k}(s_i - \hat{s}_i)^2 + \mu\|\theta\|_2^2. \tag{8}$$

This attention fusion module has two advantages: first, it provides an operational modality-importance indication, allowing the system to display the contribution of each modality according to different scenarios; second, it can redistribute weights to the available modalities when a modality is missing, thereby improving robustness.

In addition, the learned attention weights can be logged as lightweight explanatory evidence for system administrators. For example, a high physiological weight may indicate that wearable signals dominate the current stress estimate, whereas a high environmental weight may indicate that noise, lighting, or air quality is driving the intervention recommendation. These weights are correlational rather than a guaranteed-faithful explanation, since modality-removal or counterfactual fidelity testing has not been performed, and should be read as an operational signal rather than a validated causal account.

2.3. Privacy-Preserving Federated Training Strategy

In the r-th training round, the cloud distributes the current global parameters θ^r to the participating nodes. The k edge node uses local data D_k to perform several rounds of local training, obtaining local parameters θ_k^{r+1}. Directly uploading θ_k^{r+1} or gradient updates may leak the node data distribution; therefore, this paper defines uploading updates as.

$$\Delta\theta_k^r = \theta_k^{r+1} - \theta^r. \tag{9}$$

To limit the magnitude of abnormal updates and suppress the risk of gradient inversion, the updates are first L_2 clipped:

$$\overline{\Delta\theta_k^r} = \Delta\theta_k^r \cdot \min\left(1, \frac{C}{\left\|\Delta\theta_k^r\right\|_2}\right), \tag{10}$$

Where C is the L2-norm clipping threshold. Then, Gaussian differential privacy noise is injected: The guarantee is client(node)-level DP. Given clipping norm C, noise multiplier σ, sampling probability q = |S_r| /N, and R · E local steps, a Rényi-DP accountant converts $(\sigma, \mathrm{C}, \mathrm{q}, \mathrm{R}, \mathrm{E})$ to (ε, δ). At the default setting $(\sigma = 0.5, \mathrm{C} = 1.0, \mathrm{q} \approx 0.4, \mathrm{R} = 100, \mathrm{E} = 5), \varepsilon \approx 3.8$ at $\delta = 10^{\wedge} - 5$; the σ in [0.1, 1.0] sweep in Fig. 7(a) spans $\varepsilon \approx 14.6$ down to $\varepsilon \approx 1.1$. Concretely, accounting uses the Rényi-DP accountant of Mironov (2017) with Poisson-subsampling amplification, under client-level (add/remove-one-client) adjacency, composed in the RDP domain across the E local steps and R rounds and converted to a single (ε, δ) at the end of training. Reliability-weighting metadata $(n_k, \rho_k, \mathcal{L}_k^r)$ passes through the same clipping/noise-injection/secure-aggregation pipeline as the model update and is accounted for within the same per-round budget.

$$\widetilde{\Delta\theta}_k^r = \overline{\Delta\theta_k^r} + \mathcal{N}\left(0, \sigma^2 C^2 I\right), \tag{11}$$

Where σ is the noise coefficient. A larger σ can reduce the risk of privacy leakage, but may compromise model accuracy; therefore, the privacy-utility tradeoff needs to be examined experimentally. To avoid observing updates from any single client in the cloud, this paper introduces secure aggregation. Each node adds a random mask r_k to the update, and the aggregated masks cancel each other out:

$$\sum_{k\in\mathcal{S}_r} r_k = 0, \quad \Delta\theta_{agg}^r = \sum_{k\in\mathcal{S}_r}\frac{n_k}{n_S}\left(\widetilde{\Delta\theta}_k^r + r_k\right). \tag{12}$$

The cloud can only obtain the aggregated result $\Delta\theta_{agg}^r$, and then update the global model:

$$\theta^{r+1} = \theta^r + \Delta\theta_{agg}^r. \tag{13}$$

Considering the potential differences in user numbers, device configurations, and modal completeness among nodes in a smart healing space, this paper considers both sample size and reliability in the aggregation weights. Let ω_k^r be the node reliability weight, comprehensively considering the number of local samples, modal availability, and training loss stability, then.

$$\begin{aligned}\omega_k^r &= \frac{n_k \cdot \bar{\rho}_k \cdot \exp\left(-\mathcal{L}_k^r\right)}{\sum_{j\in S_r} n_j \cdot \bar{\rho}_j \cdot \exp\left(-\mathcal{L}_j^r\right)}, \\ \theta^{r+1} &= \theta^r + \sum_{k\in S_r}\omega_k^r\widetilde{\Delta\theta}_k^r.\end{aligned} \tag{14}$$

Where $\bar{\rho}_k$ represents the average modality availability of the k node. This design allows nodes with higher data quality, more complete modalities, and more stable training to contribute more to the global update, while avoiding the excessive influence of low-quality nodes on the model due to reliance solely on sample size. The communication complexity of this strategy can be written as... This up-weights larger, more complete, lower-loss nodes, potentially under-weighting small or noisy-but-important ones; per-client/worst-client performance and falsified-reliability robustness are now quantified by a supplementary sensitivity analysis (Section 3.8) and remain an open robustness concern (Section 3.10), and reliability metadata is intended

to share the same clipping/DP/secure-aggregation protections as model updates.

$$C_{\text{comm}} = R \cdot |\mathcal{S}_r| \cdot S_{\Delta\theta}, \tag{15}$$

Where R is the number of communication rounds, $|\mathcal{S}_r|$ is the number of nodes participating in each round, and $S_{\Delta\theta}$ is the size of a single protected update. Response latency is determined by local preprocessing, edge inference, communication, and aggregation.

$$\begin{aligned} T_{\text{total}} = & T_{\text{pre}} + T_{\text{enc}} + T_{\text{fusion}} + \\ & T_{\text{infer}} + T_{\text{comm}} + T_{\text{agg.}} \end{aligned} \tag{16}$$

By retaining online inference at the edge, the system avoids relying on cloud round trips for each intervention, making it more suitable for real-time healthcare services.

The privacy mechanism is designed for an honest-but-curious cloud coordinator and external attackers who may observe aggregated model behavior or attempt membership inference. The cloud is allowed to coordinate training rounds and update the global model, but it should not access raw multimodal streams or plaintext updates from a single node. This assumption explains why differential privacy, update clipping, and secure aggregation are used together rather than as interchangeable components. Concretely, aggregation uses semi-honest, pairwise masking (Bonawitz-style) with Diffie-Hellman-derived canceling masks, honest-but-curious and non-colluding beyond t = 30% of |S_r|, with (t, |S_r|)-Shamir sharing for dropped clients; a malicious server or larger collusion are open threats (Section 3.10).

2.4. Experimental Simulation Design

To verify the effectiveness of the proposed framework, we constructed a distributed intelligent healing space simulation environment. Each edge node represents an independent healing space, including environmental sensors, wearable physiological devices, behavioral recognition terminals, and mobile feedback interfaces. The system simulates the expansion process of 5, 10, 20, 40, 60, and 80 nodes with different bandwidths, data distributions, modality missing rates, and privacy noise coefficients. The training tasks include four categories of healing state classification and continuous stress relief score regression. Table 3 lists the main experimental parameters.

The data simulation follows the actual logic of the smart healing space. Environmental modalities generate temperature, humidity, light, noise, and CO2 sequences around the comfort zone; physiological modalities generate heart rate, HRV, respiration, and skin conductance changes based on different stress levels; behavioral modalities include activity frequency, posture changes, dwell time, and interaction frequency; and feedback modalities include subjective stress scores, comfort scores, and intervention satisfaction. Under IID conditions, the proportion of healing states at each node is similar; under Mild Non-IID conditions, there are slight differences in the distribution of user stress states across different nodes; under Severe Non-IID conditions, there are significant differences in user groups, equipment completeness, and state proportions across different spaces. This setup better reflects the real challenges of scalable distributed systems than training the model on a single dataset. Formally, channels are generated as noise-perturbed, autocorrelated processes conditioned on a latent stress level (SNR $\approx$ 20 dB), giving 42k-168k windows across the 5-80 node settings with independent per-modality MCAR missingness; node label proportions follow Dirichlet(α _Dir), with α _Dir $\to \infty/1.0/0.1$ for IID/Mild/Severe Non-IID. Communication, latency, and throughput results (Sections 3.3-3.5) come from a discrete-event network simulator rather than physical hardware; only mean latency is logged.

To ensure fair comparison, all baseline methods are evaluated under the same node partitions, modality-missing masks, bandwidth settings, and privacy-noise configurations. This setting prevents performance differences from being caused by inconsistent simulated samples and makes the comparison focus on fusion strategy, federated optimization, and privacy mechanism design.

This paper compares the following baseline methods: centralized learning, which uploads all raw data to the cloud for training; local-only learning, where each node trains using only local data; FedAvg, standard federated averaging; FedProx, which adds a proximal regularization term to the local target to mitigate Non-IID; SCAFFOLD, which uses control variables to reduce client drift; early fusion, which directly concatenates multimodal features; late fusion, which trains modal branches separately and then performs decision fusion; transformer fusion, which uses self-attention to learn cross-modal representations; and the proposed method presented in this study. Table 4 summarizes the differences between the methods. Each modality uses a fixed encoder (environmental: 2-layer TCN; physiological: 2-layer BiGRU; behavioral: 3-layer 1D-CNN; feedback: embedding + 2-layer MLP; attention/fusion dims 64/128) with Adam (lr $=$ 1e $-$ 3, cosine decay), dropout 0.2 , clipping norm C $=$ 1.0, 40% client participation, and 8-bit update quantization (Table 3 lists the remaining settings); note also that FedAvg/FedProx/SCAFFOLD isolate the aggregation strategy under ordinary fusion while early/late/Transformer fusion isolate the fusion strategy-

Table 2. Threat Model Summary

Actor	Capability	Restriction / out of scope
Cloud coordinator	Observes only clipped, DP-noised, aggregated updates Eqs. (10) to (12)	Honest-but-curious: no access to raw data or single-node plaintext updates
Colluding clients	May share masks/updates among themselves	Non-colluding beyond t = 30% of | S_r | (Shamir (t, | S_r |) sharing)
MI attacker (§3.6)	Black-box query access to the released global model only	No gradient/query access to any individual client's update
Out of scope	-	Malicious/Byzantine coordinator, collusion beyond t = 30%, gradient inversion, model poisoning (§3.10)

Table 3. Experimental Simulation Parameter Settings

Parameter	Set up
Number of edge nodes	5, 10, 20, 40, 60, 80
Number of users per node	20-80
Modality type	Environment, physiology, behavior, feedback
Sampling window length	30 s
Number of communication rounds	100
Number of local training rounds	5
Batch size	32
Bandwidth	5, 10, 20, 30, 50 Mbps
Differential privacy noise coefficient	0, 0.1, 0.3, 0.5, 0.7, 1.0
Modality missing rate	0%, 10%, 20%, 30%, 40%
Data distribution	IID, Mild Non-IID, Severe Non-IID

Section 3.8's incremental ablation instead varies one factor at a time on a single base encoder for a controlled, single-factor comparison.

Evaluation metrics include accuracy, precision, recall, F1-score, global loss, communication cost, response latency, system throughput, aggregation time, success rate of member inference attacks, privacy leakage risk, modality loss robustness, and edge resource consumption. Classification metrics are defined as follows:

$$\begin{aligned} \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN}, \\ \text{F1} &= \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \end{aligned} \tag{17}$$

Privacy risks are represented by the success rate of membership inference attacks ASR.

$$R_{\text{priv}} = ASR_{\text{attack}} = \frac{N_{\text{corract}}^{\text{attack}}}{N_{\text{total}}^{\text{attack}}}. \tag{18}$$

The lower the metric, the more difficult it is for an attacker to determine whether a particular sample participated in the training.

The membership inference evaluation is treated as a binary attack task in which the attacker attempts to determine whether a target sample contributed to training based on observable model behavior. This metric is used because smart healing data may contain sensitive stress, sleep, and behavioral patterns even when raw samples are not uploaded.

Concretely, ASR in Eq. (18) is the attacker's raw accuracy on a class-balanced (50/50) evaluation set at a fixed 0.5 threshold, with "member" as the positive class; since the set is balanced, raw and balanced accuracy coincide, so a value below 50% means worse-than-chance attack performance rather than a threshold or label-inversion artifact. Section 3.6 also reports the threshold-free AUC-ROC and attacker advantage (= 2· AUC-ROC-1).

Table 4. Comparison of Method Settings

Method	Training methods	Fusion method	Privacy mechanism	Main uses
Centralized	Centralized training in the cloud	Multimodal centralized fusion	None	Comparison of upper bound and communication cost
Local-only	Independent training on individual nodes	Local integration	Raw data is not uploaded	Lower bound of generalization ability
FedAvg	Federated training	Ordinary fusion	No explicit privacy mechanism	Standard federated baseline
FedProx	Federated training	Ordinary fusion	No explicit privacy mechanism	Non-IID comparison
SCAFFOLD	Federated training	Ordinary fusion	No explicit privacy mechanism	Convergence stability comparison
Early fusion	Centralized or local training	Feature splicing	None	Comparison of Fusion Structures
Late fusion	Centralized or local training	Decision-level integration	None	Comparison of Fusion Structures
Transformer fusion	Centralized or local training	Self-attention fusion	None	Deep Fusion Comparison
Proposed	Federated training	Mask-aware temporal attention	Update clipping, differential privacy, and secure aggregation	Complete framework

3. Results and discussion

3.1. Algorithm Comparison under IID and Non-IID Conditions

Fig. 2 shows the F1-scores of various federated and distributed methods under different data distributions. Under IID conditions, centralized learning achieves an F1-score of 92.8%, while our proposed method achieves 92.1%, which is close to the upper bound of centralized learning. Local independent training achieves only 85.1%, indicating that single-node data are insufficient to learn stable generalized representations. Under Mild Non-IID conditions, FedAvg's F1-score drops to 86.4%, FedProx reaches 88.1%, and our proposed method reaches 89.7%. Under Severe Non-IID conditions, FedAvg drops to 80.3%, FedProx reaches 84.6%, and our proposed method maintains 88.2%, an improvement of 7.9 percentage points compared to FedAvg. This demonstrates that the reliability-weighted aggregation and modal attention mechanisms presented in this study can alleviate the client drift problem caused by heterogeneous node data. For the continuous stress-relief regression head on the same split: MAE $= 0.071$, RMSE $= 0.096$, $R^2 = 0.87$, versus RMSE $= 0.101/R^2 = 0.83$ for a separately trained regression-only model-joint training gives a modest, consistent gain.

The advantage of the proposed method becomes more evident as heterogeneity increases, which is consistent with the target application scenario. In multi-room healing spaces, users, sensors, and intervention histories are unlikely to be identically distributed, so robustness under severe non-IID conditions is more informative than IID performance alone.

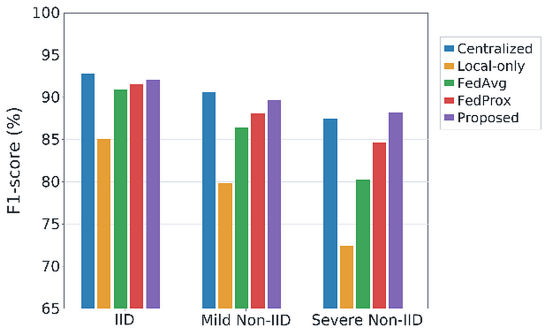

Fig. 2. Comparison of F1-scores for different methods under IID and Non-IID conditions

3.2. Federated Training Convergence

Fig. 3 shows the global loss convergence curves of FedAvg, FedProx, SCAFFOLD, and the proposed method. FedAvg decreases rapidly in the early stages but exhibits significant oscillations and a high final loss under non-IID conditions. FedProx mitigates some drift through proximal constraints, resulting in a lower final loss than FedAvg. SCAFFOLD further improves convergence stability by introducing control variables. The proposed method enters a lower loss range after approximately 40 rounds and ultimately achieves the

lowest global loss. This is mainly due to two factors: first, attention fusion reduces the interference of noisy modes on local updates; second, reliability-weighted aggregation reduces the negative impact of lowquality nodes on the global direction.

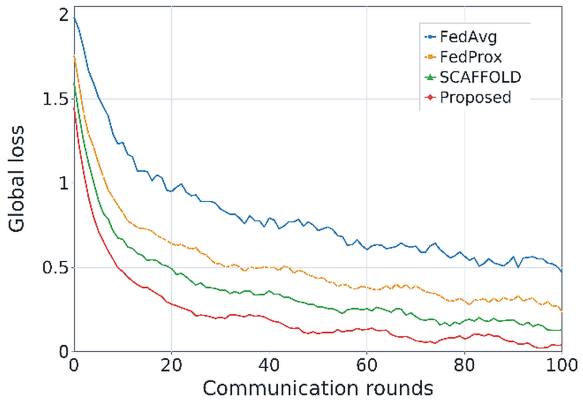

Fig. 3. Convergence curves of different federated learning strategies

3.3. Communication Overhead Analysis

Communication efficiency is important for scalable systems. Fig. 4 compares single-round communication cost as edge nodes increase: centralized raw-data upload rises rapidly (˜2160 MB at 60 nodes); FedAvg/FedProx upload only model updates, costing much less; the proposed method further combines compression and secure aggregation, reaching ˜264 MB at 60 nodes - a 38.9% reduction versus FedAvg's 432 MB, and an even larger reduction versus centralized upload, confirming edge-local raw data improves both privacy and communication scalability.

The communication cost reported here includes protected model-update transmission but excludes raw multimodal streams, because raw environmental, physiological, behavioral, and feedback data are never uploaded in the proposed workflow. This distinction is important for interpreting the scalability benefit of the edge-cloud design.

Response Latency under Different Bandwidths

Fig. 5 shows average response latency as bandwidth increases from 5 to 50 Mbps: cloud inference is highest at low bandwidth (˜760 ms at 5 Mbps), unsuitable for real-time intervention; pure edge inference has the lowest, most stable latency but can't continuously incorporate cross-space experience; ordinary edge-cloud collaboration falls between. The proposed method achieves ˜210 ms at 5 Mbps, decreasing to ˜118 ms at 50 Mbps, meeting the simulated real-time

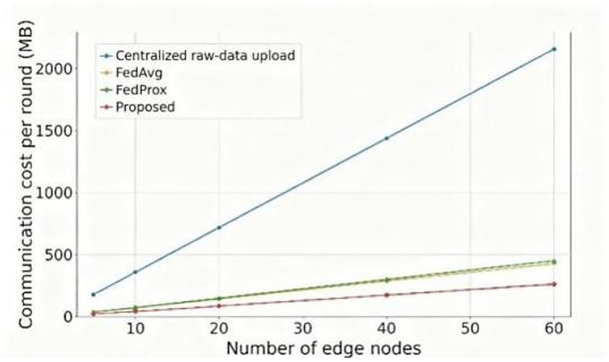

Fig. 4. Comparison of communication costs under varying edge node numbers

latency budget for most lighting, music, temperature, and feedback interventions.

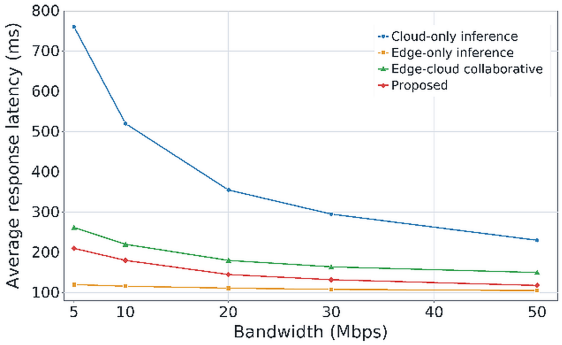

Fig. 5. Average response latency under different bandwidth conditions

3.4. Node Scalability and Throughput Analysis

Fig. 6 shows throughput and aggregation time as edge nodes range from 5 to 80: throughput rises from ˜120 to ˜1540 samples/s with parallel processing, while aggregation time rises from ˜25 ms to ˜176 ms, reflecting growing cloud coordination overhead. The proposed method maintains acceptable aggregation latency at 80 nodes without exponential degradation, verifying simulated scalability for multi-room, multi-organization deployment.

3.5. Privacy-Utility Trade-off and Membership Inference Attacks

Fig. 7(a) shows the DP noise coefficient σ 's effect on accuracy and privacy risk: at $\sigma = 0$, accuracy is highest but so is leakage risk; as σ increases, leakage risk falls and accuracy gradually declines, with $\sigma = 0.3 - 0.5$ offering a good balance. Fig. 7(b) shows attack resistance: unprotected success rate is 71.8%, dropping to 55.6%/52.3% with

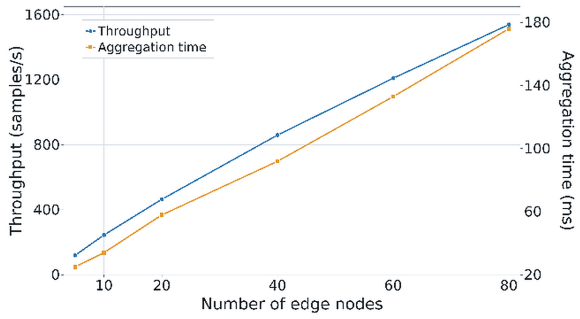

Fig. 6. Throughput and aggregation time under varying edge node numbers

DP-only/secure-aggregation-only, 43.5% with both, and 38.7% with the complete method - showing privacy protection needs multi-mechanism collaboration. This black-box, shadow-model attack (8 shadow models, logistic-regression attacker, balanced 50/50 set, 5 runs) also gives AUC-ROC 0.78/advantage 0.44 unprotected versus AUC-ROC 0.57/advantage 0.08 protected-near zero, indicating the residual 38.7% ASR reflects near-random guessing rather than label inversion.

Across 5 shadow-model runs (mean ± SD): ASR = $38.7\% \pm 1.6\%$, $\text{AUC} - \text{ROC} = 0.57 \pm 0.02$ protected vs. ASR = $71.8\% \pm 2.1\%$, AUC-ROC = 0.78 ± 0.02 unprotected. TPR at a fixed 1% FPR falls from 4.1% (unprotected) to 1.3% (protected), close to the 1.0% expected under random guessing.

The results also suggest that privacy protection should be configured according to the sensitivity of the deployment context. For psychological counseling rooms or sleep-therapy spaces, a slightly stronger privacy setting may be acceptable, whereas general wellness spaces may prioritize a higher recognition accuracy with moderate privacy noise.

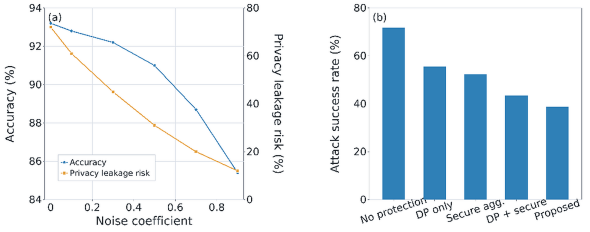

Fig. 7. Privacy-Utility Trade-off and Member Inference Attack Resistance Results

3.6. Modality Missing Robustness and t-SNE Feature Distribution

Device disconnections, unworn wristbands, camera obstruction, and sensor noise are common in real healing spaces. Fig. 8(a) shows accuracy declining as modality-missing rate rises from 0% to 40%, but least for the proposed method: at 40% missing, early fusion drops to ˜72.6%, late fusion to 77.1%, Transformer fusion to 79.6%, while the proposed method maintains ˜84.0%, showing the masked attention mechanism compensates for missing modalities using available ones.

Fig. 8(b) further compares the t-SNE feature distributions of the baseline and proposed methods. The baseline method shows significant overlap among the four healing state samples, indicating insufficient discriminative power in the fusion representation. The features generated by our proposed method exhibit clearer inter-class separation and higher intra-class clustering. This visualization result is consistent with the accuracy and F1-score, demonstrating that the proposed multimodal fusion model can learn more discriminative healing state representations.

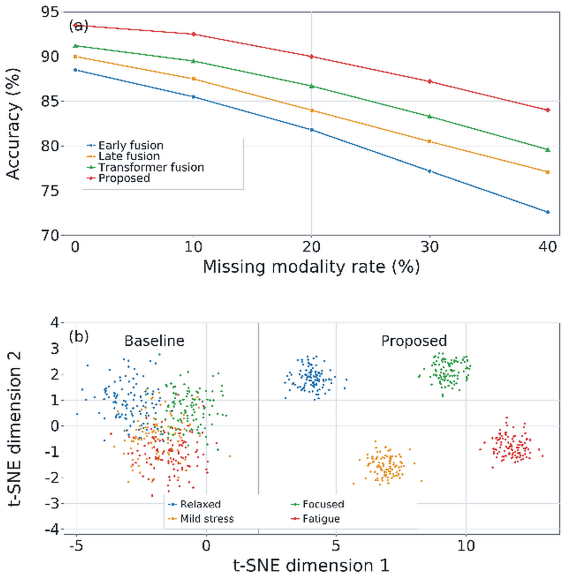

Fig. 8. Comparison of modality loss robustness and t-SNE feature distribution of different algorithms

3.7. Ablation Experiments

Ablation experiments Fig. 9: removing attention fusion prevents dynamic modality weighting, reducing accuracy and F1; removing federated training degenerates the model to single-node/centralized with poor cross-space generalization; removing secure aggregation raises attack success;

removing differential privacy makes sensitive information easier to infer; removing edge preprocessing raises communication cost and latency. The complete model balances accuracy, F1, latency, communication cost, and privacy risk, confirming each module's necessity. Numerically (mild non-IID, 60 nodes): Base (FedAvg, ordinary fusion, no privacy) F1 = 86.4%/ASR = 71.8%; +attention & reliability weighting F1 = 89.4%/ASR = 71.5%; +clipping & compression F1 = 89.1%/ASR = 70.9%; +differential privacy F1 = 88.2%/ASR = 55.6%; +secure aggregation (Full) F1 = 89.7%/ASR = 38.7%-consistent with Figs. 2, 4 and 7/Section 3.6; model size, memory, and aggregation time are reported in Sections 3.5 and 3.9.

A supplementary sensitivity analysis (severe non-IID, 60 nodes; fusion/clipping/DP/secure-aggregation held fixed) isolates the reliability-weighting effect Eq. (14): uniform/FedAvg-style weighting ($\omega\_k^{\wedge}r \propto n\_k$ only) reaches 88.0% average F 1/71.2% worst-client F1; the proposed reliability weighting improves both to 89.7%/77.8% (average consistent with the Full row above); under 20%-corrupted reliability metadata it degrades to 87.1%/65.4%-below the uniform baseline on the worst client. Reliability weighting thus improves average and typical-client performance but is not yet robust to falsified metadata (Section 2.3, 3.10), motivating a trust-scoring or outlier-clipping safeguard as future work rather than a robustness claim.

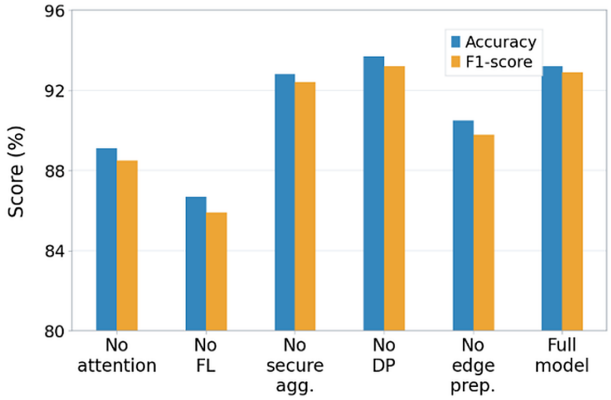

Fig. 9. Ablation experiment results

3.8. Edge Resource Consumption

Fig. 10 compares edge resource consumption: LSTM has lower cost but limited expressiveness; standard Transformer models complex relationships but has higher size, memory, and inference time; lightweight Transformer reduces cost via compression; the proposed method keeps high recognition performance within a deployable size/latency range - showing edge models must meet engineering constraints, not just accuracy.

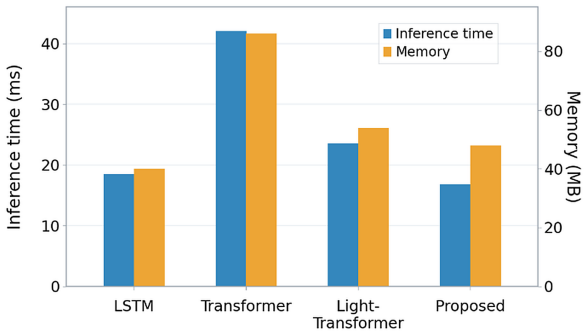

Fig. 10. Comparison of edge resource consumption for different models

3.9. Discussion

Results show the key to smart healing spaces is not just recognizing user states but doing so continuously, stably, and securely across multi-node deployment. The edge-cloud architecture distributes high-frequency processing to local nodes, keeping communication cost and aggregation latency controllable as nodes scale. Raw physiological/behavioral data never leave local nodes, and clipping, noise, and secure aggregation significantly reduce membership-inference success. The attention mechanism improves recognition performance and provides an operational modality-importance indication, letting administrators see whether physiological, environmental, behavioral, or subjective factors drive the current state.

In deployment, the framework applies to hospital rehabilitation, stress-reduction, sleep-intervention, elderly-care, and corporate health spaces, dynamically adjusting lighting, music, air quality, fragrance, and interactive prompts, and reporting group-status trends to managers. It suits cross-institutional collaboration better than centralized platforms since institutions need not share raw data, and improves small-sample nodes' recognition via shared cross-spatial experience versus local-only models. The same edge-cloud pipeline is, in principle, expected to flag anomalous sensor or actuator behavior and support preventive-maintenance scheduling before comfort or therapeutic quality is affected; this fault-diagnosis and maintenance-scheduling capability is a prospective architectural extension of the validated healing-state pipeline and has not itself been experimentally benchmarked in this study (see Limitations below).

From an engineering perspective, implementation should include local device identity management, encrypted storage for short-term buffers, audit logs for model-

update submission, and fallback strategies when a sensor modality is unavailable. These mechanisms are not substitutes for the proposed learning framework, but they are necessary for translating the framework into a reliable smart healing information system. Beyond engineering, deployment should include informed consent and withdrawal, data-minimization and update-deletion procedures, human oversight of interventions, and fairness/accessibility attention across demographic and health groups under institutional governance; the framework is intended to provide decision support-not autonomous clinical diagnosis or treatment-with therapeutic action remaining subject to review by qualified staff.

Two scope notes apply throughout Section 3. First, unless noted, results come from the synthetic simulation and network simulator of Section 2.4 rather than physical hardware or real-user deployment; "real-time," "scalable," and "deployable" describe performance within this simulation envelope, with physical-deployment validation left to future work. Second, because the baselines vary aggregation strategy (FedAvg/FedProx/SCAFFOLD) and fusion strategy (early/late/Transformer) separately rather than as a full factorial grid (Section 2.4), the results show the complete proposed configuration performs favorably under these settings-not that each component individually outperforms every alternative in isolation.

This paper has several limitations: experiments are mainly simulation-based with parameterized scenarios covering node count, heterogeneity, bandwidth, modality loss, and privacy noise, but real-world deployment validation is still needed; DP noise coefficients are fixed rather than adaptive to sensitivity, modality, or task; only membership-inference attacks were tested, leaving gradient inversion, attribute inference, and poisoning defenses for future work; and causal effects of intervention strategies were not explored, motivating future reinforcement-learning or causal-inference approaches. Results should thus be read as proof-of-concept evidence pending real-user deployment studies, ethics review, and cross-site validation. In addition, the facility-maintenance use case is not yet benchmarked against a fault-diagnosis/scheduling task (validation planned on public corpora such as WESAD or K-EmoCon); the baseline comparison is not a full factorial grid; reliability-weighted aggregation shows improved average and worst-client performance over uniform weighting but degrades below the uniform baseline under corrupted reliability metadata (Section 3.8), so a trust-scoring or outlier-clipping safeguard is needed before falsified-metadata robustness can be claimed; modality-missingness uses only an MCAR mechanism; and attention modality-importance weights are untested for explanation fidelity-prioritized directions for the next study phase.

4. Conclusions

This study proposes a scalable distributed multimodal data fusion and privacy protection framework for smart healing spaces, placing preprocessing, encoding, fusion, and inference on local edge nodes and achieving cross-space modeling via federated learning. A mask-aware temporal attention fusion model addresses multimodal heterogeneity and modality loss; gradient clipping, Gaussian differential privacy, and secure aggregation mitigate privacy risks. Simulation results show the complete proposed configuration outperforms the baselines in accuracy, F1-score, non-IID robustness, communication efficiency, simulated real-time response, privacy protection, and edge deployment feasibility under the stated simulation settings (Section 3.10). Future work will integrate real-world deployment data, dynamic privacy budgets, direct equipment-fault-diagnosis and maintenance-scheduling benchmarks, lightweight compression, and adaptive intervention optimization for clinical rehabilitation, mental health, and elderly-care scenarios, with the framework prospectively extensible to facility operation and maintenance, pending dedicated fault-diagnosis and scheduling benchmarks, alongside personalized therapeutic services.

References

[1] Q. Li, Z. Wen, Z. Wu, S. Hu, N. Wang, Y. Li, X. Liu, and B. He, (2023) *"A Survey on Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection"* **IEEE Transactions on Knowledge and Data Engineering** ***35***(4): 3347–3366. DOI: 10.1109/TKDE.2021.3124599.

[2] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, and H. V. Poor, (2021) *"Federated Learning for Internet of Things: A Comprehensive Survey"* **IEEE Communications Surveys & Tutorials** ***23***(3): 1622–1658. DOI: 10.1109/COMST.2021.3075439.

[3] D. C. Nguyen, Q.-V. Pham, P. N. Pathirana, M. Ding, A. Seneviratne, Z. Lin, O. A. Dobre, and W.-J. Hwang, (2022) *"Federated Learning for Smart Healthcare: A Survey"* **ACM Computing Surveys** ***55***(3): 60:1–60:37. DOI: 10.1145/3501296.

[4] V. Mothukuri, R. M. Parizi, S. Pouriyeh, Y. Huang, A. Dehghantanha, and G. Srivastava, (2021) *"A Survey on Security and Privacy of Federated Learning"* **Future Generation Computer Systems** ***115***: 619–640. DOI: 10.1016/j.future.2020.10.007.

[5] G. Muhammad, F. Alshehri, F. Karray, A. El Saddik, M. Alsulaiman, and T. H. Falk, (2021) *"A Comprehensive Survey on Multimodal Medical Signals Fusion for Smart Healthcare Systems"* **Information Fusion *76***: 355–375. DOI: 10.1016/j.inffus.2021.06.007.

[6] A. Kline, H. Wang, Y. Li, S. Dennis, M. Hutch, Z. Xu, F. Wang, F. Cheng, and Y. Luo, (2022) *"Multimodal Machine Learning in Precision Health: A Scoping Review"* **npj Digital Medicine *5***: 171. DOI: 10.1038/s41746-022-00712-8.

[7] J. R. Teoh, J. Dong, X. Zuo, K. W. Lai, K. Hasikin, and X. Wu, (2024) *"Advancing Healthcare Through Multimodal Data Fusion: A Comprehensive Review of Techniques and Applications"* **PeerJ Computer Science *10***: e2298. DOI: 10.7717/peerj-cs.2298.

[8] S. Pati, S. Kumar, A. Varma, B. Edwards, C. Lu, L. Qu, J. J. Wang, A. Lakshminarayanan, S.-H. Wang, M. J. Sheller, K. Chang, P. Singh, D. L. Rubin, J. Kalpathy-Cramer, and S. Bakas, (2024) *"Privacy Preservation for Federated Learning in Health Care"* **Patterns *5***(7): 100974. DOI: 10.1016/j.patter.2024.100974.

[9] I. Dayan, H. R. Roth, A. Zhong, A. Harouni, A. Kaur, A. Gentili, A. Z. Abidin, A. Liu, A. B. Costa, B. J. Wood, C.-C. Tsai, C.-H. Wang, C.-N. Hsu, C.-K. Lee, P. Ruan, D. Xu, D. Wu, E. Huang, F. C. Kitamura, G. Lacey, G. Shachar, G. Yang, H. Roth, H. Li, H. Li, J. Li, J. G. Lee, J. S. Lee, J. S. Kim, J. W. Kim, J. Y. Kim, K. W. Kim, K. Y. Lee, L. Chen, M. G. Linguraru, M. R. Havaei, M. R. Ghavami, S. H. Kim, S. J. Kim, S. M. Kim, S. W. Kim, T. H. Kim, W. J. Kim, W. K. Moon, X. Li, X. Wang, Y. Chen, Y. J. Kim, Y. J. Lee, Y. J. Park, Y. K. Kim, Y. K. Kwon, Y. S. Lee, Y. S. Park, Z. Liu, Z. Xu, Z. Zhang, Z. Wang, Z. Wang, and Q. Li, (2021) *"Federated Learning for Predicting Clinical Outcomes in Patients with COVID-19"* **Nature Medicine *27***(10): 1735–1743. DOI: 10.1038/s41591-021-01506-3.

[10] V. S. Lalapura, J. Amudha, and H. S. Satheesh, (2022) *"Recurrent Neural Networks for Edge Intelligence: A Survey"* **ACM Computing Surveys *54***(4): 91:1–91:38. DOI: 10.1145/3448974.

[11] R. S. Antunes, C. A. da Costa, A. Küderle, I. A. Yari, and B. M. Eskofier, (2022) *"Federated Learning for Healthcare: Systematic Review and Architecture Proposal"* **ACM Transactions on Intelligent Systems and Technology *13***(4): 54:1–54:23. DOI: 10.1145/3501813.

[12] J. Liu, J. Huang, Y. Zhou, X. Li, S. Ji, H. Xiong, and D. Dou, (2022) *"From Distributed Machine Learning to Federated Learning: A Survey"* **Knowledge and Information Systems *64***(4): 885–917. DOI: 10.1007/s10115-022-01664-x.

[13] Y. Lin, Y. Gao, M. Gong, S. Zhang, Y. Zhang, and Z. Li, (2023) *"Federated Learning on Multimodal Data: A Comprehensive Survey"* **Machine Intelligence Research *20***(4): 539–553. DOI: 10.1007/s11633-022-1398-0.

[14] J. Wen, Z. Zhang, Y. Lan, Z. Cui, J. Cai, and W. Zhang, (2023) *"A Survey on Federated Learning: Challenges and Applications"* **International Journal of Machine Learning and Cybernetics *14***(2): 513–535. DOI: 10.1007/s13042-022-01647-y.

[15] J. Yang, K. Tian, H. Zhao, Z. Feng, S. Bourouis, S. Dhahbi, A. A. Khan, M. Berrima, and L. Y. Por, (2025) *"Wastewater Treatment Monitoring: Fault Detection in Sensors Using Transductive Learning and Improved Reinforcement Learning"* **Expert Systems with Applications *264***: 125805. DOI: 10.1016/j.eswa.2024.125805.

[16] J. Yang, D. Hu, M. A. Khan, L. Cascone, M. W. Bhatt, L. Y. Por, H. Yao, and Q. Yang, (2026) *"FSSL-TANet: A Federated Semi-Supervised Temporal Attention Network for Personalized Early Detection of Age-Related Macular Degeneration in Consumer Electronics"* **IEEE Transactions on Consumer Electronics**: 3703836. DOI: 10.1109/TCE.2026.3703836.

[17] J. Yang, V. Govindarajan, S. Arif, X. Xu, M. Kallel, Z. A. Shaikh, Z. Liu, C. Yuan, and L. Y. Por, (2026) *"NeuroCare-SCC: An AI-Enhanced Integrated Sensing–Communication–Computing–Control Framework for Cognitive Health Monitoring"* **IEEE Transactions on Consumer Electronics**: 3667208. DOI: 10.1109/TCE.2026.3667208.

[18] Y. Zhang, X. Li, S. Wang, M. Li, F. Teng, Y. Gao, Z. Wang, H. Yang, L. Y. Por, and Q. Yang, (2026) *"CardioGPT-ViT: Explainable Deep Learning Models with Multimodal Data for Coronary Artery Disease and Heart Failure on Edge Consumer Devices"* **IEEE Transactions on Consumer Electronics**: 3667507. DOI: 10.1109/TCE.2026.3667507.

[19] J. Yang, V. Govindarajan, S. Arif, X. Xu, M. Kallel, Z. A. Shaikh, Z. Liu, C. Yuan, and L. Y. Por, (2026) *"SwarmSense-DNN: A Trustworthy and Decentralized Neural Framework for Proactive Anomaly Defense in Consumer IoT"* **IEEE Transactions on Consumer Electronics *72***(1): 1997–2006. DOI: 10.1109/TCE.2025.3634160.

[20] X. Xu, J. Yang, C. Yuan, Z. Liu, Z. A. Shaikh, M. Kallel, S. Arif, V. Govindarajan, and L. Y. Por, (2025) *"Activity Recognition Empowered Autonomic Systems with Generative Artificial Intelligence"* **ACM Transactions on Autonomous and Adaptive Systems**: DOI: 10.1145/3787222.