

Research Article

Personalized Federated Learning for ECG Classification Based on Feature Alignment

Renjie Tang ^{1,2}, Junzhou Luo ¹, Junbo Qian ², and Jiahui Jin ¹

¹School of Computer Science and Engineering, Southeast University, Nanjing, Jiangsu Province 211189, China

²China Mobile Group Zhejiang Co., Ltd., Hangzhou, Zhejiang Province 310016, China

Correspondence should be addressed to Renjie Tang; tangrj@seu.edu.cn

Received 22 June 2021; Revised 15 September 2021; Accepted 4 October 2021; Published 1 November 2021

Academic Editor: Lianbo Ma

Copyright © 2021 Renjie Tang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Electrocardiogram (ECG) data classification is a hot research area for its application in medical information processing. However, insufficient data, privacy preserve, and local deployment are still challenging difficulties. To address these problems, a novel personalized federated learning method for ECG classification is proposed in this paper. First, a global model is trained with federated learning framework on multiple local data clients. Then, we use the global model and private data to train the local model. To reduce the feature inconsistency between global and private local data and for better fitting the private local data, a novel "feature alignment" module is devised to guarantee the uniformity, which contains two parts, global alignment and local alignment, respectively. For global alignment, the graph metric of batch data is used to constrain the dissimilarity between features generated by the global model and local model. For local alignment, triplet loss is adopted to increase discriminative ability for local private data. Comprehensive experiments on our collected dataset are evaluated. The results show that the proposed method can be better adapted to local data and exhibit superior ability of generalization.

1. Introduction

Statistics of WHO report that heart disease is the most lethal chronic disease. Nearly 17.7 million people die of cardiovascular disease every year, which accounts for 31% of the total deaths in the world [1]. Electrocardiogram (ECG) is a physiological signal which is widely used in heart health monitoring. It contains many pathological information related to heart activity, and it is an effective way for monitoring and diagnosis of cardiovascular disease [2]. ECG is a long series of data and can be lasted for days, so it is very time and label consuming for monitoring and diagnosis with human experts. Therefore, it is necessary to use artificial intelligence technology for automatic cardiovascular disease diagnosis.

Using the advanced machine learning technology especially deep learning, the cardiovascular recognition model can be trained with labeled ECG data, i.e., the method proposed in [3] achieves 0.837 F1 value for 12 kinds of cardiac irregularities. However, most state of the art methods are based on public available training

datasets, which are relatively small and with limited varieties. Moreover, they are very difficult to deploy in practical applications.

In order to expand available cardiovascular information and guarantee the privacy, data from multiple medical institutions can be combined as a unified dataset, which can be used to train a superior global model with federated learning framework [4]. Federated learning is a special machine learning model using datasets that are distributed across multiple devices while preventing data leakage. It is also a privacy-preserving decentralized collaborative learning technique [5]. There have been works that adopted federated learning for medical data processing and model training [6, 7].

In this way, a centralized global model can be trained based on data from a large number of local nodes. Then, the global model is deployed to the local model for data prediction. However, there is a big challenge for which local models will get different performance by using the same global model. This is because data distributions of local clients vary with the global trained data distribution.

Therefore, personalization of the global model for each client becomes necessary to overcome the problem posed by heterogeneity of data distribution [8].

Device in terms of storage, computation, and communication capabilities will generate heterogeneous data. For ECG, sampling frequency and duration are different based on parameters setting and environment, which leads to nonuniform of pacing signal, left ventricular high voltage, and other diagnostic patterns. When there is large difference between data distributions of global server and local clients, it is hard to directly measure the feature inconsistency between them. Hence, this makes it difficult to deploy the global ECG classification model to local client with acceptable performance.

In order to solve these problems, a novel personalized federated learning framework for ECG data classification is proposed. First, a global ECG classification model is trained with a typical federated learning method across multiple local clients. Then, the global model is inherited to local model and served as the backbone part. During local model personalized training, the inherited part model is fixed. The local model is personalized based on the proposed feature alignment module. To utilize the generalization of the global model, we form the features of batch data into graph representation, so the internal structure between nodes can be preserved. The graph distance between global feature and local feature is used for global alignment constrain. From another point of view, local alignment is used to make the model better adapt to local private data. The metric learning method is adopted, and a triplet loss is designed to make data point of same class close to each other and negative data point far away. Finally, these loss functions are combined together for model training. By the proposed method, the consistency between global and local data can be learned, and the local personalized model is built with better adaptability and generalization. As far as we know, there are no related research studies for problems of personalized ECG classification. The main contributions of this paper are two folds:

- (i) In order to reduce the difference between global and local data, a novel feature alignment module is designed. With graph constraint and triplet metric, it can make the local model with better adaptability and generalization.
- (ii) Extensive experimental evaluations are carried out, and performance analyses are reported from multiple aspects.

The rest of this paper is organized as follows. Section 2 gives the related works. Section 3 describes the methodology. Experimental evaluation and analysis are given in Section 4. Section 5 concludes this paper.

2. Related Works

Related works are introduced in this section, including ECG data classification method, federated learning framework, and personalization in federated learning.

2.1. ECG Classification Method. With well-designed feature representation, ECG classification can be realized by models based on Bayes, K-means, Decision tree, and Linear Discriminate classifiers [9–11], along with commonly used optimization techniques [12, 13]. Features like cycle and higher order of QRS wave were extracted in reference [14]. Then, a fuzzy neural network was trained as classifier. Wavelet transform was used for feature extraction in reference [15]. Reference [16] adopted support vector machine for ECG classification.

A convolutional neural network method suitable for multilead ECG data was proposed in reference [17]. The stacked denoising autoencoder recognition model was used in reference [18] for ECG data. In reference [19], deep belief networks were used to construct the model for arrhythmia diagnosis. The convolutional neural networks model was used in reference [20] for heart beat classification. In reference [21], the deep factor decomposition method was used to decrease the influence of complex noise on ECG signal. Deep autoencoder was first used for signal denoising and reconstruction, and then fully convolutional network was trained as classifier. In reference [22], nonnegative matrix factorization was used for data dimension reduction and feature was extracted with sparse representation. Feature representation with multiple scales was proposed in reference [23], and progressive decisions were fused for final classification.

2.2. Federated Learning. In order to utilize massive distributed data storage and keep the privacy, federated learning is a useful framework to provide efficient training from data island and make model collaboration [4]. In reference [24], local models are trained at each local node, and only the updated parameters set is shared for global training. A multitask-based federated learning method was proposed in reference [25], which can solve the problem of high communication cost and fault tolerance. In reference [26], a safe client-server was first constructed. Data were allocated according to different local users. A homomorphic encryption method was designed for model parameter aggregation so as to improve server security [27]. A differential privacy method for federated learning was introduced in reference [28]. It provides protection for client data by hiding the customer's contribution during training. A ternary quantization method was proposed in reference [29], which optimizes the quantized networks and reduces lots of redundant parameters and excessive communication costs. To address problems of unlabeled and unannotated on-device data, reference [30] used a deep temporal neural network to train an auxiliary task by optimizing a contrastive objective with multiview strategy on diverse data sets.

2.3. Personalization in Federated Learning. Federated learning can be used to train a global model by utilizing distributed local data. However, for different local clients, the benefits they get from global model may vary greatly for various data distributions. To cope with non-IID data distributions of clients, personalized federated learning has

been proposed to improve performance of each local client by training a personalized local model. Reference [31] reported that local models' performances were hard to improve during federated learning, which may even worse than training only using local data. Therefore, it is important for model personalization according to specific local node. In reference [32], local nodes were first clustered, and each group models were training separately. In reference [33], part parameters of the global model were copied to the local model, and then local model was finetuned using local data. The metalearning method was adopted in reference [34], which treated model personalization as a meta testing procedure. In reference [35], there was a balance between global and local models. Global and local models were combined for final classification. In reference [36], the local model and global model were considered as two experts, and the personalized model was trained by mixed output of the personalized model and global model. Similar work was studied in reference [37]. Reference [38] introduced an attentive message passing mechanism to facilitate the collaboration effectiveness between clients.

3. Methodology

In this section, the proposed personalized federated learning for ECG classification based on feature alignment is described. Figure 1 gives the main framework of the proposed method. First, a global model M_G' is constructed with a typical federated learning framework from multiple local clients. Then, for each client c_i , we train a personalized local model with private dataset. The local model contains three parts, including M_G , M_C , and M_f . M_G is inherited from the global model M_G' . M_C is a convolution neural networks module for local feature extraction and representation. M_f is a fully connected layer or softmax layer to form the final classifier. During local model training, M_G is fixed to maintain the generalization ability of the global model. Specially, two alignment modules, global alignment and local alignment, are designed to constrain the feature distribution between local model and global model, which are realized by constraints of graph representation of feature generated by M_G and M_C , along with metric learning for intraclass and interclass losses. Finally, global alignment loss L_a , local alignment loss L_t , and cross-entropy loss L_c are all incorporated to optimize the objective function. Details are described in the following subsections.

3.1. Global Model Training. In order to make the best use of distributed data and privacy guarantee, federated learning is a popular way for model training. In the first step, the global model M_G' is trained with a most widely used federated learning framework FedAvg [39], which is synchronous update for each communication round.

Figure 2 demonstrates the framework for global model training in our work. There are n clients, and each contains a local dataset P_i and a local model M_i . Initially, the global

server sends the model parameters to all clients. Then, each client c_i performs local model training based on local data and then sends parameter update w_i back to the server. The server collects all local updates, and the global model parameters are optimized, and this process is repeated for multiple times. For efficiency, the global model can be updated when part of local updates is collected.

$$lc_i(w_i) = \frac{1}{n_i} \sum_{k=1}^{n_i} l(x_k, y_k, w_i), \quad (1)$$

$$\min_{w_i} lc_i(w_i). \quad (2)$$

For client c_i , its model parameter w_i is trained with local data $(x_k, y_k) \in P_i$. Equation (1) gives the loss function lc_i , which is the mean value of loss $l()$ for all training data. n_i is the total number of data in P_i . Then, equation (2) gives the minimize objective by adjusting w_i with SGD and BP method.

When clients finish their training, the server updates the global model parameter w_G by averaging all client model's parameters w_i . As shown in equations (3) and (4), N is the number of clients and t_i is the weight threshold for each model.

$$e_G = \frac{1}{N} \sum_{i=1}^N t_i * w_i, \quad \sum_{i=1}^N t_i = 1, \quad (3)$$

$$\min_{w_G} \frac{1}{N} \sum_{i=1}^N lc_i(w_G). \quad (4)$$

w_G is distributed to all clients after one iteration, and each client uses s_G as the base model to make further training using equations (1) and (2). After multiple iterations, the global model M_G' can be trained with optimal performance. Moreover, other federated learning frameworks can also be adopted for global model training.

3.2. Local Model Training. The global model M_G' is trained with federated learning over distributed data in the last subsection. However, M_G' cannot be directly deployed to local clients for local data inference when there is large difference between data distributions of global server and local clients. In this subsection, a personalized model adaptation method is designed based on the global trained model and private data of a specific client. The local model M_L contains 3 main components, $M_L = \{M_G, M_C, M_f\}$. M_G is inherited by the global trained model M_G' , which serves as the backbone and is fixed during local model training. M_C and M_f are used for local feature representation and final classifier of the local model.

For better fitting the local private data, a special constraint strategy "feature alignment" is devised to guarantee the uniformity between global and local models. The alignment module is further divided into two parts, global alignment and local alignment, which are described as follows.

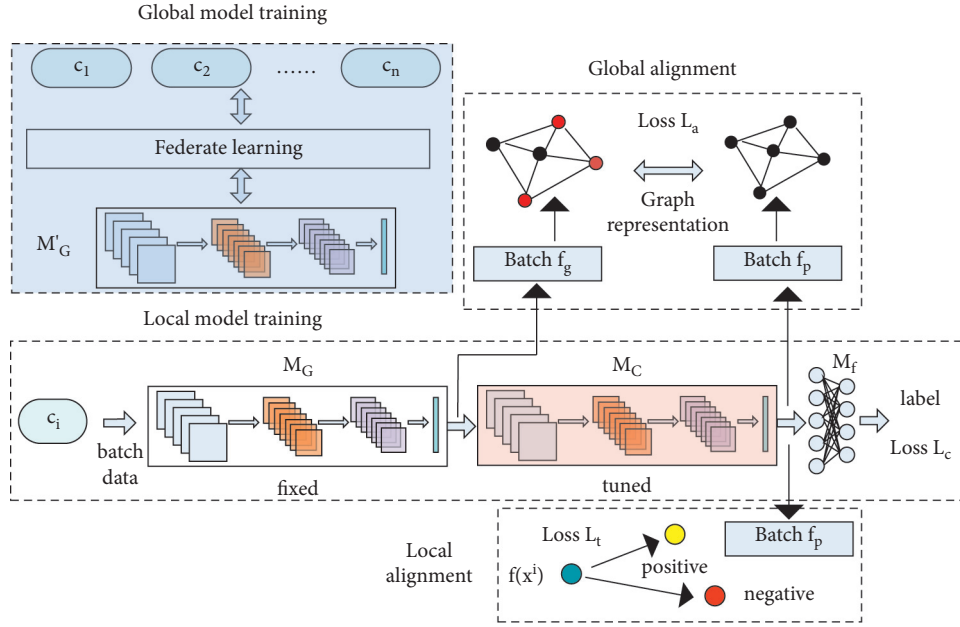


FIGURE 1: Main framework of the proposed method.

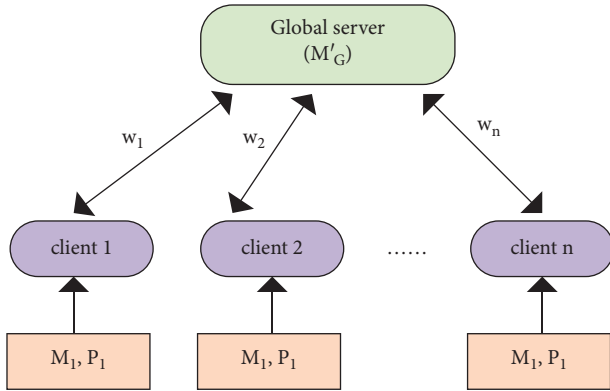


FIGURE 2: Global model training with federated learning.

3.2.1. Global Alignment. A global alignment module is first designed to constrain the local model training by constructing the structure between feature nodes. Different from other methods, the batch training data are used to form a graph structure, which can represent the relationship between data nodes. In this way, effect of single data feature shift can be reduced and relation between data nodes is retained. As shown in Figure 3, for i th training batch data samples (X_i, Y_i) in a private dataset, its global feature $f_g(X_i)$ and local feature $f_p(X_i)$ are extracted through models M_G and M_C . Then, the batch training data are treated as the basic group for global alignment operation.

The features of samples are then used to construct graph representation using $f_g(X_i)$ and $f_p(X_i)$. Suppose there are n samples in a training batch, then the graph representation contains n nodes and n^2 edges. The nodes of graph are represented with the features of samples in a training batch, and the edges of graph are represented by the distances between nodes. $gm(X_i)$ and $gp(X_i)$ are used to denote the graph representation of i th batch data by global and local models, respectively.

In order to measure the similarity between two batch data, matrix format is used to represent the graph structure representation. In this paper, only edges between nodes are incorporated for representation. It is hypothesized that the internal skeleton between nodes of a graph is more important and should be learned from the global model. If the node feature is used, then there will be strong probability that the local model is more like the global model. The white and yellow matrices in Figure 3 are used to denote $gm(X_i)$ and $gp(X_i)$, which are formulated as equations (5) and (6). e represents the edge between two nodes. x_j^i and x_k^i are two training samples in X_i .

$$gm(X_i) = \bigcup_{x_j^i, x_k^i \in X_i} e(fg(x_j^i), fg(x_k^i)), \quad (5)$$

$$gp(X_i) = \bigcup_{x_j^i, x_k^i \in X_i} e(fp(x_j^i), fp(x_k^i)). \quad (6)$$

Equation (7) gives the distance metric for two graph representations of a given batch data X_i . $d()$ means a distance computation method, and $|X_i|$ denotes the size of batch data. Basically, all edges of two feature graph representations are compared.

$$\begin{aligned} d(X_i) &= d(gm(X_i), gp(X_i)) \\ &= \frac{1}{|X_i|} \sum_{x_j^i, x_k^i \in X_i} d(e(fg(x_j^i), fg(x_k^i)), \\ &\quad \cdot e(fp(x_j^i), fp(x_k^i))). \end{aligned} \quad (7)$$

3.2.2. Local Alignment. For model personalization with local private data, a local alignment module is also designed, which aims to increase the classification performance for local private data.

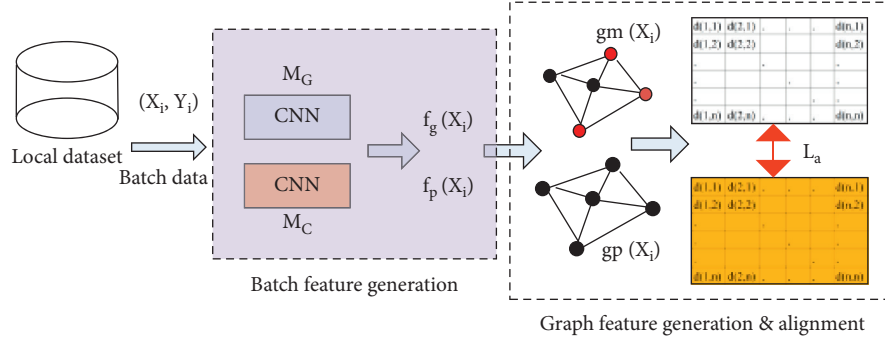


FIGURE 3: Demonstration of the global alignment module.

As shown in Figure 4, for a sample training batch data (X_i, Y_i) in a private dataset, its local feature $f_p(X_i)$ is extracted through models M_C . Then, the batch training data are used as the basic group for local alignment operation. Using metric learning method, we try to decrease distance between features from the same class; otherwise, the distance is increased. A training sample is randomly selected from $f_p(X_i)$, which is denoted as $f_p(x_i^a)$. It is called anchor sample. Then, another two samples are selected. One sample has the same label with anchor, and the other sample has different label. These three samples constitute a triplet $(f_p(x_i^a), f_p(x_i^p), f_p(x_i^n))$, where $f_p(x_i^p)$ denotes the positive sample and $f_p(x_i^n)$ denotes the negative sample. Through training, it is expected to decrease the distance between $f_p(x_i^a)$ and $f_p(x_i^p)$ and increase the distance between $f_p(x_i^a)$ and $f_p(x_i^n)$. The constraint is shown in the following equation:

$$\begin{aligned} & |f_p(x_i^a) - f_p(x_i^p)|_2^2 + \alpha < |f_p(x_i^a) - f_p(x_i^n)|_2^2, \\ & \forall (f_p(x_i^a), f_p(x_i^p), f_p(x_i^n)) \in \Gamma, \end{aligned} \quad (8)$$

where α is the threshold for minimal distance and Γ is the triplet set.

3.3. Training Objective Function. The final loss function of our proposed model contains three parts, global alignment loss L_a , local alignment loss L_t , and cross-entropy loss L_c , respectively.

Global alignment loss L_a is given in equation (9), which is inherited from equation (7). M means the number of batch data in all training dataset and i indicates the index of batch data.

$$L_a = \sum_{i=1}^M d(X^i). \quad (9)$$

Local alignment loss L_t is given in equation (10), which is on the basis of equation (8). “+” means that function value is 0 when content in [] is smaller than 0; otherwise, it is the normal loss function value. Γ_i is the triplet set for i th batch data.

$$L_t = \sum_{i=1}^M \sum_{\forall (f_p(x_i^a), f_p(x_i^p), f_p(x_i^n)) \in \Gamma_i} \left[|f_p(x_i^a) - f_p(x_i^p)|_2^2 - |f_p(x_i^a) - f_p(x_i^n)|_2^2 + a \right]_+. \quad (10)$$

Cross-entropy loss L_c is given in equation (11). $\text{CE}()$ is the standard cross-entropy function. M and i are the same as equation (8). $Y(X^i)$ is the ground truth label of batch data X^i , and $Y'(X^i)$ is the output value of the local model.

$$L_c = \sum_{i=1}^M \text{CE}(Y(X^i), Y'(X^i)). \quad (11)$$

The final loss function L is a combination of L_a , L_t , and L_c , as shown in equation (12). w_a , w_t , and w_c are weighted hyperparameters. L is used to update the parameter of M_C and M_f modules in the local personalized model:

$$L = w_a * L_a + w_t * L_t + w_c * L_c. \quad (12)$$

4. Experimental Evaluation

In this section, dataset description and experiment setting are first given. Then, the performance of the proposed method for personalized ECG classification is evaluated in terms of various environments.

4.1. Dataset and Experiment Setting. As there are no research studies on personalized federated learning for ECG classification, we construct a specific circumstance to evaluate the proposed algorithm. We collect about 120,000 ECG data from 8 hospitals. Table 1 gives the detail description of dataset. Six symptoms such as sinus rhythm, sinus arrhythmia, sinus tachycardia, sinus bradycardia, T-wave

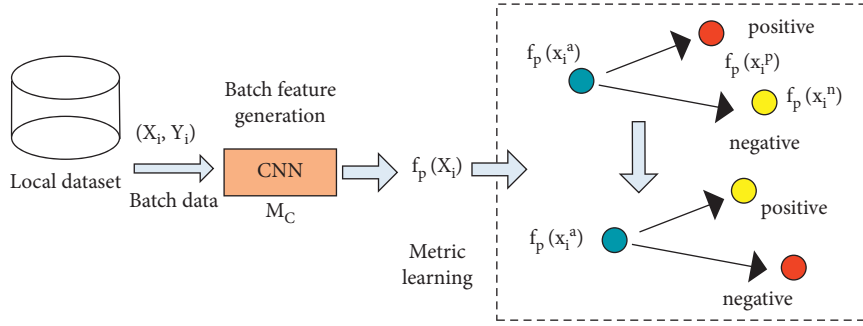


FIGURE 4: Demonstration of the local alignment module.

TABLE 1: Dataset used in our experiment.

| No. | Type | Quantity |
|-----|-------------------|----------|
| 1 | Sinus rhythm | 20201 |
| 2 | Sinus arrhythmia | 18581 |
| 3 | Sinus tachycardia | 13682 |
| 4 | Sinus bradycardia | 15854 |
| 5 | T-wave alternans | 14211 |
| 6 | Sinus normal | 38524 |

alternans, and normal are selected, which are the commonest types and data rich. There are 20201, 18581, 13682, 15854, 14211, and 38524 for types of sinus rhythm, sinus arrhythmia, sinus tachycardia, sinus bradycardia, T-wave alternans, and normal, respectively. Data distributions of each medical institution are also listed in Table 2.

In our research, each medical institution is corresponded to a local node, and these local nodes provide private data for global training. Hence, the federated learning circumstance is set up.

4.2. Base Model Training. The server (global model) employs the FedAvg to train the model globally whereas each local client updates its model locally after successive global aggregations using the SGD style algorithm. The CNN model with ResNet-34 is used as the base network structure for both global model and local model.

Each experiment is run for 100 global aggregations, with $e = 4$ epochs for SGD between successive global aggregations. The constant learning rate of 0.01 is used across global aggregations and clients.

Table 3 gives the classification result of each local node after federated learning. Average classification rate is used as the metric in our work. There are two node types, local and global. The global model has a classification rate of 82.6%. For local nodes, there are two evaluations. Local nodes 1 to 8 obtain the classification rate of 89.48%, 88.76%, 90.25%, 91.54%, 88.31%, 89.57%, 90.18%, and 90.54%, respectively, on their corresponding local private data set. Meanwhile, local nodes 1 to 8 obtain the classification rate of 54.65%, 57.18%, 49.75%, 51.43%, 48.92%, 46.35%, 52.45%, and 50.20%, respectively, on global testing data set. It can be obviously seen that the performance of local models is better than that of global model on private data. Local models get low performance on global data data, with about 30% lower than the global model. These

indicate that local models are more preferable to local data, while the global model trained with traditional federated learning framework needs further improvement. Therefore, it is urgent requirement to make model personalization.

4.3. Model Personalization Evaluation. In this subsection, the proposed personalization model-based feature alignment is evaluated.

The global model M_G' obtained in the above subsection is first downloaded to each local node, and then the personalized model for each local node is trained on the basis of local data set D_p and M_G' . Complying with Section 2, w_a , w_t , and w_c are set with 0.3, 0.3, and 0.4, respectively. Batch size is set with 16, and learning rate is 0.001.

Table 4 gives the result of model personalization. Column 1 is the node index. Column 2 and 3 are average performance on local data with models of training with only local data and training with personalization. Column 4 and 5 are average performance on global data with models of training with only local data and training with personalization. It can be seen that the performance of model with personalization is decreased with about 3%. This demonstrates that the personalization model is less deviated to distribution of local data. For global testing data, the performance of the model with personalization is greatly increased with about 15–18%. It is a good validation that the personalization model is more generalized.

4.4. Comparisons with Other Methods. In this subsection, some related model personalization methods are compared with our proposed model. We implement algorithms of [31, 34, 35], and the average classification rate on local and global testing data is evaluated.

Table 5 demonstrates the comparison result. Methods in [33–35] get an average performance of 84.41%, 82.70%, and 83.55% on local node testing data and 79.80%, 78.68%, and 76.45% on global testing data. There are about 4% and 6% compared with our proposed method, and this validates the effectiveness of the proposed personalization framework and feature alignment module.

4.5. Effect of Global Alignment and Local Alignment. In this subsection, effect of global alignment and local alignment is evaluated. Global alignment and local alignment are two

TABLE 2: Data distributions of each medical institution.

| Type | Node1 | Node2 | Node3 | Node4 | Node5 | Node6 | Node7 | Node8 |
|-------------------|-------|-------|-------|-------|-------|-------|-------|-------|
| Sinus rhythm | 2924 | 1940 | 2308 | 1483 | 3610 | 3103 | 3645 | 1188 |
| Sinus arrhythmia | 2451 | 1720 | 1893 | 1264 | 3171 | 2728 | 3513 | 1841 |
| Sinus tachycardia | 1913 | 1283 | 1407 | 1308 | 2262 | 2178 | 2619 | 712 |
| Sinus bradycardia | 2046 | 1523 | 1660 | 1418 | 2898 | 2665 | 2857 | 787 |
| T-wave alternans | 1850 | 1295 | 1477 | 1227 | 2418 | 2260 | 2767 | 917 |
| Normal | 5533 | 3866 | 3920 | 3653 | 6652 | 6196 | 7473 | 1231 |

TABLE 3: Performance of each local node after federated learning.

| Node type | No. | Classification rate (local) | Classification rate (global) |
|-----------|-----|-----------------------------|------------------------------|
| Local | 1 | 89.48 | 54.65 |
| Local | 2 | 88.76 | 57.18 |
| Local | 3 | 90.25 | 49.75 |
| Local | 4 | 91.54 | 51.43 |
| Local | 5 | 88.31 | 48.92 |
| Local | 6 | 89.57 | 46.35 |
| Local | 7 | 90.18 | 52.45 |
| Local | 8 | 90.54 | 50.20 |
| Global | | | 82.6 |

TABLE 4: Performance of each local node after federated learning.

| Node | Avg. (local) | Avg. (local with personalization) | Avg. (global) | Avg. (global with personalization) |
|------|--------------|-----------------------------------|---------------|------------------------------------|
| 1 | 89.48 | 86.25 | 54.65 | 84.42 |
| 2 | 88.76 | 87.18 | 57.18 | 83.25 |
| 3 | 90.25 | 87.50 | 49.75 | 84.58 |
| 4 | 91.54 | 88.42 | 51.43 | 83.60 |
| 5 | 88.31 | 85.85 | 48.92 | 82.46 |
| 6 | 89.57 | 85.34 | 46.35 | 83.15 |
| 7 | 90.18 | 87.68 | 52.45 | 84.56 |
| 8 | 90.54 | 88.15 | 50.20 | 84.10 |

TABLE 5: Comparisons with other methods.

| Algorithm | Avg. (local) | Avg. (global) |
|--------------------------------|--------------|---------------|
| On-device personalization [33] | 84.41 | 79.80 |
| Agnostic meta learning [34] | 82.70 | 78.68 |
| Mixture models [35] | 83.55 | 76.45 |
| Proposed method | 87.85 | 83.92 |

novel operations proposed in our work, which aims to catch the generalization ability of the global model and extract the discrimination ability of local private data. Here, the effect of global alignment and local alignment by assigning them different weights is evaluated.

Table 6 demonstrates the comparison result. Five parameter settings for w_a , w_t , and w_c are adopted. It can be seen from the table that with the raise of w_a , value avg. (local) increases; meanwhile, the value of avg. (global) decreases. There are similar appearances for w_t . w_a and w_t are two parameters to trade off the performance balance between

global data and local data. Parameter setting with 0.3, 0.3, and 0.4 obtains the optimal performance.

4.6. Evaluation of Execution Time. In this subsection, personalized model execution time is evaluated. Three CNNs structures with various training batch sizes are tested. Table 7 shows that the VGG16 model costs the most execution time, with 1.26 s, 1.67 s, 2.31 s, and 3.74 s for the batch size of 8, 12, 16, and 24 for each training iteration. ResNet-34 gets the least execution time, with about 58% of VGG16's. For model inference, three models take 0.120 s,

TABLE 6: Effect of global alignment and local alignment.

| w_a | w_t | w_c | Avg. (local) | Avg. (global) |
|-------|-------|-------|--------------|---------------|
| 0.5 | 0.1 | 0.4 | 90.5 | 60.28 |
| 0.4 | 0.2 | 0.4 | 85.8 | 80.7 |
| 0.3 | 0.3 | 0.4 | 87.85 | 83.92 |
| 0.2 | 0.4 | 0.4 | 86.0 | 81.1 |
| 0.1 | 0.5 | 0.4 | 80.3 | 83.2 |

TABLE 7: Execution time (seconds).

| Execution time | Batch size | VGG16 | Inception-v4 | ResNet-34 |
|----------------|------------|-------|--------------|-----------|
| Training time | 8 | 1.26 | 1.03 | 0.97 |
| | 12 | 1.67 | 1.29 | 1.18 |
| | 16 | 2.31 | 1.85 | 1.72 |
| | 24 | 3.74 | 2.26 | 2.15 |
| Testing time | x | 0.12 | 0.076 | 0.063 |

0.076 s, and 0.063 s for testing data, respectively. ResNet-34 is the preferred model for its excellent performance and acceptable cost.

5. Conclusions

This work proposes a novel personalized federated learning method for ECG classification. We explore feature alignment for personalization strategies on both global and local sides. Through experiments on our collected dataset, it shows that personalization benefits the local model with high performance and more generalization. To our knowledge, this is the first evaluation of personalization federated learning for ECG data analysis.

Our future works will focus on two aspects: (1) we will make more in-depth research on the personalization method with specific structure and (2) external dataset should be used to improve model performance, such as webly grabbed data [40].

Data Availability

Dataset used in this research is private.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was supported by National Key R&D Program of China under Grant no. 2017YFB1003000, National Natural Science Foundation of China under Grant nos. 61632008, 62072099, 61972085, 61872079, and 61972083, Jiangsu Provincial Key Laboratory of Network and Information Security under Grant no. BM2003201, and Key Laboratory of Computer Network and Information Integration of Ministry of Education of China under Grant no. 93K-9 and partially supported by Collaborative Innovation Center of Novel Software Technology and Industrialization, Collaborative Innovation Center of Wireless Communications

Technology, and the Fundamental Research Funds for the Central Universities.

References

- [1] World Health Statistics 2020, *Monitoring Health for the SDGs, Sustainable Development Goals*, World health organization, Geneva Switzerland, 2020.
- [2] S. K. Berkaya, A. K. Uysal, E. S. Gunal, S. Ergin, S. Gunal, and M. B. Gulmezoglu, "A survey on ECG analysis," *Biomedical Signal Processing and Control*, vol. 43, pp. 216–235, 2018.
- [3] A. Y. Hannun, P. Rajpurkar, M. Haghpanahi et al., "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Medicine*, vol. 25, no. 1, pp. 65–69, 2019.
- [4] J. Konečný, H. B. McMahan, D. Ramage, and P. Richtárik, "Federated Optimization: Distributed Machine Learning for On-Device intelligence," 2016, <http://arxiv.org/abs/1610.02527>.
- [5] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning," *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 2, pp. 1–19, 2019.
- [6] T. S. Brisimi, R. Chen, T. Mela, A. Olshevsky, I. C. Paschalidis, and W. Shi, "Federated learning of predictive models from federated Electronic Health Records," *International Journal of Medical Informatics*, vol. 112, pp. 59–67, 2018.
- [7] O. Choudhury, Y. Park, T. Salonidis, A. G. Divanis, I Sylla, and A. K Das, "Predicting adverse drug reactions on distributed health data using federated learning," in *Proceedings of the AMIA . Annual Symposium proceedings*, pp. 313–322, Washington, D.C., USA, November 2019.
- [8] V. Kulkarni, M. Kulkarni, and A. Pant, "Survey of Personalization Techniques for Federated Learning," 2020, <https://arxiv.org/abs/2003.08673>.
- [9] R. J. Muirhead and R. D. Puff, "A Bayesian classification of heart rate variability data," *Physica A: Statistical and Theoretical Physics*, vol. 336, no. 3, pp. 503–513, 2004.
- [10] I. Christov, I. Jekova, and G. Bortolan, "Premature ventricular contraction classification by theKth nearest-neighbours rule," *Physiological Measurement*, vol. 26, no. 1, pp. 123–130, 2005.
- [11] O. Behadada and M. Chikh, "An interpretable classifier for detection of cardiac arrhythmias by using the fuzzy decision tree," *Artificial Intelligence Research*, vol. 2, no. 3, pp. 45–58, 2013.

- [12] L. Ma, S. Cheng, and Y. Shi, "Enhancing Learning Efficiency of Brain Storm Optimization via Orthogonal Learning Design," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 11, 2020.
- [13] L. Ma, M. Huang, S. Yang, R. Wang, and X. Wang, "An adaptive localized decision variable analysis approach to large-scale multi-objective and many-objective optimization," *IEEE Transactions on Cybernetics*, 2021.
- [14] S. Osowski and T. H. H. L. Tran, "ECG beat recognition using fuzzy hybrid neural network," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 11, pp. 1265–1271, 2001.
- [15] I. Guler and E. Übeyli, "ECG beat classifier designed by combined neural network model," *Pattern Recognition*, vol. 38, no. 2, pp. 199–208, 2005.
- [16] E. D. Übeyli, "ECG beats classification using multiclass support vector machines with error correcting output codes," *Digital Signal Processing*, vol. 17, no. 3, pp. 675–684, 2007.
- [17] P. Xiong, H. Wang, M. Liu, F. Lin, Z. Hou, and X. Liu, "A stacked contractive denoising auto-encoder for ECG signal denoising," *Physiological Measurement*, vol. 37, no. 12, pp. 2214–2230, 2016.
- [18] M. M. A. Rahhal, Y. Bazi, H. AlHichri, N. Alajlan, F. Melgani, and R. R. Yager, "Deep learning approach for active classification of electrocardiogram signals," *Information Sciences*, vol. 345, pp. 340–354, 2016.
- [19] Z. Y. Wu, X. Q. Ding, and G. R. Zhang, "A novel method for classification of ECG arrhythmias using deep belief networks," *International Journal of Computational Intelligence and Applications*, vol. 15, no. 4, Article ID 1650021, 2016.
- [20] U. R. Acharya, S. L. Oh, Y. Hagiwara et al., "A deep convolutional neural network model to classify heartbeats," *Computers in Biology and Medicine*, vol. 89, pp. 389–396, 2017.
- [21] G. Wang, L. Yang, M. Liu et al., "ECG signal denoising based on deep factor analysis," *Biomedical Signal Processing and Control*, vol. 57, 2020.
- [22] R. Li, G. Yang, K. Wang, Y. Huang, F. Yuan, and Y. Yin, "Robust ECG biometrics using GNMF and sparse representation," *Pattern Recognition Letters*, vol. 129, pp. 70–76, 2020.
- [23] J. Zhang, J. Tian, Y. Cao, Y. Yang, and X. Xu, "Deep time-frequency representation and progressive decision fusion for ECG classification," *Knowledge-Based Systems*, vol. 190, Article ID 105402, 2020.
- [24] R. Shokri and V. Shmatikov, "Privacy-preserving deep learning," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, pp. 1310–1321, ACM, New York, NY, October 2015.
- [25] V. Smith, C. K. Chiang, M. Sanjabi, and S. T. Ameeet, "Federated multi-task learning," in *Proceedings of the Annual Conference on Neural Information Processing Systems*, pp. 4424–4434, Long Beach, CA, USA, December 2017.
- [26] H. B. McMahan, E. Moore, D. Ramage, and Y. A. B. Agüera, "Federated Learning of Deep Networks Using Model averaging," 2016, <http://arxiv.org/abs/1602.05629>.
- [27] Y. Aono, T. Hayashi, L. Wang, and S. Moriai, "Privacy-preserving deep learning via additively homomorphic encryption," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 5, pp. 1333–1345, 2018.
- [28] R. C. Geyer, T. Klein, and M. Nabi, "Differentially Private Federated Learning: A Client Level perspective," 2017, <http://arxiv.org/abs/1712.07557>.
- [29] J. Xu, W. Du, Y. Jin, W. He, and R. Cheng, "Ternary compression for communication-efficient federated learning," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2021.
- [30] A. Saeed, F. D. Salim, T. Ozcelebi, and J. Lukkien, "Federated self-supervised learning of multisensor representations for embedded intelligence," *IEEE Internet of Things Journal*, vol. 8, no. 2, pp. 1030–1040, 2021.
- [31] T. Yu, E. Bagdasaryan, and V. Shmatikov, "Salvaging Federated Learning by Local Adaptation," 2020, <https://arxiv.org/abs/2002.04758>.
- [32] Y. Mansour, M. Mohri, J. Ro, and A. T. Suresh, "Three Approaches for Personalization with Applications to Federated Learning," 2020, <https://arxiv.org/abs/2002.10619>.
- [33] K. Wang, R. Mathews, C. Kiddon, H. Eichner, F. Beaufays, and D. Ramage, "Federated Evaluation of On-Device Personalization," 2019.
- [34] Y. Jiang, J. Konecny, K. Rush, and S. Kannan, "Improving Federated Learning Personalization via Model Agnostic Meta Learning," *arXiv preprint arXiv:1909.12488*, 2019, <https://arxiv.org/abs/1909.12488>.
- [35] F. Hanzely and P. Richtarik, "Federated Learning of a Mixture of Global and Local Models," 2020, <https://arxiv.org/abs/2002.05516>.
- [36] B. Guo, Y. Mei, D. Xiao, and W. Wu, "PFL-MoE: personalized federated learning based on mixture of experts," *Web and Big Data*, vol. 1, pp. 480–486, 2021.
- [37] Y. Deng, M. M. Kamani, and M. Mahdavi, "Adaptive Personalized Federated Learning," 2020, <https://arxiv.org/abs/2003.13461>.
- [38] Y. Huang, L. Chu, Z. Zhou et al., *Personalized Cross-Silo Federated Learning On Non-Iid Data*, pp. 7865–7873, AAAI, CA, USA, 2021.
- [39] B. McMahan, E. Moore, D. Ramage, S. Hampson, and Y. B. A. Blaise, "Communication-efficient learning of deep networks from decentralized data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, pp. 1273–1282, Ft. Lauderdale, FL, USA, April 2017.
- [40] Y. Zhang, M. Long, J. Wang, and I. J. Michael, "On Localized Discrepancy for Domain Adaptation," 2020, <https://arxiv.org/abs/2008.06242>.