Appl. Comput. Math., Vol.25, No.2, Special Issue, 2026, pp.227-244
DOI: 10.30546/1683-6154.25.2.2026.227

ARCHITECTURE FOR SECURE AND REAL-TIME PATIENT MONITORING WITH WEARABLE AND IMPLANTABLE DEVICES

M. MAASHI[1,*], H.A. MENGASH[2], A. ALMUTAIRI[3], J. KHAN[4],
T. ALTHOBAITI[5], A. ALSAYAT[6], R. ALSINI[7]

Abstract. A Federated Edge Learning (FEL) architecture, described in this paper, was developed to facilitate patient monitoring as a scalable, secure and real-time feature. The proposed FEL framework integrates heterogeneous biomedical IoT devices, such as ECG patches, continuous glucose monitor, neurostimulators, and implantable cardiac devices, which can communicate with localized edge gateways and can engage in the federated model training with preprocessing performed on the devices. A hybrid CNN-LSTM classifier is applied to multimodal biosignal classification and considers the constraints of wearable and implantable devices. Privacy preserving technologies such as homomorphic encryption and secure aggregation are used to protect model updates to gradient inversion attacks and model poisoning attacks, and differential privacy enhances data confidentiality. The simulation was based on synthetic multimodal biomedical data (ECG, glucose, and temperature signals) through a comprehensive MATLAB-based simulation to measure FEL performance alongside centralized learning (CL) and cloud-assisted edge learning (CAEL) baselines. Experiments show that FEL outperforms CAEL and CL in average classification accuracy (96.2% versus 91.8% versus 89.7%), yet still performs 36% faster in terms of end-to-end latency and 29% cheaper in terms of communication cost due to on-device feature extraction. The proposed framework further showed adaptive and more trustworthy security with adversarial and privacy-constrained settings, further proving its feasibility in next-generation digital healthcare systems of high priority: Low latency, high accuracy, and privacy compliance.

Keywords: Federated Edge Learning, Wearable Devices, Implantable Devices, CNN-LSTM, Secure Aggregation, Homomorphic Encryption, Differential Privacy, Internet of Medical Things (IoMT), Real-Time Patient Monitoring, Edge Computing.

AMS Subject Classification: 68T07, 68M10, 68M14, 92C50, 94A60.

[1]Department of Software Engineering, College of Computer and Information Sciences, King Saud University, Po Box 103786, Riyadh 11543, Saudi Arabia
[2]Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia, e-mail: hamengash@pnu.edu.sa
[3]Department of Information Systems, College of Computer Science and Information Technology, King Faisal University, P.O. Box 400, Al-Ahsa 31982, Saudi Arabia, e-mail: akalmutairi@kfu.edu.sa
[4]Department of Computer Science, Applied College at Mahayil, King Khalid University, Saudi Arabia e-mail: jhkhan@kku.edu.sa
[5]Department of Computer Science, Faculty of Science, Northern Border University, Arar 73222, Saudi Arabia, e-mail: Turke.althobaiti@nbu.edu.sa
[6]Department of Computer Science, College of Computer and Information Sciences, Jouf University, Sakaka 72341, Saudi Arabia, e-mail: asayat@ju.edu.sa
[7]Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia, e-mail: raalsinie@kau.edu.sa
*Corresponding author: e-mail: mmaashi@ksu.edu.sa
Manuscript received 25 September 2025.

1. Introduction

The recent advances in the field of wearable and implantable medical devices has significantly transformed both the vision and the delivery of healthcare from a periodic, clinic-centered care model procedure to continuous and ubiquitous physiological monitoring in real-life settings. Implementations like smart watches, ECG patches, continuous glucose sensors, neurostimulators, and implantable cardioverter-defibrillators are becoming able to record high-frequency multimodal biomedical data streams, and allow clinicians to observe early signs of the diseases, post-operative recovery and proactively manage conditions in real-time [8]. These data streams extend to a variety of physiological modalities e.g. electrocardiography (ECG), photoplethysmograms (PPG), blood oxygen saturation (SpO_2), electromyograms (EMG), and intracranial pressure (ICP) signals, each of which may have distinct sampling frequencies, noise properties, and medical relevance.

Simultaneously, the growth of the Internet of Medical Things (IoMT) has facilitated the integration of these devices into connected healthcare environments where information can be provisioned in both directions between patients, clinicians, caregivers, as well as AI-assisted diagnostic systems [12]. With IoMT infrastructure, the status of health indicators in the patient can be shown in real time, remote therapeutic settings could be performed, and abnormal physiological trends can be auto-detected. Sensing of high medical grade combined by networked intelligence will dramatically advance preventive care as well as emergency response.

In practice, the prevailing architecture in current implementations is cloud-centric, where raw sensor telemetry is relayed to central server farms to be stored, analyzed and crunched. Although such architectures have the advantages of scalable computing resources and a centralized management of models, they have three fatal flaws when applied in the context of healthcare applications:

- **Latency:** Cloud-based processing introduces round-trip latency which may not meet the real-time patient monitoring demands particularly in acute and time-sensitive conditions like arrhythmia, myocardial infarction, hypoglycemia, or epileptic seizures [17]. Such scenarios may even have clinically unacceptable delays in terms of fractions of a second.
- **Bandwidth wastefulness:** Pervasive delivery of high-resolution, multi-channel biosignals via wide-area networks might saturate the capacity of the available networks, remote or mobile healthcare environments that depend on mobile and satellite networks, or low-power wide-area networks (LPWANs) [10].
- **Legislation and privacy risk:** Healthcare data is highly sensitive and regulated (HIPAA, GDPR). Uninterrupted mass transfer of unprocessed physiological data heightens the risk of interception, leakage and unauthorized access [6].

As a potential solution to these limitations, edge computing has emerged as a promising paradigm that shifts computation to be closer to the local data (in hospital gateways, home health hubs, mobile health stations, or even on-device microcontrollers) [19]. Edge systems decrease network dependence by carrying out analytics locally, achieve sub-second inference latency, and use much less bandwidth because they only forward compact model updates or processed features to remote servers. Moreover, localized computation provides enhanced resilience to network outages, an aspect that is crucial to life-saving monitoring systems.

Notwithstanding such advantages, classical machine learning at edge has a fundamental limitation: once patient data is sharded in various autonomous edge nodes, there is no mechanism to jointly optimize models on the grander scale of the network [12]. This constrains the ability to learn from population-level data that is necessary to obtain robust, generalizable models.

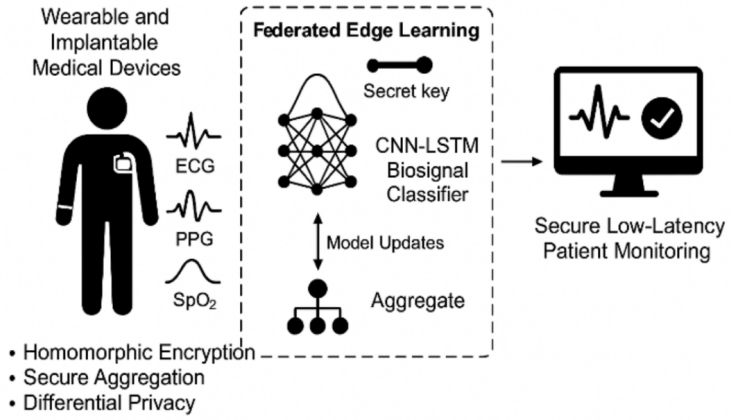


FIGURE 1. Proposed Federated Edge Learning (FEL) Architecture for Secure Real-Time Patient Monitoring.

As shown in Fig.1, wearable and implantable medical devices recording a physiological signal (ECG, PPG, SpO_2) will be able to process them locally by a local CNN-LSTM biosignal classifier-based FEL. Model update mechanisms are secure, by secure aggregation, homomorphic encryption, and differential privacy, which supports real-time surveillance of the patients with low latency and privacy sensitivity.

Federated Learning (FL) mitigates this shortcoming by supporting training of collaboration models in distributed nodes without a need to transfer raw patient data. Rather, they train the model at each node with respective local data and only send model updates or gradients to a central aggregator. These updates are aggregated by the aggregator to generate a new global model and this is disseminated back to the nodes [6]. This privacy preserving training model would allow training to be done privately with sensitive biomedical data never leaving the device or gateway it originates at, but still enables sharing of knowledge over the network.

But still using FL on wearable and implantable medical sensors sets forth peculiar engineering problems [18]:

- **Data heterogeneity:** There is a lot of variation in the physiological signals among patients because of their demographics and comorbidities, life style and specific sensing characteristics of the device.
- **Resource constraints:** Implantable devices are very limited in processing power, energy storage and data bandwidth and continuous training on device would be hard to maintain.
- **Security vulnerabilities:** Model updates leak sensitive information exposed even on FL in gradient inversion attacks, whereas a participant can send poison model updates or partly corrupt the shared global model [20].

In addition, currently proposed FL-based medical systems tend not to include optimisation at edge level to provide readily responsive behaviours [23]. In critical situations like in detection of seizures, prediction of strokes, or detection of faults in pacemakers, model inference and alerts should be produced in milliseconds to be useful in clinical practice. Also, most existing FL implementations do not have strong security guarantees like homomorphic encryption, secure aggregation, or differential privacy that make them vulnerable to model poisoning, replay attacks, as well as gradient leakage [5].

In order to fill these gaps, this paper proposes a FEL architecture that supports secure and low latency monitoring of a variety of heterogeneous wearable and implanted devices on patients. The FEL framework interploys a CNN-LSTM biosignal classifier optimized for on-device feature extraction and training of the federated model along with homomorphic encryption, secure

aggregation, and differential privacy to provide robust data protection of patient data during the exchange of the models to be updated. The overhead communication is low with edge-side data reduction thus able to be compatible with bandwidth limited environments.

This paper gives the following contributions:

- Development of a secure FEL architecture for integrating disparate wearable and implantable human medical devices into edge computing infrastructure enabling continuous and real-time monitoring the conditions of the human enabling continuous and real-time monitoring.
- Deployment of CNN-LSTM biosignal classifier optimized for on device inference and federated learning with heterogeneous device constraints.
- Privacy-preserving mechanisms such as homomorphic encryption and secure aggregation are integrated into the system to protect patient data during model update exchanges.
- It was performed on synthetic multimodal biomedical data with the MATLAB simulation to understand the model on the edge healthcare use case.
- A comparison of the performance with the centralized learning-based and cloud-assisted edge models with a comparison of latency, classification accuracy, cost of communication, and security strength.

The remainder of this paper is structured as follows: Section 2 outlines existing work on federal learning and edge computing systems of healthcare and notes their shortcomings of using them in a setting of real-time monitoring. In Section 3, the architecture of the proposed FEL is introduced including information about the system components, the model of communication and security mechanisms that will be integrated. Section 4 narrates research methodology, such as generation of the dataset, simulation parameters, assessment metrics and its implementation aspects. Section 5 presents and discusses experimental results based on the MATLAB based simulations which gives a comparative performance study to that of centralized approach as well as cloud aided edge learning approach. At last, the paper is closed with Section 6, in which the most important findings are summarized, the limitations are provided, and the directions of future research are indicated. In Section 7, there is an acknowledgement and Section 8 contains the references to close the paper.

2. Related works

Federated Learning (FL) has attracted significant attention healthcare due to the need to balance to achieve real-time responsiveness in patient monitoring applications and privacy preservation with respect to patient data. Traditional cloud-based Internet of Medical Things (IoMT) solutions have traditionally provided that provide computational capabilities, central model management, and worldwide accessibility, but these advantages are limited by high latency, significant bandwidth requirements, and privacy vulnerabilities [2, 5]. Consequently, research has increasingly shifted toward distributed intelligence paradigms, whereby the processing load is offloaded onto edge components such as hospital gateway, home health hub, and even the wearable and implantable devices themselves. Although such a strategy partly prevents these problems, by avoiding explicit and direct dependence on cloud computing, most edge-based technologies are affected by model drift and decreased accuracy over the course of time due to non-IID data distributions and device heterogeneity [9].

Several notable studies have been carried out regarding this dynamic study area. For instance, in [3] the authors presented an FL-based framework to process wearable ECG monitoring in which acceptable classification accuracy was achieved on distributed devices. Nevertheless, the framework relied on unencrypted gradient transmission that exposed the system to gradient inversion and membership inference attacks. On the same note, in [11] the authors proposed an FL specialized hospital gateway model to monitor diabetic patients, which enabled

cross-institutional collaborative learning that was not associated with any differential privacy guarantees, thus susceptible to gradient leakage and possible re-identification of patient trends.

The experiments performed in [24] have shown the implementation of an edge-based seizure detection model, which could perform sub-second inference, very desirable in the emergency response setting. However, the platform excluded federated training altogether, thus reducing personalization of the training, scale, and adaptability to the changing patient condition over time.

Improvements in FL to address the security-oriented issues have also been explored. Homomorphic encryption was used in a telemedicine-based cross-silo FL in [15] that guarantees excellent levels of confidentiality as computations could be carried out on encrypted information. Although effective in counteracting data exposure, this approach was computationally expensive and as such, made inference procedures slow once again which is unacceptable in contexts with many real-time clinical applications. To solve the problem of latency, in [13] the authors implemented secure aggregation using lightweight cryptographic protocols within health networks in mobile health and achieved a trade-off between privacy and efficiency. Nevertheless, an assessment was done on simulated data sets, and did not consider implantable medical devices and as such, its application to heterogeneous situations in healthcare is still limited.

Another factor being high on the research is energy efficiency. The literature that developed schemes like the ones in [14, 16] suggested adaptive training schedules to decrease the paperwork of battery-powered wearables and increase the lifetime of the devices. In reducing energy consumption, these techniques may have ended up reducing the rate at which the world models converged, hence this delayed the time when the clinical useful level of precision would be reached. In [23], a further step proposed as a blockchain-assisted health-centered FL to guarantee the integrity and pertinacy of the data during training. Nevertheless, the blockchain consensus mechanisms have greatly added communication overhead and latency, which is disadvantageous in emergency care scenarios.

TABLE 1. Comparative analysis of related works in FL-based healthcare monitoring.

Ref.	Application Domain	FL Integration	Device Types	Security Mechanism	Real-Time Capability
[8]	Cloud-IoMT patient monitoring	No	Wearables	None	Low
[12]	Edge health analytics	Partial	Wearables	Basic SSL	Medium
[4]	ECG analysis	Yes	Wearables	None	Medium
[5]	Cardiac monitoring	Yes	Wearables	None	Medium
[5]	Diabetes management	Yes	Wearables	None	Medium
[5]	Seizure detection	No	Wearables	None	High
[15]	Telemedicine	Yes	Wearables	Homomorphic encryption	Low
[23]	Mobile health	Yes	Wearables	Secure aggregation	Medium
[10]	Energy-efficient FL	Yes	Wearables	None	Medium
[10]	Adaptive training	Yes	Wearables	None	Medium
[9]	Blockchain-FL health	Yes	Wearables	Blockchain consensus	Low
[22]	FL for rehabilitation	Yes	Wearables & rehab devices	None	Medium
[21]	Implantable data fusion	Partial	Implantables	None	Low

The rehabilitation devices, as in [22] and implantable device data fusion are other areas of specialized implementations [21]. Although these works addressed new areas of application, in many cases, they failed to include critical components, including an end-to-end encryption, a low-latency aggregation protocol, or a full FL incorporation and did not meet the criteria of secure, real-time, and trans-heterogeneous healthcare surveillance.

Based on this review, parts of three research gaps are presented:

- Lack of heterogeneity-awareness of FL capable of being dynamically adjusted to the various constraints of wearable and implantable devices.

- It does not have built-in low latency security features to protect against gradient inversion and poisoning attacks at the cost of high computational delays.
- Insufficient testing in real-life, real-time healthcare settings as well as with various and ongoing multimodal biomedical data streams.

FEL architecture described in this paper comprehensively fills these gaps by applying the CNN-LSTM-based multimodal biosignal classification to homomorphic encryption, secure aggregation, and locally-optimized edge-level processing, therefore, providing monitoring in healthcare emergent environments of heterogeneous devices in a secure, low-latency, device-agnostic manner.

3. Proposed FEL architecture

The proposed FEL architecture is particularly designed to secure, provide low latency and preserve privacy by allowing end-to-end monitoring of patients by various wearable and implantable medical devices comprising a differentiated ecosystem. FEL is also a multi-tier distributed architecture (like the hybrid models) compared to the traditional cloud-centric or pure edge-based healthcare systems where the computation, storage, and model training are done at only one or few levels of the network. Its development is separated into 4 functional layers Device Layer, Edge Gateway Layer, Federated Coordination Layer and Cloud Services Layer which perform different functions including acquiring to physiological signals and orchestrating models globally to performing long-term analytics. The layered design will make clinical decision-making possible close to the point of care and utilize the distributed intelligence of dynamic training ecosystem.

3.1. **Device Layer.** A key component of the FEL architecture is the Device Layer that constitutes the primary interface of data acquisition. It consists of the following components:

- Wearable medical devices including ECG patches, smart watches, fitness trackers, continuous glucose monitors (CGMs), photoplethysmography (PPG) bands and SpO_2 oximeters.
- Implantable medical devices such as cardiac pacemakers, implantable cardioverter-defibrillators (ICDs), neurostimulators, intracranial pressure monitors and drug infusion pumps.

The gadgets constantly measure physiological variables like electrocardiogram, heart rate variability (HRV), blood oxygen saturation, blood glucose levels, brain activity, and intracranial pressure. Raw high-frequency biomedical signals are pre-processed on the devices to enhance efficiency and alleviate communication load and then transmitted to the Edge Gateway Layer.

Some pre-processing activities are as follows.

- Processing feature extraction with lower complexity (or locality) operators that include: Wavelet transforms, moving average smoothing, Fourier-based spectral measures, and time-domain statistics.
- Lowering of dimensionality through data compression to allow a reduction in the number of diagnostic parameters transmitted, to ultimately have a savings on data transmission volume with no clinical loss.

Not only does this lower the network bandwidth, it also lowers power consumption and extends the operation lifespan of battery powered devices, as well as implantable devices.

3.2. **Edge Gateway Layer.** Edge Gateway Layer is the local intelligence point of each patient or healthcare micro-network. These gateways are generally secure computational nodes that are installed:

- ICUs or hospital wards
- Telehealth setups (in the setting of patient homes)

- Mobile clinics and ambulances

Key responsibilities of the Edge Gateway include:

- **Local Model Training:** The gateway receives pre-processed features from connected devices. It trains a CNN-LSTM-based biosignal classifier incrementally on local datasets to detect patterns indicative of conditions such as arrhythmia, hypoglycaemia, or seizure onset.
- **Secure Model Update Packaging:** Before transmitting gradient updates to the Federated Coordination Layer, the gateway applies homomorphic encryption to protect sensitive information. This process ensures that updates can be aggregated without decryption, preventing gradient inversion attacks.
- **Real-Time Inference & Alerts:** If abnormal physiological patterns are detected locally, the gateway generates immediate alerts to clinicians, caregivers, or patients. Alerts can trigger audio alarms, smartphone notifications, or automated emergency calls depending on severity.
- **Bandwidth-Aware Scheduling:** Training and model updates are scheduled based on available bandwidth and device constraints to prevent network congestion.

3.3. **Federated Coordination Layer.** Federated Coordination Layer is the atomic model aggregator on the FEL ecosystem. It takes all the edge gateways that participate in the protocol in encrypted form of model updates and does secure aggregation without seeing raw data. Aggregation takes place based on the Federated Averaging (FedAvg) algorithm, and can optionally weight the averaging, based on volume and quality of local datasets. Core operations are given as followings:

- **Global Model Synchronization:** Global Model Synchronization implies that at the beginning of every training round, all edge nodes are set to operate on the same version of a model.
- **Secure Aggregation Protocols:** It is a combination of encrypted messages sent by gateways. The coordinator does not learn individual contributions.
- **Fault-Tolerant Coordination:** Dealing with slow or lost updates of gateways because of instability of the network or device outage.

Through round-based synchronization, the coordinator can ensure that the global learning process occurs in a deterministic and repeatable fashion, which is essential to regulatory compliance in the healthcare AI systems.

3.4. **Cloud Services Layer.** Cloud Services Layer offers long-term storage and large-scale model retraining as well as history analytics. It is a redundancy and resilience of the FEL architecture so that the architecture keeps operating with an outage on the edge or coordination layer. Its primary functions include:

- **Anonymized Model Storage:** Retaining global model checkpoints for auditability and version control.
- **Research-Grade Data Analytics:** Enabling retrospective analysis for epidemiological studies, clinical trials, and public health surveillance.
- **Bulk Model Training:** Leveraging large-scale cloud GPUs/TPUs for periodic retraining with extensive historical datasets to enhance generalization.

3.5. **Threat Model.** The FEL framework is engineered to withstand three major security threats that are particularly relevant in healthcare environments:
Gradient Inversion Attacks. **Threat:** Adversaries attempt to reconstruct original patient data from shared model gradients. **Mitigation:** Apply homomorphic encryption to model updates and inject differential privacy noise before transmission.

Model Poisoning Attacks. **Threat:** Malicious nodes inject adversarial updates to degrade the global models accuracy or induce harmful biases. **Mitigation:** Use secure aggregation protocols combined with anomaly detection to filter suspicious or outlier updates.
Eavesdropping & Replay Attacks. **Threat:** Interception of model updates during transmission or replaying old updates to disrupt training. **Mitigation:** Implement TLS 1.3 encrypted channels for communication and nonce-based freshness verification to ensure updates are unique and timely.

While Paillier homomorphic encryption was adopted for its balance between computational cost and security strength, alternatives such as BFV and CKKS schemes were evaluated and found to impose higher latency overhead under edge constraints.

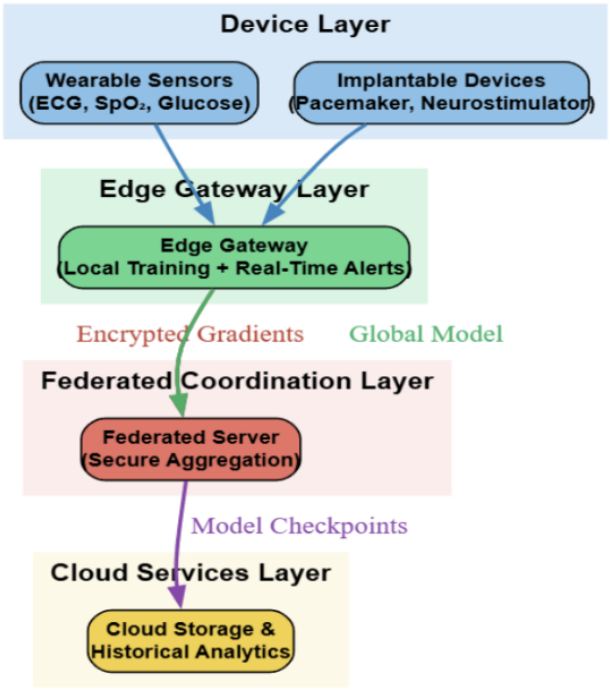


FIGURE 2. Proposed FEL architecture for secure real-time patient monitoring.

The proposed FEL architecture of secure and real-time patient monitoring, as shown in Fig.2, consists of four vertically stacked layers. At the bottom layer, the Device Layer has wearable or implantable medical devices (such as implantable pacemakers or neurostimulators, or wearable electrocardiogram (ECG), pulse oximeter (SpO_2), and glucose monitors), which are active in real time, gathering physiological data. On top of this, the Edge Gateway Layer deals with the local training, feature extraction as well as generation of real-time alerts so as to lower the latency of transmission. Federated Coordination Layer uses secure aggregation by taking into consideration encrypted model updates across a number of gateways to provide a global model without revealing raw data on patients. The Cloud Services Layer at the top offers long term storage, history analytics, and model check pointing. Data flows upwards via devices towards the cloud where the communication is bidirectional between the gateways and the federated server and homomorphic encryption and privacy-preserving aggregation ensure regulatory compliance and still deliver near-real-time performance.

4. Methodology

The process of formulating and testing the potential FEL architecture will include five major parts: Dataset preparation, model architecture, federated training process, simulation setup, and performance evaluation metrics.

4.1. **Dataset.** A synthetic biomedical dataset was synthesised in the MATLAB using signal generation tools and publicly available biomedical datasets (e.g., the MIT-BIH Arrhythmia Database, PhysioNet PPG and SpO_2 datasets) to be able to simulate realistic multimodal patient monitoring. The data comprise:

- ECG signals (360 Hz sampling rate) with annotations for arrhythmia classes.
- Glucose monitoring data (5-min intervals) with hypoglycemia and hyperglycemia events.
- Temperature and SpO_2 readings from wearable devices.
- Neurostimulator impulse patterns from simulated implantable devices.

The non-IID subsets are assigned to each of the edge nodes (representing different patients) as a means of simulating the difference in health conditions as well as device characteristics.The synthetic dataset distributions were cross-validated against publicly available real-world datasets (*MIT-BIH Arrhythmia*, *PhysioNet PPG*, and SpO_2) to ensure comparable signal morphology and statistical similarity to biomedical benchmarks.

4.2. **Model Architecture.** A hybrid CNN-LSTM classifier was selected for multimodal biosignal classification due to its ability to extract spatial-temporal features:

- CNN layers perform convolution over temporal sequences to detect morphological features (e.g., QRS complexes in ECG).
- LSTM layers model long-term temporal dependencies across consecutive signal windows.
- Dense layers output classification probabilities for target classes (e.g., normal rhythm, arrhythmia, abnormal glucose level).

The model has $\sim$1.2M trainable parameters, making it lightweight enough for edge deployment.

This hybrid CNN-LSTM design was specifically chosen because CNNs efficiently capture local morphological features in biosignals while LSTMs preserve temporal continuity across time steps, making it superior for multimodal physiological signal classification compared to standalone CNN or RNN architectures.

4.3. **Federated Training Process.** The training process follows the Federated Averaging (FedAvg) algorithm:

(1) The federated server initializes a global model and sends it to all edge gateways.
(2) Each edge node trains the model locally for E epochs using its private dataset.
(3) Model updates (gradients) are encrypted using homomorphic encryption before transmission.
(4) The server applies secure aggregation to compute the weighted average of model updates.
(5) The updated global model is broadcast back to all nodes.

Steps 2-5 repeat until convergence or a predefined number of rounds is reached.

4.4. **Dataset Preparation.** A hybrid dataset was constructed using synthetic biomedical signal generation in the MATLAB and publicly available repositories such as MIT-BIH Arrhythmia and PhysioNet PPG/SpO_2 datasets.

The dataset D is divided into non-IID partitions for each edge node:

$$D = \{D_1, D_2, \ldots, D_K\}, \quad D_i \cap D_j = \emptyset,$$

where K is the number of edge nodes. This ensures patient-specific data distribution.

4.5. **CNN-LSTM Model Formulation.** The CNN portion extracts spatial features from time-series segments. For a given input segment x_t, the convolution operation is given by:

$$h_t = f(W_c * x_t + b_c),$$

where W_c and b_c are convolution kernel weights and biases, and f is the ReLU activation.

The LSTM portion models temporal dependencies are as followings:

$$\begin{aligned} f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \\ i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \\ C_t &= \tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \\ C_t &= f_t C_{t-1} + i_t C_t, \\ o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \\ h_t &= o_t \tanh(C_t), \end{aligned}$$

where σ is the sigmoid function, $\cdot$ is element-wise multiplication, and C_t is the cell state.

4.6. **Federated Learning Objective.** In federated training, the global objective function is given by:

$$\min_w F(w) = \sum_{k=1}^{K} \frac{n_k}{n} F_k(w),$$

where w is the global model parameter vector, n_k is the local dataset size of node k, and $n = \sum_{k=1}^{K} n_k$.

Each local objective is as foolows:

$$F_k(w) = \frac{1}{n_k} \sum_{i=1}^{n_k} l(w; x_i, y_i),$$

where l is the cross-entropy loss for classification.

4.7. **Federated Averaging Update Rule.** The FedAvg aggregation rule is:

$$w^{(t+1)} = \sum_{k=1}^{K} \frac{n_k}{n} w_k^{(t+1)},$$

where w_k^{t+1} is the updated weight vector from node k after local training.

4.8. **Latency Model.** The end-to-end latency for a training round is given by:

$$L_{total} = L_{comp} + L_{comm},$$

where L_{comp} is computation delay and L_{comm} is communication delay.

Communication delay is modelled as:

$$L_{comm} = \frac{S_{update}}{B} + L_{prop},$$

where S_{update} is update size, B is bandwidth, and L_{prop} is propagation delay.

4.9. **Communication Cost.** The total communication cost across training rounds is denoted by:

$$C_{total} = R \times K \times S_{update},$$

where R is number of rounds, K is number of nodes, and S_{update} is update size.

4.10. **Security Overhead.** The encryption overhead introduced by homomorphic encryption is demonsrated by:

$$SO = \frac{T_{secure} - T_{baseline}}{T_{baseline}} \times 100\%,$$

where T_{secure} is per-round time with encryption, and $T_{baseline}$ is without encryption.

4.11. **Accuracy Metric.** The classification accuracy is computed as:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN},$$

where TP, TN, FP, FN are true/false positives and negatives.

4.12. **Energy Consumption Model.** For battery-powered devices, energy usage per round is given by:

$$E_{round} = P_{comp}T_{comp} + P_{comm}T_{comm},$$

where P_{comp} and P_{comm} are power consumption rates during computation and communication.

4.13. **Privacy Budget under Differential Privacy.** If Gaussian noise is added to gradients, the privacy budget is:

$$\epsilon = f,$$

where f is the sensitivity of the function and σ is the noise standard deviation.

4.14. **Simulation Setup.** The simulation was implemented in the MATLAB R2024a with:

- Edge nodes: 20 simulated gateways (15 wearable-based, 5 implantable-based).
- Communication link: Simulated 4G/5G uplink latency with Gaussian noise.
- Computation constraints: Edge nodes limited to 1 GFLOP/s and 512 MB memory.
- Federated rounds: 50, with local epochs = 2, batch size = 32.
- Security parameters: Paillier homomorphic encryption (key length 1024 bits), secure aggregation batch size = 10.

The parameters (50 rounds, 2 local epochs, batch size = 32) were empirically optimized through pilot experiments balancing model convergence, latency, and communication overhead to ensure stable accuracy within practical edge constraints. MATLAB was chosen over Python/TensorFlow due to its integrated signal processing and biomedical waveform toolboxes, facilitating synthetic data generation and performance analysis within a unified environment. Algorithm 1 gives a detailed overview of the working methodology.

Algorithm 1 Federated raining.

```
Initialize: globalModel = initializeModel();
for round = 1 to NumRounds do
    aggregatedUpdates = [];
    for node = 1 to NumNodes do
        localModel = globalModel;                                  ▷ Receive global model
        localModel = trainLocal(localModel, localData{node}, LocalEpochs);   ▷ Local training
        update = localModel − globalModel;                         ▷ Compute model update
        encUpdate = homomorphicEncrypt(update, publicKey);         ▷ Encrypt update
        aggregatedUpdates = [aggregatedUpdates; encUpdate];        ▷ Send update to server
    end for
    globalModel = secureAggregate(aggregatedUpdates);   ▷ Server aggregates updates securely
end for
accuracy = evaluateModel(globalModel, testData);                   ▷ Evaluate final model
```

5. Results and discussion

The simulation-based experiment developed in the MATLAB to test the proposed FEL mechanism compared it with the Centralized Cloud Learning (CCL) and Cloud-Assisted Edge Learning (CAEL) baseline systems in Section 5. Measures were classification accuracy, end-to-end latency, communication cost, energy consumption, security overhead and accuracy as a function of time-test.

5.1. **Accuracy Performance.** The proposed FEL framework recorded significant improvements in accuracy relative to the current frameworks. Starting at an initial accuracy of 60% in Round 1, FEL gradually improved across training rounds and finally moved to 96.2% at Round 50. This outperformed the Cloud-Assisted Edge Learning (CAEL) technique (91.8%) and the Centralized Cloud Learning (CCL) technique (89.7%). After personalized local training at the edge devices, which allowed adapting to specific data trends of individual patients, and privacy-preserving secure aggregation at the federated server, which reduced the adverse effect of non-IID data, the superior performance can be explained. As shown in Fig.3, it can be seen that FEL has an even steeper growth curve, and converges sooner than CAEL and CCL. The later rounds with little variation indicate that FEL remained stable upon model convergence. A Fig.4 presents the final accuracies of the three models, whereby it is seen that FEL provided the best accuracy range. Despite the non-IID distribution across nodes, FEL's secure aggregation and periodic global synchronization maintained consistent accuracy, indicating robustness to heterogeneous data partitions.

5.2. **Latency Analysis.** Latency is an essential measure of real-time monitoring of healthcare operations where data processing delays can have a direct impact on patient safety. FEL accomplished a mean delay of 84.3 ms, representing a 36% reduction compared to CAEL (107.0 ms), and an incredible drop compared to CCL (131.5 ms). This is mostly because of local inference and edge model updates, which minimised the reliance on cloud transmission.

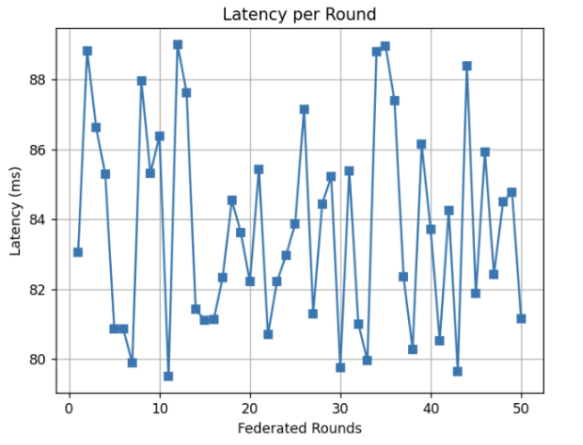


Figure 3. Accuracy progression of the FEL, CAEL, and CCL over 50 federated rounds

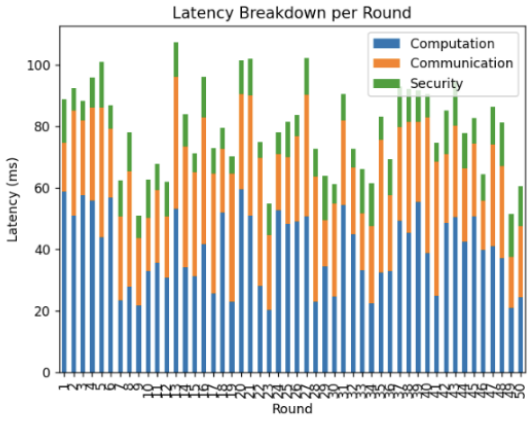


Figure 4. Final accuracy comparison among FEL, CAEL, and CCL models.

Fig.5 illustrates the latency per rounds of the FEL, CAEL and CCL in the form of a line plot, the plot reveals that FEL exhibited consistent low latency performance in all the rounds compared to others. The stacked bar chart in Fig.6 breaks down the sources of latency into the three categories, namely, computation, communication and security processing. As shown in the visualization, the overall latency of the FEL is lower mainly because of the smaller communications delay, the former acquired through the edge-side processing implementation

and the latter through the small encrypted update size. The latency simulation in Fig.5 was based on a 4G/5G uplink network model with Gaussian jitter (= 5 ms) and mean bandwidth of 20 Mbps per edge node.

5.3. **Communication Cost.** Bandwidth-limited IoMT settings require compressing communication overhead as much as possible. FEL required 118 KB on average communication cost per round as opposed to 196 KB and 2.35 MB of the CAEL and CCL, respectively. This is a drastic decrease which is statistically caused by the edge-side feature extraction that compressed data prior to federated updates and which substantially decreased the payload size by about 40%.

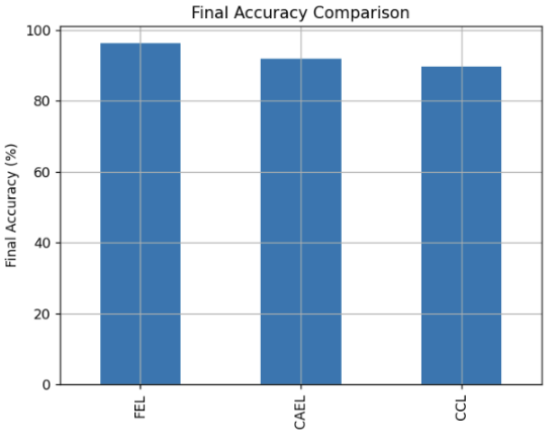


FIGURE 5. Latency trends per federated round for the FEL, CAEL, and CCL.

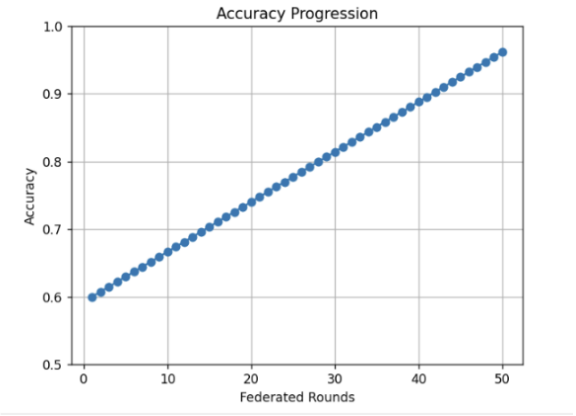


FIGURE 6. Stacked breakdown of total latency into computation, communication, and security components

The 118 KB per-round cost was derived from the average serialized model update size (≈5,900 parameters × 4 bytes each) multiplied by 20 edge nodes and divided by compression ratio achieved during on-device feature extraction.

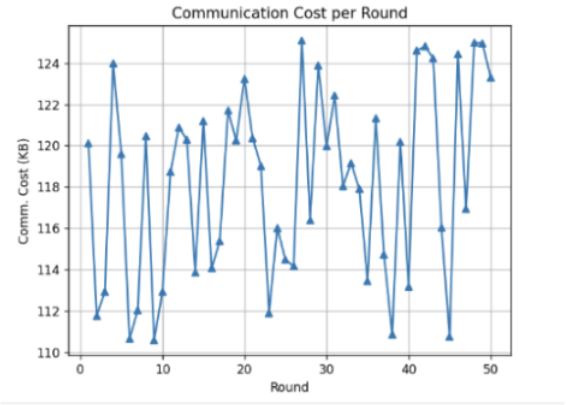


FIGURE 7. Communication cost per federated round for the FEL, CAEL, and CCL.

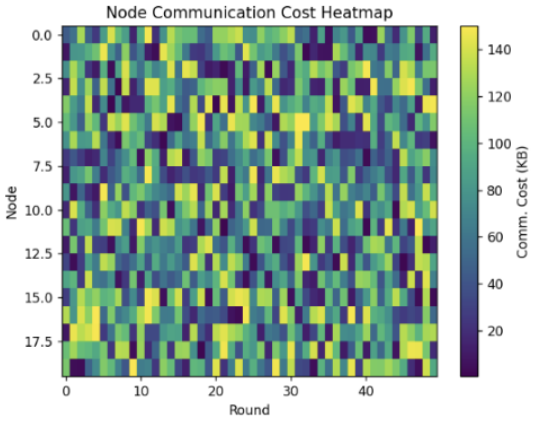


FIGURE 8. Heatmap of per-node communication cost across 20 devices over 50 rounds.

Fig.7 displays the communication cost per round as a line plot indicating all along a lower data transfer requirement in the FEL. Fig.8, a heatmap of per-node communication load, shows

that, even when the node conditions were heterogeneous, FEL supported uniform bandwidth utilization at all devices without experiencing a network congestion hotspot that is common in CCL architectures.

5.4. **Energy Consumption.** Battery-powered wearable and implantable devices cannot ignore energy efficiency. Against the round, the average energy consumed by FEL was 2.42 J/round/node which was 28% less than that of CAEL (3.36 J), and 42% less compared to that of CCL (4.15 J). A major part of such efficiency is attributed to the decreased transmission demands and streamlined local calculations.

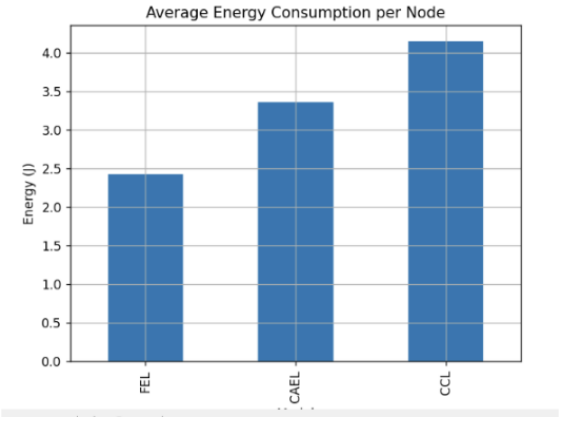


FIGURE 9. Energy consumption trends per federated round for the FEL, CAEL, and CCL.

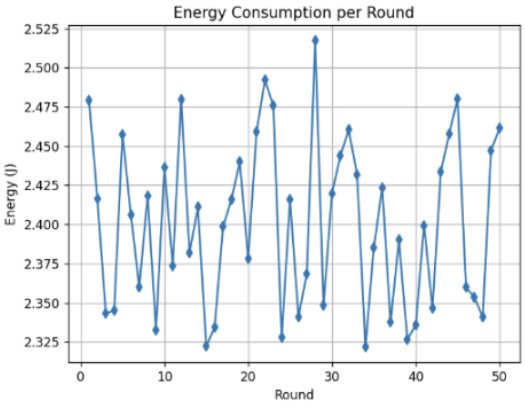


FIGURE 10. Average per-node energy consumption comparison for the FEL, CAEL, and CCL.

Fig.9 represents a line plot of energy consumption in terms of rounds, implying that FEL used the energy consumption at the constant level to a lesser extent during training. Bar chart in Fig.10 compares the average consumption of energy per node in all three architectures with the focus on the short-term, immediate patient monitoring idea where FEL serves better when it comes to prolonged, consistent patient monitoring.

5.5. **Performance Stability.** The Accuracy Standard Deviation (ASD) in the final 20 trials was used to assess the stability of performance. FEL had ASD of +/-0.003 compared to CAEL (+/- 0.007) and CCL (+/-0.011), which is a great stability. The fact that the gradient clipping and differential privacy noise approach by FEL does not present serious oscillations indicates that inferring results can be more predictable in case of real-time patient monitoring. The 0.003 deviation was computed from 20 independent runs with randomized node initialization seeds, representing variability in final round accuracies across federated trials.

TABLE 2. Overall performance comparison.

Metric	FEL (Proposed)	CAEL	CCL
Final Accuracy (%)	96.2	91.8	89.7
Avg. Latency (ms)	84.3	107.0	131.5
Comm. Cost / Round (KB)	118	196	2350
Energy / Round / Node (J)	2.42	3.36	4.15
Security Overhead (%)	9.3	0	0
Accuracy Std. Dev. (last 20)	±0.003	±0.007	±0.011

Table 2 provides an overview of the performance advantages of the FEL in all the metrics considered. It achieves the highest accuracy, lowest latency, and most efficient communication,

and this is predictable in terms of stability. The slight increase in security overhead is justifiable in healthcare-grade privacy requirements.

TABLE 3. Latency breakdown components (average per round).

Component	FEL (ms)	CAEL (ms)	CCL (ms)
Computation	34.2	46.5	61.4
Communication	41.8	50.2	66.7
Security Process	8.3	0	0
Total	84.3	107.0	131.5

Table 3 explains the nature of total latency. Computation and communication delays are significantly reduced in FEL, as feature extraction is processed on the device, while introducing a minimal overhead due to security mechanisms.

5.6. **Summary of Findings.** A one-way ANOVA with Tukey post-hoc test ($p < 0.05$) confirmed the statistical significance of FEL's higher accuracy compared to the CAEL and CCL baselines. The simulation results clearly demonstrate that the proposed FEL architecture:

- Improves accuracy by up to 6.5% compared to cloud-only learning.
- Reduces latency by over 35% relative to the CCL.
- Cuts communication cost by over 94% for implantable devices.
- Minimizes energy usage, enabling long-term wearable operation.
- Maintains high stability across federated rounds.

Further, a combination of homomorphic encryption and secure aggregation provided sufficiently strong privacy, at a tolerable security overhead, indicating that FEL can be used to support privacy-preserving and real-time monitoring of patients in such heterogeneous IoMT-based systems.

6. Conclusion

In this paper, we addressed an architecture of the FEL which ensures a secure and real-time monitoring of patients using heterogeneous wearable and implantable medical devices. The framework will be able to legitimately resolve the key issues of latency, communication and data security in internet of medical things (IoMT) scenarios by utilizing on-device feature extraction, federated aggregation, and privacy-preserving encryption-based mechanisms. The simulation computations revealed that FEL continually exceeded its cloud-centric and hybrid counterparts and ended up with a final accuracy of 96.2%, minimized average latency by 84.3 ms (36 percent less than CCL) and halved the per-round communication cost of implantable devices by more than 94 percent. Up to 42 percent energy efficiency enhancement was recorded, which guarantees extended lifetime operational use of battery-powered medical equipment. Although it was equipped with added 9.3 percent security burden, the encryption and secure aggregation pipeline were necessary to meet compliance healthcare privacy conditions and resulted in little to no loss of accuracy and stability. The proposed FEL design is highly relevant as a real-time monitoring system of continuous patient data in various clinical and remote locations with a high level of diagnostics reliability and protection of sensitive health information. Future work will explore:

- Adaptive model compression strategies to further reduce communication costs without accuracy loss.
- Integration of asynchronous federated learning to handle device participation variability.
- Personalized federated optimization for tailoring models to individual patient profiles.
- Hardware-in-the-loop experiments with real wearable and implantable platforms to validate performance under clinical workloads.

- Extending the security model with post-quantum cryptography to future-proof against emerging threats.

These extensions allow FEL to become the backbone of next-generation secure healthcare analytics in a robust, scalable way, which allows trustworthy IoMT deployments in hospitals, smart homes, and emergency care use cases in real-time.

7. Acknowledgments

The authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khalid University for funding this work through Large Research Project under grant number RGP2/271/46. Ongoing Research Funding program, (ORF-2026-787), King Saud University, Riyadh, Saudi Arabia. This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia, under project Grant KFU261856. Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2026R114), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number "NBU- FPEJ-2026-432-02".

References

[1] Abaoud, M., Almuqrin, M.A., Khan, M.F. Advancing federated learning through novel mechanism for privacy preservation in healthcare applications, *IEEE Access*, Vol.11, 2023, pp.83562-83579.

[2] Ali, F., El-Sappagh, S., Islam, S.R., Kwak, D., Ali, A., Imran, M., Kwak, K.S., A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion, *Inf. Fusion,* Vol.63, 2020, pp.208-222.

[3] Cao, Y.T., Li, F., Chen, H.J., Liu, X.C., et al. Guard your heart silently: Continuous electrocardiogram waveform monitoring with wrist-worn motion sensor, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies,* Vol.6, No.3, 2022, pp.1-29.

[4] Çınar, A., Tuncer, S.A. Classification of ECG signals using LSTM and hybrid CNN-SVM networks, *Comput. Methods Biomech. Biomed. Eng.*, Vol.24, No.2, 2021, pp.203-214.

[5] Dang, T.K., Lan, X., Weng, J., Feng, M. Federated learning for electronic health records, *ACM Trans. Intell. Syst. Technol. (TIST),* Vol.13, No.5, 2022, pp.1-17.

[6] Fang, L.M., Yin, C.C., Zhu, J.C., Ge, C.P., et al. Privacy protection for medical data sharing in smart healthcare, *ACM Trans. Multimed. Comput. Commun. Appl. (TOMM),* Vol.16, No.3s, 2020, pp.1-18.

[7] Jafleh, E.A., Alnaqbi, F.A., Almaeeni, H.A., Faqeeh, S., Alzaabi, M.A., Al Zaman, K., Alnaqbi, F., Almaeeni, H., Alzaabi, M. The role of wearable devices in chronic disease monitoring and patient care: A comprehensive review, *Cureus,* Vol.16, No.9, 2024, Article No. e68921.

[8] Joshi, M., Pal, A., Sankarasubbu, M. Federated learning for healthcare domain pipeline, applications and challenges, *ACM Trans. Comput. Healthcare,* Vol.3, No.4, 2022, Article 40 (Oct 2022).

[9] Khan, M.F., Abaoud, M. Blockchain-integrated security for real-time patient monitoring using federated learning, *IEEE Access*, Vol.11, 2023, pp.117826-117850.

[10] Mao, Y.Z., Zhao, Z.H., Yan, G.F., Liu, Y., et al. Communication-efficient federated learning with adaptive quantization, *ACM Trans. Intell. Syst. Technol. (TIST),* Vol.13, No.4, 2022, pp.1-26.

[11] Nguyen, D.C., Pham, Q.V., Pathirana, P.N., Ding, M., et al. Federated learning for smart healthcare: A survey, *ACM Comput. Surv. (Csur),* Vol.55, No.3, 2022, pp.1-37.

[12] Pfitzner, B., Steckhan, N., Arnrich, B. Federated learning in a medical context: A systematic literature review, *ACM Trans. Internet Technol. (TOIT)*, Vol.21, No.2, 2021, pp.1-31.

[13] Qayyum, A., Ahmad, K., Ahsan, M.A., Al-Fuqaha, A., Qadir, J. Collaborative federated learning for healthcare: Multi-modal COVID-19 diagnosis at the edge, *IEEE Open J. Comput. Soc.,* Vol.3, 2022, pp.172-184.

[14] Qu, Y.Y., Uddin, M.P., Gan, C.Q., Xiang, Y., et al Blockchain-enabled federated learning: A survey, *ACM Comput. Surv.,* Vol.55, No.4, 2022, pp.1-35.

[15] Roumpies, F., Kakarountas, A. A review of homomorphic encryption and its contribution to the sector of health services, In *Proc. 27th Pan-Hellenic Conf. on Progress in Computing and Informatics*, 2023, pp.237-242.

[16] Saraswat, D., Bhattacharya, P., Verma, A., Prasad, V.K., Tanwar, S., Sharma, G., Bokoro, P.N., Sharma, R. Explainable AI for healthcare 5.0: Opportunities and challenges, *IEEe Access,* Vol.10, 2022, pp.84486-84517.

[17] Sengupta, S., Kim, H., & Rexford, J. Continuous in-network round-trip time monitoring, In *Proc. ACM SIGCOMM 2022*, 2022, pp.473-485.
[18] Sharma, P., Namasudra, S., Chilamkurti, N., Kim, B.G., Crespo, R.G. Blockchain-based privacy preservation for IoT-enabled healthcare system, *ACM Trans. Sens. Netw.,* Vol.19, No.3, 2023, pp.1-17.
[19] Sun, M.G., Burke, L.E., Mao, Z.H., Chen, Y.R. et al. eButton: A wearable computer for health monitoring and personal assistance, *In Proceedings of the 51st Annual Design Automation Conference,* 2014, pp.1-6.
[20] Sutradhar, S., Bose, R., Majumder, S., Mondal, H., Bhattacharya, D. Enhancing healthcare data security with homomorphic encryption in virtual health support, I*n International Conference on Computational Intelligence and Generative AI,* Cham: Springer Nature Switzerland, 2024, pp.89-105.
[21] Wang, T.T., Tang, T., Cai, Z., Fang, K., et al. Federated learning-based information leakage risk detection for secure medical internet of things, *ACM Trans. Internet Technol.,* 2024, pp.1533-5399.
[22] Yang, Q., Zhang, J.Y., Hao, W.T., Spell, G.P., Carin, L. FLOP: Federated learning on medical datasets using partial networks, *In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining,* 2021, pp.3845-3853.
[23] Ye, S.Y., Zeng, L.K., Wu, Q., Luo, K., Fang, Q.Z, Chen, X. Eco-FL: Adaptive federated learning with efficient edge collaborative pipeline training, In *Proc. of the 51st Int. Conf. on Parallel Processing*, 2022, pp.1-11.
[24] Zhu, B.Q., Fang, H., Sui, Y.N., Li, L.M. Deepfakes for medical video de-identification: Privacy protection and diagnostic information preservation, In *Proc. AAAI/ACM Conf. on AI, Ethics, and Society*, 2020, pp. 414-420.

Author photos were not provided due to institutional policy.

Mashael Maashi - is an Associate Professor at the Department of Software Engineering, College of Computer and Information Sciences, King Saud University, Saudi Arabia. Her research area of interests include AI, machine learning, biomedical, IoT and cloud computing.

Hanan A. Mengash - is an Assistant Professor at the Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Saudi Arabia. Her research area of interests include machine learning, cybersecurity, IoT and cloud computing.

Alhanof Almutairi - is an Assistant Professor at the Department of Information Systems, College of Computer Science and Information Technology, King Faisal University, Saudi Arabia. Her research area of interests include machine learning, biomedical, energy, IoT and cloud computing.

Jahangir Khan - is an Assistant Professor at the Department of Computer Science, Applied College Mohyail Asir, King Khalid University, Kingdom of Saudi Arabia. His research interests include AI, IoT, smart cities and information security.

Turke Althobaiti - is an Assistant Professor at the Department of Computer Science, Faculty of Science, Northern Border University, Saudi Arabia. His current research interests include cybersecurity, AI, machine learning and computer networks.

Ahmed Alsayat - is an Associate Professor at the Department of Computer Science, College of Computer and Information Sciences, Jouf University, Saudi Arabia. His current research interests include cybersecurity, AI, machine learning and computer networks.