# Federated Learning-Empowered Disease Diagnosis Mechanism in the Internet of Medical Things: From the Privacy-Preservation Perspective

Xiaoding Wang, Jia Hu, Hui Lin, Wenxin Liu, Hyeonjoon Moon, and Md. Jalil Piran, *Senior Member, IEEE*

***Abstract*—The deep integration of the Internet of Things (IoT) and the medical industry has given birth to the Internet of Medical Things (IoMT). In IoMT, physicians treat a patient's disease by analyzing patient data collected through mobile devices with the assistance of an artificial intelligence (AI)-empowered systems. However, the traditional AI technologies may lead to the leakage of patient privacy data due to its own design flaws. As a privacy-preserving federated learning (FL) can generate a global disease diagnosis model through multiparty collaboration. However, FL is still unable to resist inference attacks. In this article, to address such problems, we propose a privacy-enhanced disease diagnosis mechanism using FL for IoMT. Specifically, we first reconstruct medical data through a variational autoencoder and add differential privacy noise to it to resist inference attacks. These data are then used to train local disease diagnosis models, thereby preserving patients' privacy. Furthermore, to encourage participation in FL, we propose an incentive mechanism to provide corresponding rewards to participants. Experiments are conducted on the arrhythmia database Massachusetts Institute of Technology and Beth Israel Hospital (MIT-BIH). The experimental results show that the proposed mechanism reduces the probability of reconstructing patient medical data while ensuring high-precision heart disease diagnosis.**

***Index Terms*—Disease diagnosis, federated learning (FL), Internet of Medical Things (IoMT), privacy protection.**

Manuscript received 1 April 2022; revised 18 August 2022; accepted 24 September 2022. Date of publication 30 September 2022; date of current version 20 June 2023. This work was supported in part by the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry and Fisheries (IPET) through Digital Breeding Transformation Technology Development Program, funded by Ministry of Agriculture, Food and Rural Affairs (MAFRA) under Grant 322063-03-1-SB010, and in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant 2020R1A6A1A03038540. Paper no. TII-22-1395. *(Corresponding author: Hui Lin; Md. Jalil Piran.)*

Xiaoding Wang, Hui Lin, and Wenxin Liu are with the College of Computer and Cyber Security, Fujian Normal University, Engineering Research Center of Cyber Security and Education Informatization, Fujian Province University, Fuzhou, Fujian 350117, China (e-mail: wangdin1982@fjnu.edu.cn; linhui@fjnu.edu.cn; sixwenxin@163.com).

Jia Hu is with the University of Exeter, EX4 4PY Exeter, U.K. (e-mail: j.hu@exeter.ac.uk).

Hyeonjoon Moon and Md. Jalil Piran are with the Department of Computer Software and Engineering, Sejong University, Seoul 05006, South Korea (e-mail: hmoon@sejong.ac.kr; piran@sejong.ac.kr).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TII.2022.3210597.

Digital Object Identifier 10.1109/TII.2022.3210597

## I. Introduction

NOWADAYS, technologies such as cloud computing, big data, the Internet of Things (IoT), and artificial intelligence (AI) have gradually penetrated the medical industry. With the help of these technologies, mobile medical care has gradually emerged, and it has begun to reconstruct the medical and health industry chain and service model. At the same time, various mobile smart terminals have been developed to provide people with standardized, information-based and professional health management, which has become a new trend in the medical and health industry [1]. With the gradual development of the IoT technology, mobile medical technology provides people with a new service model and medical experience, i.e., users monitor their own health status through smart terminals at home.

According to statistics in 2017, 60% of global healthcare organizations have implemented IoT solutions in their processes, while another 27% of organizations expect to adopt this technology in the short term. This is because IoMT technology can effectively save medical costs while meeting the ever-increasing demand for remote patient monitoring [2]. More importantly, the treatment of major diseases, i.e., heart attacks that cause more than 375 000 death in the US every year, will be significantly improved in IoMT [3]. For instance, connecting wearable medical devices through the IoMT can provide doctors and patients with data needed to better manage heart disease after diagnosis, and ultimately reduce heart disease-related mortality.

Medical wearable devices with AI can increase the speed and accuracy of heart disease diagnosis exponentially through remote monitoring [4], which gives birth to a promising AI-empowered patient monitoring architecture for IoMT. An instance AI-empowered patient monitoring model is represented in Fig. 1. Researchers train machine learning (ML) models and study the results of previous patient scans and data on whether the patient will continue to have a heart attack in the future to identify signs of heart disease, thereby increasing the diagnostic accuracy rate to 90%. Since AI systems require large amounts of data, collecting data from unreliable sources may adversely affect the effectiveness of AI solutions. However, what if these

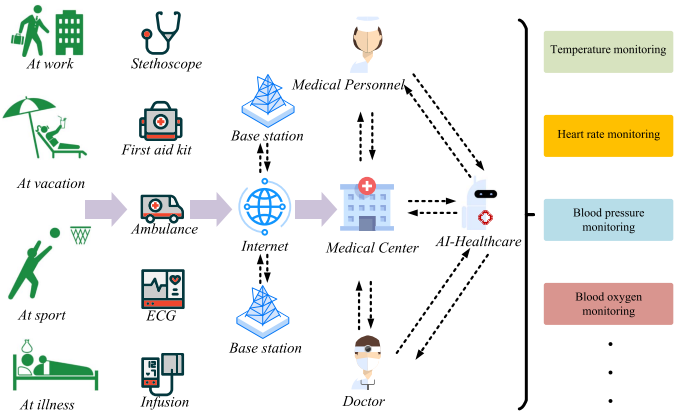


Fig. 1. AI-empowered patient monitoring architecture for IoMT.

data sources refuse to provide their data because of concerns about data leakage?

According to a report in "HIPAA Magazine," in 2020, the rate of medical data leakage has increased by 25% compared to 2019. Therefore, for hospitals, clinics and other medical institutions using the IoMT system, data security threats are still the biggest challenges [5]. To avoid this situation, the government and medical organizations have established regulations that every hospital must comply with the "Healthcare Information Portability and Accountability Act" and the "Economic and Clinical Health Information Technology Act". Although these regulations reduce the risk of patients' privacy leakage to a certain extent, due to its own shortcomings in traditional ML, patients' privacy data will still be leaked.

As one of the important technologies in the privacy computing system, federated learning (FL) [6], [7] is a mechanism that uses a central server to train a shared global model, while keeping all sensitive data in the local institution to which the data belongs, and it holds great promise for connecting decentralized medical data sources and privacy protection for IoMT. EHRs have become an important source of real-world medical data, used for important biomedical research, including ML research. FL is a viable way to connect healthcare provider EHR data, allowing healthcare providers to share experiences, not data, with guaranteed privacy. In these scenarios, the performance of ML models will be significantly improved by iteratively improving learning on large and diverse medical datasets. Several tasks have been studied in the FL setting in the medical domain, such as patient similarity learning across institutions, patient representation learning, spilt neural network, and predictive modeling [8].

FL also enables predictive modeling based on disparate sources, which can provide clinicians with more insights into the risks and benefits of treating patients early. For example, one case is using FL to predict patient resistance to certain treatments and drugs, and their survival rates for certain diseases, and another tested a privacy-preserving framework based on FL for predicting in-hospital deaths of patients admitted to the intensive care unit [9]. However, FL is subject to inference attacks [10], i.e., the data used for model training is reconstructed without any prior knowledge, which means that the privacy of FL participants may be leaked. If we design an incentive mechanism to provide corresponding rewards to the participants of FL, this can solve the problem of users no longer providing training data due to privacy leakage to a certain extent.

TABLE I
TABLE OF ACRONYMS

| Acronyms | Phrase |
|---|---|
| $IoT$ | Internet of Things |
| $IoMT$ | Internet of Medical Things |
| $AI$ | Artificial intelligence |
| $EHR$ | Electronic medical Record |
| $ECG$ | Electrocardiogram |
| $FL$ | Federated learning |
| $LSTM$ | Long short-term memory |
| $VAE$ | Variational autoEncoder |
| $DP$ | Differential privacy |

To address above problems, we propose a privacy-enhanced disease diagnosis mechanism using FL for IoMT. The main contributions of this article are summarized as follows.

1) To provide privacy protection in FL, we first reconstruct the patient data through a variational autoencoder (VAE), and add Laplace noise to the reconstructed data to achieve differential privacy protection for the patient data. On this basis, a disease diagnosis model is trained through FL. The global model trained on privacy-enhanced data can effectively resist inference attacks from adversaries.
2) To encourage the participation of FL, an incentive mechanism is designed to comprehensively evaluates the diagnosis models according to the quality, similarity, and richness of the training data, based on which participants are then given the corresponding rewards.
3) Simulation experiments and theoretical analysis are used to verify the performance of the proposed mechanism. The experimental results show that although our mechanism reduces the probability of patient medical data being reconstructed, due to the need to reconstruct and add noise to the data, the number of FL rounds will increase under the premise of ensuring the diagnostic accuracy, thereby increasing the calculation overhead as we expected. Furthermore, the proposed mechanism is able to achieve a tradeoff between privacy protection and accuracy of heart disease diagnosis, i.e., we manage to trade 6% of diagnosis accuracy for privacy protection on the MIT-BIH dataset.

The list of acronyms used in this article is given in Table I. The rest of this article is organized as follows. Related work is given in Section II. Section III introduces the system model and threat model. Section IV details the implementation of this mechanism. The performance evaluation is given in Section V. Finally, Section VI concludes the article.

## II. Related Work

FL, as a distributed ML method with privacy protection capability, is widely used in various fields. However, existing FL techniques are vulnerable to some secure threats, such as

poisoning attacks, backdoor attacks, and inference attacks [11], [12]. In response to these security threats, extensive research work have been conducted.

To resist poisoning attacks, Zhao et al. [13] used the generative adversarial networks (GANs) to generate an anomaly model detection mechanism for audit datasets to achieve poisoning attack mitigation. For backdoor attacks, Wu et al. [14] proposed a joint pruning method to remove redundant neurons in the network and mitigate backdoor attacks by integrating extreme weights of the model.

To alleviate the privacy leakage problem against inference attacks in FL, many privacy-preserved strategies have been proposed. Alamri et al. [16] designed a scheme to protect the privacy of medical data by using blockchain and smart contracts. This solution uploads the patient's encrypted private medical data to a cloud server, uses the recorded hash value as a data index, and stores it in a smart contract for patient access control.

Hamza et al. [17] developed a chaos-based privacy protection encryption system in IoMT to ensure the security of the medical key frames extracted from the endoscopy procedure. Alraja et al. [18] proposed a scheme to protect the privacy of user data in an IoMT. This scheme specifies the type and accuracy of data that can be shared, and compared the privacy risks and benefits from data sharing, so as to control data sharing and protect privacy. Can et al. [19] applied FL to the heart activity data collected by the smart bracelet, and deployed the scheme in the wearable biomedical monitoring system to monitor the stress levels in different events, thereby protecting the privacy of the data. Liu et al. [20] used the sparsity characteristics of the feature map in the network model to represent the raw local data of the participants to realize the privacy protection of the raw data.

Furthermore, various incentive mechanisms were proposed to mitigate the impact of the aforementioned security threats on FL models. Qu et al. [21] proposed an incentive mechanism based on the amount of data and sample size in the blockchain-based FL system. The FL equipment and miners can respectively obtain rewards that are linearly proportional to the amount of data provided and the amount of mining. Zhang et al. [22] proposed a method to evaluate the reputation of participants through the quality of the model. Specifically, the quality of the model was mainly evaluated by cross-entropy, the calculation execution time of the participants, and the amount of training data.

Although scholars have put forward a lot of excellent works to relieve the security threats suffered by FL, the existing models did not combine the design of incentive mechanism with the protection of data privacy of FL participants. In addition, the richness of participant data was not considered when the incentive mechanism is designed. Therefore, this article jointly considers privacy protection, incentive mechanism, and data richness to improve the enthusiasm of patients to contribute medical data, thereby improving the accuracy of disease diagnosis models.

## III. System Model and Security Model

### A. System Model

In this article, we aim to realize the privacy-preserving disease diagnosis for IoMT using FL. To achieve this goal, we consider the following entities: patients, medical centers, doctors, evaluation server, hospital, and adversary.

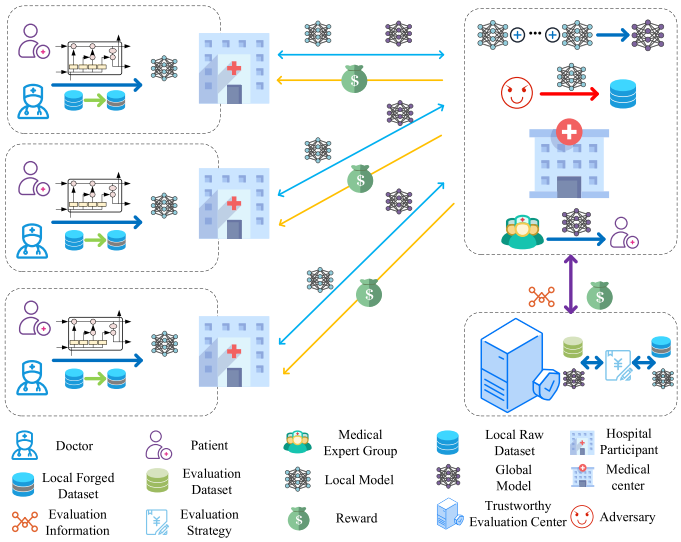


Fig. 2. Proposed system model.

1) Patient: A person who has a disease and wants to go to the hospital for a diagnosis.
2) Medical center: An institution used to share diagnostic information among different hospitals or doctors.
3) Doctor: An expert who makes a medical diagnosis of a patient through the patient's information.
4) Evaluation server: A device that evaluates and rewards diagnostic models uploaded by different hospitals.
5) Hospital: An institution that trains local disease diagnosis models from doctors' diagnostic information and patient information.
6) Adversary: Someone who launches an inference attack to reconstruct a patient's private data.

The system model is shown in Fig. 2. An important reference index for doctors to diagnose heart disease is the patient's ECG. Whereas the ECG signal is time series data, the LSTM network can well analyze time series related data and learn the relationship between them. Therefore, analyzing a patient's ECG through LSTM can help doctors diagnose the patient's symptoms [23], [24]. In order to protect patient privacy, the hospital reconstructs the patient's ECG through a VAE [15], which is used to train an LSTM-based heart disease diagnosis model. Each hospital uploads different diagnostic models to the medical center, and the medical center aggregates through FL [25] to generate a global diagnostic model. During this process, the evaluation server will request the medical center for diagnostic models of different hospitals and evaluate them. Based on this, the medical center pays different hospitals for contributing diagnostic models.

### B. Security Model

In the process of FL, there is a possibility that the privacy data of participants may be leaked. In this article, the leaked data are ECG data of different cardiac patients, and related work shows that adversaries can infer and reconstruct training data from gradients or models without any prior knowledge about the training data [10]. Furthermore, we assume that the adversary

TABLE II
NOTATIONS AND EXPLANATIONS

| Notations | Explanations |
|---|---|
| $p_i$ | Participant of FL |
| $D_i$ | Original dataset of $p_i$ |
| $\hat{D}_i$ | Reconstructed dataset of $p_i$ |
| $D_i^{'}$ | Privacy-enhanced dataset of $p_i$ |
| $z_j$ | Latent variable of sample $x_j$ |
| $\mu_j$ | Mean of latent variable $z_j$ |
| $\sigma_j^2$ | Variance of latent variable $z_j$ |
| $\epsilon$ | Privacy budget |
| $\delta$ | Confidence parameter |
| $\Delta f$ | Sensitivity |
| $M_k^t$ | Diagnosis model of hospital $k$ at round $t$ |
| $Quantity_i$ | Evaluation accuracy of local model of $p_i$ |
| $Similarity_i$ | Feature difference of the same type data about $p_i$ |
| $Richness_i$ | Normalized relevant category number of dataset of $p_i$ |

**Algorithm 1:** Privacy-Enhanced Disease Diagnosis Mechanism Using FL.

1: Execute the privacy-enhanced disease diagnosis model generation by Algorithm 2
2: Execute the incentive mechanism by Algorithm 3

is capable of white-box inference [26], i.e., the adversary can access the local model uploaded by any participant.

The FL model considered in this article is LSTM, and the training dataset is a time series medical dataset, which is different from the traditional convolutional neural network model. For the inference attack, the ECG data used to train the LSTM model first needs to be reconstructed by VAE, and the differential privacy noise is further added. On this basis, FL can effectively protect the privacy of patients.

## IV. IMPLEMENTATION DETAILS OF THE PROPOSED MECHANISM

A number of studies have shown that the training process of FL involves the risk of dataset being recovered from the model causing the privacy leakage. In response to this problem, the privacy protection of patient data are achieved through VAE-based data reconstruction followed by Laplacian noise injection. Meanwhile, an incentive mechanism is designed based on the dataset quantity, richness, and similarity to encourage participation. The privacy-enhanced disease diagnosis mechanism using FL is summarized in Algorithm 1. The main notations and explanations used in the proposed mechanism are listed in Table II.

### A. Privacy-Enhanced Disease Diagnosis Model Generation

*1) Privacy-Enhancement on Patients' Data:* For a participant $p_i$ in FL, the original dataset can be represented by $D_i = \{(x_1, y_1), \ldots, (x_j, y_j)\}$. To achieve privacy protection, we consider the VAE a mapping function $VAE(\cdot)$ from the original dataset to the reconstructed dataset as follows:

$$\hat{D}_i \leftarrow VAE\left(D_i\right), \tag{1}$$

where $\hat{D}_i$ represents the reconstructed dataset with $\hat{D}_i = \{(\hat{x}_1, \hat{y}_2), \ldots, (\hat{x}_j, \hat{y}_j)\}$.

This is because the VAE consists of an encoder $Enc$ and a decoder $Dec$. When we train the VAE, the training sample $x_j$ is encoded as a set of low-dimensional vectors by the encoder $Enc$, part of which is fitted to the mean $\mu_j$ of the low-dimensional latent variable, denoted by $z_j$, distribution of the sample $x_j$, and the other part is fitted to the variance $\sigma_j^2$ of $z_j$, while the decoder $Dec$ restores the sampling result from the distribution $(\mu_j, \sigma_j^2)$ to a high-dimensional generated sample $\hat{x}_j$. And our goal is to minimize the objective function $\mathcal{L}$ on the generated sample $\hat{x}_j$, training samples $x_j$ and latent variable $z_j$, i.e., $\mathcal{L} = RE(x_j, \hat{x}_j) + KL(P(z_j|x_j)||N(0, I))$, where $P(\cdot|\cdot)$ is the distribution of the latent variable exclusive to the sample, $RE(\cdot, \cdot)$ represents the reconstruction error, $N(0, I)$ represents standard normal distribution, and $KL(\cdot)$ represents the Kullback–Leibler divergence.

To further improve the privacy protection, we use the Laplace mechanism to achieve differential privacy protection for the reconstructed datasets because the Laplace mechanism provides a strict $(\epsilon, 0)$ differential privacy, while the Gaussian mechanism provides a relaxed $(\epsilon, \delta)$ differential privacy. Specifically, we add the $noise$ satisfying the Laplacian distribution to the reconstructed dataset to achieve differential privacy protection as follows:

$$D_i' = \hat{D}_i + noise, \tag{2}$$

$$s.t. noise \sim Laplace\left(0, \frac{\Delta f}{\epsilon}\right), \tag{3}$$

where the $\epsilon$ denotes privacy budget, the $\Delta f$ denotes sensitivity. Then, the reconstructed dataset after applying differential privacy protection is denoted by $D_i' = \{(x_1', y_2'), \ldots, (x_j', y_j')\}$.

Note that adding Laplacian noise when training the local model is to achieve differential privacy protection for the reconstructed dataset and further prevent the original dataset from being maliciously reconstructed by adversaries.

*2) Privacy-Enhanced FL:* Participant $p_i$ download global model $M^t$ of the $t$th round FL from the model aggregation server, and use the gradient descent algorithm to train the $t+1$th round new local model $M_i^{t+1}$ with the reconstructed data $D_i'$ as follows:

$$M_i^{t+1} = M^t - \eta \times \nabla \mathcal{L}(M^t; D_{i,\text{batch}}'), \tag{4}$$

where $\mathcal{L}(\cdot;\cdot)$ represents the local loss function (i.e., the LSTM in our case), $D_{i,\text{batch}}'$ denotes batches of data sampled from $D_i'$, and $\eta$ is the learning rate of local model training.

When all participants complete local training, the server select $N$ participants and aggregate their trained local models using the Fedavg algorithm to generate the new global model $M^{t+1}$

**Algorithm 2:** Privacy-Enhanced Disease Diagnosis Model Generation.

**Input:** $K$ hospitals indexed by $k$, $E$ is the number of local epochs, and $\eta$ is the learning rate
**Output:** global diagnosis model $M$

Reconstruct patients' data $\hat{D}_i$ from the original data $D_i$ by (1)
Obtain differential privacy protected data $D_i'$ from $\hat{D}_i$ by (2)
$\mathbf{ServerinMedicalCenterexecutes}$:
initialize $M^0$
**for** each round $t = 1, 2, \ldots$ **do**
  **for** each hospital $k$ **in parallel do**
    $M_k^{t+1} \leftarrow \mathbf{HospitalUpdate}(k, M^t)$
    $M^{t+1} \leftarrow \frac{1}{n}\sum_i^n M_i^{t+1}$
  **end for**
**end for**
$\mathbf{HospitalUpdate}(\mathbf{k}, \mathbf{M})$://Run on Hospital $k$
**for** each local epoch $j$ from 1 to $E$ **do**
  **for** each batch $b \in D_i'$ **do**
    $M \leftarrow M - \eta \times \nabla \mathcal{L}(M; b)$
  **end for**
**end for**

as follows:

$$M^{t+1} = \frac{1}{n}\sum_{i}^{n} M_i^{t+1}. \tag{5}$$

When the aggregation server completes the model aggregation, the participants redownload the global model and perform a new round of federated training until the end condition is reached. In each round of FL, the aggregation server sends the required data (including test data, local model, global model, etc.) to the trusted model evaluation server.

We summarize the generation of the privacy-enhanced global disease diagnosis model in Algorithm 2.

Since the VAE learns the distribution of the low-dimensional latent variable space of the training samples, the data reconstructed by the VAE has the same feature distribution as the training data. This means that training the model with the reconstructed data can reduce the risk of the participants' real data being reconstructed. In addition, we add appropriate Gaussian noise to the patient's data to provide differential privacy protection. As FL progresses, this noise-induced error decreases with iterations of FL rounds. Since privacy and accuracy cannot be improved at the same time, this means that we need to find a balance between the two, i.e, to enhance privacy as much as possible while reducing the impact of reconstructed data on the reliability of the global model. At the same time, we design a data-based three-incentive mechanism for evaluating tuple features.

## B. Incentives

To encourage FL participants to actively contribute to the local model, we propose a data feature-based incentive mechanism to motivate more participants to contribute to the local model. Our motivation for proposing this incentive mechanism is that FL exploits the data features of different actors to improve the generalization ability of the global model, which is the key to generating the global model. In addition, we introduce a trusted third-party evaluation server to evaluate the local training situation of different participants and calculate the corresponding rewards. In the FL task, the evaluation server computes the corresponding evaluation based on the evaluation dataset provided by the model aggregation server and the evaluation data provided by the participants. Therefore, the reward is calculated based on the evaluation results, independent of the number of participants.

In the incentive mechanism, the reward for the participant $p_i$ consists of two parts, namely basic reward $Reward_{i,\text{basic}}$ and extra reward $Reward_{i,\text{extra}}$. The basic reward is accumulated after each round of successful submission of the local model by the participants, so the final basic reward can be calculated as follows:

$$Reward_{i,\text{basic}} = success\_num \times Reward_{\text{basic}}^{\text{iter}}, \tag{6}$$

where $success\_num$ represents the number of times the participant has successfully uploaded the model, and $Reward_{\text{basic}}^{\text{iter}}$ represents the basic reward for each round of FL.

For the extra reward calculation, we consider the local dataset held by the participant $p_i$, denoted by $C_i^l$, has the form as $C_i^l = \{c_i^1, c_i^2, \ldots, c_i^x\}$, where $c_i^x$ represents a certain type of dataset in the participant $p_i$ dataset, and $x$ represents the total number of categories in the dataset. And the test dataset held by the aggregation server, denoted by $C_s$, has the similar form as $C_s = \{c_s^1, c_s^2, \ldots, c_s^X\}$, where $c_s^X$ represents a certain type of dataset in the test dataset in the aggregation server, $X$ represents the total number of categories in the dataset. On this basis, we use $C^{\text{same}}$ to represent how revelent the participant dataset to the aggregation server test dataset, i.e., $C^{\text{same}} = C_i^l \cap C_s$.

When the amount of data in the training model is insufficient, the trained model is easy to over-fit. When other datasets are used for testing, the prediction accuracy will be at a low level. So when a model is less accurate in prediction, the number of datasets that train this model might be small. For the participant $p_i$, in a certain round of FL, the average prediction accuracy of the uploaded local model, denoted by $Avgacc_i^{\text{iter}}$, can be calculated by $Avgacc_i^{\text{iter}} = \frac{1}{X_{\text{same}}}\sum_j^{X_{\text{same}}} acc_j^{\text{iter}}$, where $acc_j^{\text{iter}}$ represents the prediction accuracy of the same type of dataset, and $X_{\text{same}}$ represents the number of elements in the set $C^{\text{same}}$.

When the FL ends, we calculate the evaluation accuracy of the local model $Quantity_i$ provided by participant $p_i$ as follows:

$$Quantity_i = \frac{1}{success\_num}\sum_{iter}^{success\_num} Avgacc_i^{\text{iter}}, \tag{7}$$

where $success\_num$ represents the number of times that the participant successfully uploaded the local model during the FL process. Then, we define the normalized relevant number of categories of dataset of $p_i$ to train the local model by $Richness_i$. Thereby, we have

$$Richness_i = \frac{X_{\text{same}}}{X}. \tag{8}$$

**Algorithm 3:** Incentive Mechanism.

**Input:** test data, local model $M_i^t$, global model $M^t$
**Output:** total reward $Reward_{i,total}$ for each participant $p_i$

1: Calculate the basic reward $Reward_{i,basic}$ by (6)
2: Calculate the extra reward $Reward_{i,extra}$ based on $Quantity_i$, $Similarity_i$, $Richness_i$ and $Reward_{basic}$ by (7)–(10)
3: Calculate the total reward by for each participant $p_i$ by $Reward_{i,total} = Reward_{i,basic} + Reward_{i,extra}$

We also define the feature difference of the same type of data between the local training dataset and the evaluation dataset as $Similarity_i$, i.e,

$$Similarity_i = \frac{1}{X_{\text{same}}} \sum_{j}^{X_{\text{same}}} \left( AF\left(c_i^j\right) - AF\left(c_s^j\right) \right)^2, \quad (9)$$

where $AF(\cdot)$ represents the average feature value of a certain type of dataset. As a result, the extra reward for the participant $p_i$ can be computed by

$$Reward_{i,\text{extra}} = Reward_{i,\text{basic}} \times \left( \frac{Quantity_i + Richness_i}{\rho + Similarity_i} \right). \quad (10)$$

Obviously, the total reward for the participant $p_i$, denoted by $Reward_{i,\text{total}}$, equals to the sum of the basic reward $Reward_{i,\text{basic}}$ and the extra reward $Reward_{i,\text{extra}}$. We summarize the incentive mechanism in Algorithm 3.

### C. Security Analysis

Different from ciphertext-only attack, known-plaintext attack, chosen-plaintext attack, and chosen-ciphertext attack, which recover the key from the plain-ciphertext pairs to crack the ciphertexts, this article focuses on inference attacks launched by adversaries. These attacks attempt to recover the patient's private information from the disease diagnosis model. To protect patients' privacy, this article proposes a privacy-enhanced disease diagnosis mechanism based on FL against inference attacks. Since the patient's data are reconstructed by a VAE, and Laplace noise is added to the reconstructed data, differential privacy protection can be added to the data. Using these data to train a local disease diagnosis model, and on this basis, train a global disease diagnosis model through FL. According to the post-processing property of differential privacy, we can provide differential privacy protection for the generated global model. Since the local model is trained on the reconstructed and noisy data, even if the adversary restores the patient data from the local model through inference attacks, according to the nature of the VAE and the principle of differential privacy, these data might have similar distributions with respect to certain characteristics as the patient's original data while the data themselves are different. In addition, the ultimate goal of ciphertext-only attack, known-plaintext attack, chosen-plaintext attack, and chosen-ciphertext attack is to crack the ciphertext to obtain the plaintext. In contrast, since our proposed mechanism is able to provide differential privacy protection, according to the above analysis, it can be deduced that it is difficult for attackers to obtain the real data of patients, thereby protecting their privacy as above four attacks.

## V. Experiment

### A. Experiment Setup

This section comprehensively evaluates the proposed mechanism through the scientific computing libraries Tensorflow in python. The experimental environment is configured on the computer of Intel(R) Core(TM) i5-10300H CPU @ 2.50 GHz, and the version of Tensorflow used is 2.3.1.

In this experiment, we use the MIT-BIH database, which is the generally available set of standard test material for evaluation of arrhythmia detectors, which contains 48 half-hour excerpts of two-channel Holter recordings from 47 subjects studied in the BIH Arrhythmia Laboratory between 1975 and 1979 [27]. This dataset has been used for that purpose as well as for basic research into cardiac dynamics at more than 500 sites worldwide. The dataset contains ECGs of the following five types of heartbeats:

1) Normal (N);
2) Supraventricular ectopic beat (S);
3) Ventricular ectopic beat (V);
4) Fusion beat (F);
5) Unknown beat (Q).

To evaluate the proposed mechanism, we consider a three-party FL system.

We adopt the same method as that used in [28] to process the dataset. First, the ECG signal is decomposed into six levels of db4 wavelet basis functions for noise reduction. Then, features are directly extracted from the morphology of ECG signal in the time domain, mainly located characteristic points, i.e., the peak of P wave, QRS complex, and the peak of T wave, recorded as P, Q, R, S, T, and the intervals and statistics features are used. Next, the beat annotations contained in the database are used as initial evidence for the segmentation stage, treating them as R peaks. According to the mean RR interval, 499 samples from the left side of the QRS mid-point, 500 samples after QRS mid-point, and the QRS mid-point itself were chosen as a segment, thereby at least two cardiac cycles are included. At last, the raw data are standardized and mapped to $[0-1]$ by a linear transformation.

In our mechanism, each participant's private dataset is a part of the MITBIH dataset. The global model refers to the model generated by the aggregation server through aggregating local models in each round of FL. And the local model is LSTM. The input shape of this model is (1,187), and it contains an LSTM layer and a dense layer. The shape of the kernel in the LSTM cell is (187, 200), the shape of the recurrent_kernel is (50, 200), the activation function is Tanh, and the shape of the kernel in the dense is (50, 5), and the activation function is Softmax. Among them, the recurrent kernel is the set of each gate parameter in the LSTM cell. We also define the prediction accuracy of the global model and the local model in FL as the global accuracy and the local accuracy, respectively.

### B. Experiment Result

First, we test the VAE model in the reconstructed training data and the results are shown in Fig. 3. As the number of training rounds increases, the VAE model learns more fully the latent variable space of the training data, and the training loss of the model gradually decreases until it converges. The numerical results show that the VAE model can be trained normally even with the addition of random samples.

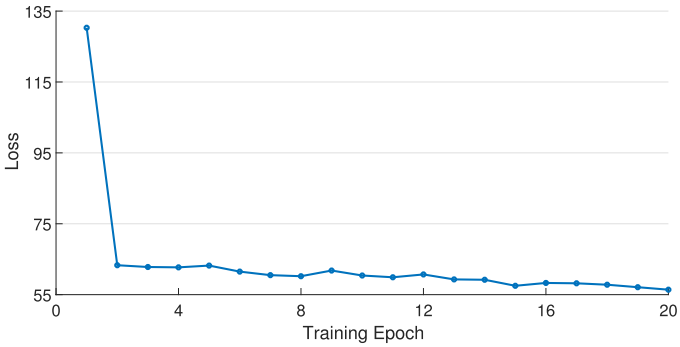


Fig. 3. Training loss of VAE.

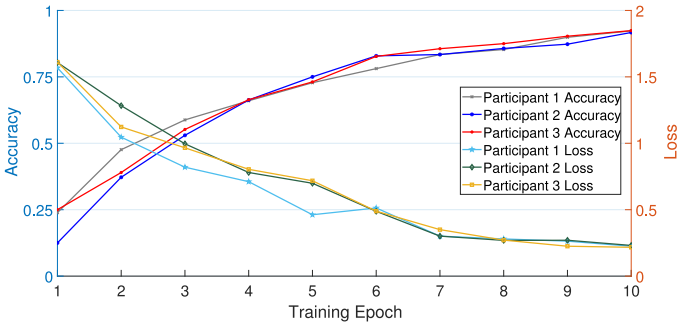


Fig. 4. Participants' local training.

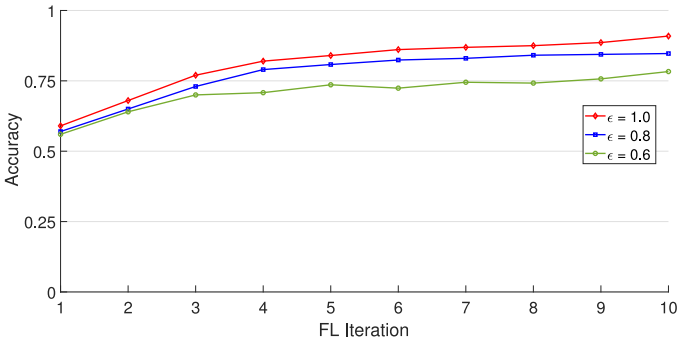


Fig. 5. Influence of the privacy budget $\epsilon$ on the global model.

We then evaluated three participants for local training on the reconstructed data. As shown in Fig. 4, as the number of local training rounds of the participants increases, the model gradually learns the distribution characteristics of the training data, so the model can predict the training samples as the corresponding correct labels, so the loss will gradually decrease until the algorithm converges. In addition, as the loss of the training model gradually decreases, the prediction accuracy of the local model will gradually improve. For different participants, the prediction accuracy of local model training convergence can reach about 90%.

Next, we conduct experiments on the global prediction accuracy, when the differential privacy budget $\epsilon$ is 1.0, 0.8, and 0.6 in the FL system. Fig. 5 shows the simulation results. As shown in the figure, as the privacy budget gradually increases,

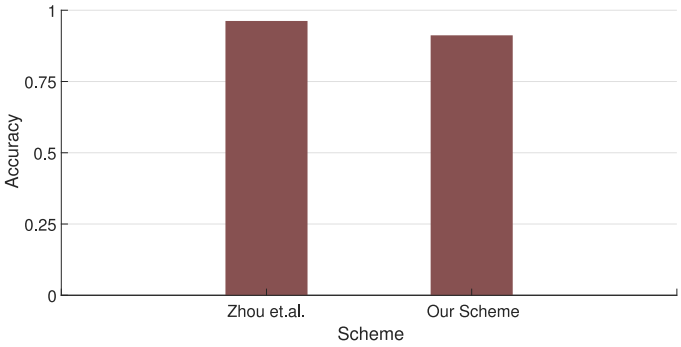


Fig. 6. Global model accuracy comparison between different mechanisms.

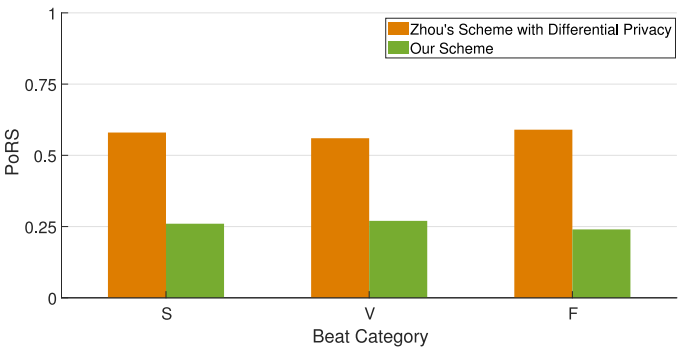


Fig. 7. Probability of successful data reconstruction.

the prediction accuracy of the global model after 10 rounds of FL also gradually decreases. When $\epsilon$ is 1.0, the prediction accuracy of the global model is 90.9%; when $\epsilon$ is 0.8, the prediction accuracy of the global model is 84.7%; when $\epsilon$ is 0.6, the prediction accuracy of the global model is 78.3%. This is because with the gradual increase of the differential privacy budget $\epsilon$, the Gaussian noise added before local training also increases gradually, and more noise will affect the prediction accuracy of the FL global model.

We compare the proposed mechanism with that of Zhou et al. [28] in terms of prediction accuracy and the results are shown in Fig. 6. As shown in the figure, the prediction accuracy of the mechanism proposed by Zhou et al. in the MITBIH dataset is around 97%, while our mechanism is around 91%. Since privacy and prediction accuracy cannot have both, we trade a small fraction of model prediction accuracy for privacy protection of real data. Probability of reconstruction success (PoRS) is used to evaluate the privacy-preserving ability of our mechanism. We compare the proposed mechanism with Zhou's mechanism with DP [28]. We let the differential privacy mechanism be implemented by adding $(\mu = 0, b = 1)$ Gaussian noise to the training dataset. We simulated PoRS on the ECGs of S-, V-, and F-types in the MIT-BIH dataset, and the results are shown in Fig. 7. The differential privacy protection of our mechanism is accomplished in the data generated by the VAE, which can reduce the probability that the real dataset of the participants is reconstructed compared with the real data. The numerical results show that this scheme can reduce the reconstruction success rate

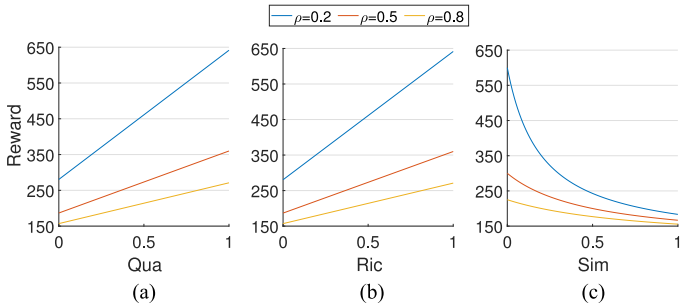


Fig. 8. Incentives evaluation.

of the original training dataset by about 50% compared with Zhou's with DP.

The incentive mechanism is evaluated and the experimental results are shown in Fig. 8, with the value of the basic reward obtained by the participants is set to 100. In Fig. 8(a), the data richness $Ric$ and similarity $Sim$ are fixed at 0.5 and 0.07, respectively, and the total reward obtained by the participants will become higher and higher with the increase of the data quality $Qua$. In Fig. 8(b), the data quality $Qua$ and similarity $Sim$ are fixed at 0.5 and 0.07, respectively, and the total reward obtained by the participants will become higher and higher with the increase of the data richness $Ric$. In Fig. 8(c), the data quality $Qua$ and the richness $Ric$ are fixed at 0.5 and 0.5, respectively. The more similar the feature distribution between the data, the smaller the value of the similarity $Sim$, so the total reward obtained by the participants will increase with the data similarity. In addition, we find that the change of the total reward of the participants is negatively correlated with the reward factor $\rho$. Numerical results show that the incentive mechanism can comprehensively evaluate and incentivize local training and local models according to the triple characteristics of the training data.

## VI. Conclusion

In IoMT, patients' data collected by mobile smart terminals are systematically analyzed through AI technology to assist doctors for patients' diseases diagnosis. Since traditional AI technology may cause the privacy leakage of patients, FL emerges as a privacy-protected and multiparty collaborative ML model. However, FL is subject to inference attacks. To solve above problems, we proposed a privacy-enhanced disease diagnosis mechanism using FL for IoMT. Specifically, we reconstructed patient data through VAE, add differential privacy noise, and train a disease diagnosis model through FL on this basis. In addition, we designed an incentive mechanism to encourage patients to provide medical data. We conduct experiments on the MIT-BIH arrhythmia database, and the experimental results show that the proposed mechanism guarantees high accuracy for heart disease diagnosis and low success rate for adversarial inference attacks. The accuracy of the global disease diagnosis model depends to a certain extent on the process of local model aggregation in FL. Our future research direction will be how to improve the accuracy of the global model through adaptive aggregation weight adjustment in FL.

## References

[1] X. Xu et al., "Privacy-preserving federated depression detection from multi-source mobile health data," *IEEE Trans. Ind. Informat.*, vol. 18, no. 7, pp. 4788–4797, Jul. 2022, doi: 10.1109/TII.2021.3113708.
[2] C. Guo, P. Tian, and K. K. R. Choo, "Enabling privacy-assured fog-based data aggregation in E-healthcare systems," *IEEE Trans. Ind. Informat.*, vol. 17, no. 3, pp. 1948–1957, Mar. 2021.
[3] X. Yuan, J. Chen, K. Zhang, Y. Wu, and T. Yang, "A stable AI-based binary and multiple class heart disease prediction model for IoMT," *IEEE Trans. Ind. Informat.*, vol. 18, no. 3, pp. 2032–2040, Mar. 2022, doi: 10.1109/TII.2021.3098306.
[4] K. Guo et al., "MDMaaS: Medical-assisted diagnosis model as a service with artificial intelligence and trust," *IEEE Trans. Ind. Informat.*, vol. 16, no. 3, pp. 2102–2114, Mar. 2020.
[5] Y. Sun, J. Liu, K. Yu, M. Alazab, and K. Lin, "PMRSS: Privacy-preserving medical record searching scheme for intelligent diagnosis in IoT healthcare," *IEEE Trans. Ind. Informat.*, vol. 18, no. 3, pp. 1981–1990, Mar. 2022, doi: 10.1109/TII.2021.3070544.
[6] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," 2016, *arXiv:1610.05492*.
[7] J. Chen, J. Zhang, Y. Zhao, H. Han, K. Zhu, and B. Chen, "Beyond model-level membership privacy leakage: An adversarial approach in federated learning," in *Proc. 29th Int. Conf. Comput. Commun. Netw.*, 2020, pp. 1–9.
[8] J. Xu et al., "Federated learning for healthcare informatics," *J. Healthcare Informat. Res.*, vol. 5, no. 1, pp. 1–19, 2021.
[9] D. C. Nguyen et al., "Federated learning for smart healthcare: A survey," *ACM Comput. Surv.*, vol. 55, no. 3, pp. 1–37, 2022.
[10] L. Zhu and S. Han, "Deep leakage from gradients," in *Federated Learning*. Cham, Switzerland: Springer, 2020, pp. 17–31.
[11] M. Alazab, S. P. RM, P. M, P. K. R. Maddikunta, T. R. Gadekallu, and Q.-V. Pham, "Federated learning for cybersecurity: Concepts, challenges, and future directions," *IEEE Trans. Ind. Informat.*, vol. 18, no. 5, pp. 3501–3509, May 2022, doi: 10.1109/TII.2021.3119038.
[12] B. Ghimire and D. B. Rawat, "Recent advances on federated learning for cybersecurity and cybersecurity for federated learning for internet of Things," *IEEE Internet Things J.*, vol. 9, no. 11, pp. 8229–8249, Jun. 2022.
[13] Y. Zhao et al., "Detecting and mitigating poisoning attacks in federated learning using generative adversarial networks," *Concurrency Comput.: Pract. Exp.*, vol. 34, 2020, Art. no. e5906.
[14] C. Wu et al., "Mitigating backdoor attacks in federated learning," 2020, *arXiv:2011.01767*.
[15] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," 2013, *arXiv:1312.6114*.
[16] B. Alamri, I. T. Javed, and T. Margaria, "Preserving patients' privacy in medical IoT using blockchain," in *Proc. Int. Conf. Edge Comput.*, Cham, 2020, pp. 103–110.
[17] R. Hamza, Z. Yan, K. Muhammad, P. Bellavista, and F. Titouna, "A privacy-preserving cryptosystem for IoT e-healthcare," *Inf. Sci.*, vol. 527, pp. 493–510, 2020.
[18] M. N. Alraja, H. Barhamgi, A. Rattrout, and M. Barhamgi, "An integrated framework for privacy protection in IoT - Applied to smart healthcare," *Comput. Elect. Eng.*, vol. 91, 2021, Art. no. 107060.
[19] Y. S. Can and C. Ersoy, "Privacy-preserving federated deep learning for wearable IoT-based biomedical monitoring," *ACM Trans. Internet Technol.*, vol. 21, no. 1, pp. 1–17, 2021.
[20] B. Liu, Y. Guo, and X. Chen, "PFA: Privacy-preserving federated adaptation for effective model personalization," in *Proc. Web Conf.*, 2021, , pp. 923–934.
[21] Y. Qu et al., "Decentralized privacy using blockchain-enabled federated learning in fog computing," *IEEE Internet Things J.*, vol. 7, no. 6, pp. 5171–5183, Jun. 2020.
[22] Q. Zhang, Q. Ding, J. Zhu, and D. Li, "Blockchain empowered reliable federated learning by worker selection: A trustworthy reputation evaluation method," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops*, 2021, pp. 1–6.
[23] P. Zhang et al., "A united CNN-LSTM algorithm combining RR wave signals to detect arrhythmia in the 5G-Enabled medical Internet of Things," *IEEE Internet Things J.*, vol. 9, no. 16, pp. 14563–14571, Aug. 2022.
[24] R. He et al., "Automatic detection of QRS complexes using dual channels based on U-net and bidirectional long short-term memory," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 4, pp. 1052–1061, Apr. 2021.

[25] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. Aguera y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. 20th Int. Conf. Artif. Intell. Statist.*, Lauderdale, FL, USA, 2017, pp. 1273–1282.
[26] H. Hu, Z. Salcic, G. Dobbie, and X. Zhang, "Membership inference attacks on machine learning: A survey," *ACM Comput. Surveys*, vol. 54, no. 11, pp. 1–37, 2022.
[27] A. L. Goldberger et al., "Components of a new research resource for complex physiologic signals," *PhysioBank, PhysioToolkit, PhysioNet*, 2000.
[28] R. Zhou et al., "Arrhythmia recognition and classification through deep learning-based approach," *Int. J. Comput. Sci. Eng.*, vol. 19, no. 4, pp. 506–517, 2019.

**Xiaoding Wang** received the Ph.D. degree from the College of Mathematics and Informatics, Fujian Normal University, Fuzhou, China, in 2016. He is currently an Associate Professor with the College of Computer and Cyber Security, Fujian Normal University. His research interests include network optimization and fault tolerance.

**Hui Lin** received the Ph.D. degree in computing system architecture from the College of Computer Science, Xidian University, China, in 2013.

He is currently a Professor with the College of Computer and Cyber Security, Fujian Normal University, Fuzhou, China. He is also an M.E. Supervisor with the College of Computer and Cyber Security, Fujian Normal University, Fuzhou, China. He has authored or coauthored more than 50 papers in international journals and conferences. His research interests include mobile cloud computing systems, blockchain, and network security.

**Jia Hu** received the B.Eng. and M.Eng. degrees in electronic engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2006 and 2004, respectively, and the Ph.D. degree in computer science from the University of Bradford, Bradford, U.K., in 2010.

He is currently a Senior Lecturer in computer science with the University of Exeter, Exeter, U.K. His research interests include edge-cloud computing, resource optimization, applied machine learning, and network security. He has authored or coauthored more than 90 research papers published in prestigious international journals and reputable international conferences.

Dr. Hu is on the Editorial Board of Elsevier *Computers and Electrical Engineering* and has guest-edited many special issues on major international journals (e.g., *IEEE Internet of Things Journal*, *Computer Networks*, and *Ad Hoc Networks*). He was the General Co-Chair of IEEE CIT'15 and IUCC'21, and Program Co-Chair of IEEE ISPA'20, ScalCom'19, SmartCity'18, CYBCONF'17, and EAI SmartGIFT'2016. He was the recipient of the Best Paper Awards at IEEE SOSE'16 and IUCC'14

**Wenxin Liu** received the bachelor's degree in information security from the Xi'an University of Posts and Telecommunications, Xi'an, China, in 2019. He is currently working toward the master's degree with the College of Computer and Cyber Security, Fujian Normal University, Fuzhou, China.

His research interests include deep learning, cyber security, and blockchain.

**Hyeonjoon Moon** received the B.S. degree in electronics and computer engineering from Korea University, Seoul, South Korea, in 1990, and the M.S. and Ph.D. degrees in electrical and computer engineering from the State University of New York at Buffalo, Buffalo, NY, USA, in 1992 and 1999, respectively.

From January 1996 to October 1999, he was a Senior Researcher with the Electro-Optics/Infrared Image Processing Branch, U.S. Army Research Laboratory (ARL), Comment: Author: Please provide the subject in which the author Xiaoding Wang received the Ph.D. degree. Adelphi, MD, USA. He developed a face recognition system evaluation methodology based on the face recognition technology Program. From November 1999 to February 2003, he was a Principal Research Scientist with Viisage Technology, Littleton, MA, USA. He has extensive background on still image and real-time video-based computer vision and pattern recognition. Since March 2004, he has been with the Department of Computer Science and Engineering, Sejong University, where he is currently a Professor and the Chairperson. His research focuses on the research and development of real-time facial recognition system for access control, surveillance, and big database applications. His current research interests include image processing, biometrics, artificial intelligence, and machine learning.

**Md. Jalil Piran** (Senior Member, IEEE) received the Ph.D. degree in electronics and information engineering from Kyung Hee University, Seoul, South Korea, in 2016. He then continued his research career as a Postdoctoral Fellow with Networking Laboratory, Kyung Hee University. He is currently an Assistant Professor with the Department of Computer Science and Engineering, Sejong University, Seoul, South Korea. He has authored a substantial number of technical papers in well-known international journals and conferences in the field of intelligent information and communication technology, specifically in the fields of machine learning, data science, wireless communications and networking, 5G/6G, Internet of Things, and cyber security.

Dr. Piran is an Editor of various journals, including IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE ACCESS, Elsevier *Journal of Physical Communication*, and Elsevier Journal of Computer Communication. Moreover, he was the Chair of the 5G and Beyond Communications Session at the 2022 IEEE International Conference on Communications (ICC), and a Technical Committee Member of Several Conferences. In the worldwide community, he is an Active Delegate from South Korea to the Moving Picture Experts Group (MPEG). He was the recipient of the IAAM "Scientist Medal of the Year 2017" award for notable and outstanding research in new age technology and innovation, in Stockholm, Sweden. In 2017, he was recognized by the Iranian Ministry of Science, Technology, and Research as an Outstanding Emerging Researcher. In addition, his Ph.D. dissertation has been selected as the Dissertation of the Year 2016 by the Iranian Academic Center for Education, Culture, and Research in the Engineering Group.