Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 4410894, 10 pages
https://doi.org/10.1155/2021/4410894

Research Article

Mobile Edge Computing Enabled Efficient Communication Based on Federated Learning in Internet of Medical Things

Xiao Zheng,[1] Syed Bilal Hussain Shah,[2] Xiaojun Ren,[3] Fengqi Li,[2] Liqaa Nawaf,[4] Chinmay Chakraborty,[5] and Muhammad Fayaz[6]

[1]*School of Computer Science and Technology, Shandong University of Technology, Zibo, Shandong, China*
[2]*School of Mechanical and Electronic Engineering, Dalian Jiaotong University, Dalian, China*
[3]*Blockchain Laboratory of Agricultural Vegetables, Weifang University of Science and Technology, Weifang, Shandong, China*
[4]*Computer Science School of Technologies, Cardiff Metropolitan University, UK*
[5]*Birla Institute of Technology Ranchi Jharkhand, Jharkhand, India*
[6]*Department of Computer Science, University of Central Asia, Naryn, Kyrgyzstan*

Correspondence should be addressed to Muhammad Fayaz; muhammad.fayaz@ucentralasia.org

Received 23 September 2021; Accepted 14 October 2021; Published 27 October 2021

Academic Editor: Shalli Rani

The rapid growth of the Internet of Medical Things (IoMT) has led to the ubiquitous home health diagnostic network. Excessive demand from patients leads to high cost, low latency, and communication overload. However, in the process of parameter updating, the communication cost of the system or network becomes very large due to iteration and many participants. Although edge computing can reduce latency to some extent, there are significant challenges in further reducing system latency. Federated learning is an emerging paradigm that has recently attracted great interest in academia and industry. The basic idea is to train a globally optimal machine learning model among all participating collaborators. In this paper, a gradient reduction algorithm based on federated random variance is proposed to reduce the number of iterations between the participant and the server from the perspective of the system while ensuring the accuracy, and the corresponding convergence analysis is given. Finally, the method is verified by linear regression and logistic regression. Experimental results show that the proposed method can significantly reduce the communication cost compared with the general stochastic gradient descent federated learning.

1. Introduction

The Internet of Medical Things (IoMT) is using a variety of communication systems to connect many devices to form best-in-class systems that can detect, collect, exchange, analyze, and transmit valuable communications [1, 2], helping companies manage smarter and deliver faster business solutions. IoMT can build a large number of applications through various "smart" sensors such as artificial intelligence and machine learning (ML) technology, thereby revolutionizing the ubiquitous computing system [3, 4]. Secure communications and sensing technologies can leverage a participatory approach to implement integrated solutions while establishing new applications relevant to the industry, particularly healthcare. One of the key applications of 5G-based IoMT is healthcare, which is aimed at maintaining patients' medical information in electronic environments (such as cloud and edge cloud) systems through the latest telecom paradigm [5, 6]. For healthcare applications, ML models are typically trained on enough user data to track health status information. Traditional machine learning methods such as support vector machine (SVM), decision tree (DT), and hidden Markov model (HMM) can be used in a variety of healthcare applications [7]. Patterns are analyzed and classified based on the construction of explicit or implicit models, and its ML method has been used to

improve the detection rate of malicious data [8]. However, they still have many problems in detecting new or evolving malicious data, and the accuracy of unsupervised anomaly detection used to detect new security is low [9]. As the number of variants grew, this became a major bottleneck, mainly because of the amount of work required to gather enough datasets. In addition, when new features from different network layers need to be combined to deal with the evolving malicious data, the learned classifier cannot be directly used to test the data with different feature spaces [10]. This paper attempts to overcome these challenges, which involve data aggregation with security and privacy protection. First, in the real world, data often exists in separate, decentralized forms. Although there is a lot of data in different sensors, it is not shared due to privacy and security concerns [11]. If the same user uses data from two different sensors, the data stored in different clouds cannot be exchanged, making it difficult to train powerful models with valuable data. Another important issue is personalization based on feature data, most of which are based on a common server model for almost all users. After capturing enough user data, train a satisfactory machine learning model, which itself is distributed to all user devices that can track health information on a daily basis, but the program lacks personalization. It can be seen that different users have different characteristics and daily behavior models. As a result, general models cannot deliver personalized healthcare. Based on this idea, a federated transfer learning algorithm is proposed, which is an IoMT-enabled intelligent healthcare framework named FT-IoMT Health [12]. FT-IoMT Health can solve the problem of data decentralization and model personalization through federated learning and homomorphic multiparty encryption methods [13]. FT-IoMT Health aggregates data from different systems to build powerful machine learning models and appropriately protect user privacy. After building the cloud model, FT-IoMT Health utilizes migration learning to implement a personalized model for each network entity [14]. Transfer learning is a novel machine learning technology, which utilizes knowledge learned from related training (source) sets to improve the prediction accuracy of test (target) sets with almost no label data [15] and enables the framework to update gradually. FT-IoMT Health is scalable and used in many healthcare applications, enabling them to constantly update their learning capabilities every day.

In short, the main contributions of the paper are as follows:

(1) This paper proposes an algorithm, FT-IOMT Health, which is the first federated migration learning mechanism based on IoMT. This mechanism aggregates data from different entities without compromising privacy and security and obtains relatively personalized models by means of transfer learning

(2) On the basis of known data analysis, transfer learning technique is used to detect new unknown data analysis. The use of transfer learning itself is the main advantage of enhancing the adaptability of the detection model

(3) This paper validates FT-IoMT Health's superior performance in identifying human activity on UCI smartphones. The experimental results show that FT-IoMT Health greatly improves the recognition accuracy compared with traditional ML methods

2. Related Work

In traditional healthcare applications, it is important to note that models are typically built by aggregating all user data. In practice, however, data is often separated and difficult to share due to privacy issues, and the models built by applications lack the characteristics of model personalities. A well-known network data detection technique is signature-based detection, which is based on the deep information of the specific characteristics of each detection. Another technique used for network data detection is supervised learning [16, 17]. Both studies were less accurate in detecting new data because they typically relied on known cases of detection. Federated machine learning was first proposed by Google [18]; since the phone is distributed throughout its life cycle, Google trains the machine learning model on this machine, with the primary purpose of protecting user data in the program. Federated learning is a technical approach to solve the problem of data discreteness through the training of privacy models in networks. The goal of transfer learning is to transfer information from known related fields to new fields, so as to achieve the purpose of analogical reasoning, and the main goal is to reduce the distribution differences between different fields. Therefore, there are two main implementation methods: instance reweighting [19] and feature matching [20]. Recently, deep transfer learning technology has made great achievements in many applications. FT-IoMT Health mostly involves deep transfer learning. Many methods assume the feasibility of training data, which is obviously unrealistic. FT-IoMT Health builds deep migration learning into a federated learning framework, eliminating the need to access raw user data. Therefore, this achieves the goal of greater security.

The point of federated transfer learning here is that samples or features do not have more in common. In recent years, a number of researchers have begun to dabble in the field. In [12], Liu et al. put forward a secure federated transfer learning algorithm in a two-party privacy protection environment, which paid more attention to data security. Most studies also propose a federated domain adaptive approach, which extends the domain adaptive approach to federated setting constraints to achieve data privacy and domain transformation. Although a great deal of research work continues to develop rapidly, there are still many challenges in the practical application of federated transfer learning. The work in this paper is the first federated transfer learning mechanism designed specifically for IoMT applications and will therefore be extended by a variety of transfer learning technologies.

3. System Model

3.1. Problem Definition. Take data from N different users, the user is represented as $\{U_1, U_2, \cdots, U_N\}$, and the reading value of the sensor providing the data is defined as $\{\mathscr{D}_1, \mathscr{D}_2, \cdots \mathscr{D}_N\}$. The conventional method trains the general model $\mathscr{M}_{\mathrm{GEN}}$ by combining all the data $\mathscr{D} = \{\mathscr{D}_1 \cup \mathscr{D}_2 \cdots \cup \mathscr{D}_N\}$. All data should have different distributions. In response to our proposed problem, we aim to gather on all data to train the federated model $\mathscr{M}_{\mathrm{FED}}$, in which no user U_i will disclose these data $\mathscr{D}_i$ to each other. If we define the accuracy as $\mathscr{A}$, the goal of FT-IoMT Health is to guarantee that the accuracy of federated learning is approximately or better than the accuracy of the following conventional learning: $\mathscr{A}_{\mathrm{FED}} - \mathscr{A}_{\mathrm{GEN}} > \delta$, where δ is a very small positive real number.

FT-IoMT Health is aimed at leveraging joint transfer learning technology to obtain accurate personal healthcare information without compromising user privacy. Figure 1 shows a profile of the mechanism. Suppose there exist N users (network data) and one server, and then, expand them to a more general situation. The main components of the framework are described below. First, train the cloud model on the server based on a common dataset. Therefore, the cloud model is distributed to all users so that each user will train his model on his own dataset. The user model is then uploaded to the cloud for training the new cloud model using model aggregation. Finally, each user will implement personalized models to train users based on cloud models, network data, and predictive future data. In this process, due to the large distribution difference between server data and user data, the transfer learning method is adopted to make the model more suitable for users, as shown on the right end of Figure 1. It is important to note that none of the parameter sharing processes will include user data leaked through homomorphic encryption.

The federated learning model is an important computation model for the entire FT-IoMT Health mechanism. Its role in the whole process is to deal with model construction and parameter sharing. The server model will be directly applied to users after the learning and training process. This is exactly a way based on traditional healthcare applications applied to model learning. Obviously, the probability distribution of the samples in the server and the data generated by each user is very different. Therefore, it is difficult for the general model to achieve personalized settings of the data model. In addition, due to privacy security issues, the user model cannot easily achieve continuous model updates.

3.2. Federated Learning. FT-IoMT Health uses the federated learning paradigm to implement training and sharing of encryption models, and its steps mainly involve the following two key parts: namely, cloud and user model learning. For FT-IoMT Health, deep neural networks are used to learn cloud and user models. The deep neural network uses the original input of user data as the network input for end-to-end feature learning and classifier training, where f represents the server model to be learned, and the learning goal is

$$\underset{\theta}{\operatorname{argmin}} L = \sum_{i=1}^{N} L(y_i, f(x_i)), \tag{1}$$

where $L(*, *)$ indicates network loss function such as cross-entropy loss for classification tasks, $\{x_i, y_i\}$ is a sample of server data, and its size is N. θ represents all the parameters to be learned, namely, weights and bias.

After obtaining the cloud model, distribute it to all users. From the obstacle in Figure 1, direct sharing of user info will be prohibited. The process exploits homomorphic encryption to prevent info leakage. Due to the fact that encryption is not a subject to be considered, only the procedure of homomorphic encryption applying the addition of real numbers is explained. Therefore, this can complete parameter sharing without leaking any user information. We apply federated learning to aggregate user data without compromising privacy and security. Therefore, the learning goal for user u is defined as

$$\underset{\theta^u}{\operatorname{argmin}} L_u = \sum_{i=1}^{N} y_i^u, f_u(x_i^u). \tag{2}$$

After completing the training of all user models f_u according to the shared cloud model, upload them to the server for aggregation. It can be seen from the evaluation that in the case of shared initialization, the method of federated averaging [21] can be adopted to average the model to achieve good performance in reducing loss. Therefore, following [21], align the user model by the model average value, and then, perform the cloud model update average value on b user models in each training round. The updated cloud model is expressed as

$$\bar{f}(w) = \frac{1}{B} \sum_{b=1}^{B} f_{u_b}(w), \tag{3}$$

where w is the parameter of the network and B is the number of users. After enough iterations, the updated server model $\bar{f}$ has better generalization capabilities. Then, new users can join the next round of server model training. Therefore, FT-IoMT Health has incremental learning functions.

3.3. Transfer Learning. Apply transfer learning technology to improve the detection of new network data analysis by transferring the information learned from known network data analysis, so as to distinguish between the common coarse feature model for all users and the fine-grained feature model for personalized user. The expression source and target are used to define the training and test datasets in the machine learning task, respectively. Both source and target data are represented by normal flow records and abnormal flow records. The purpose of this transfer learning is to adapt source data to assist distinguish new detections from the target, thereby building a personalized model for each user.

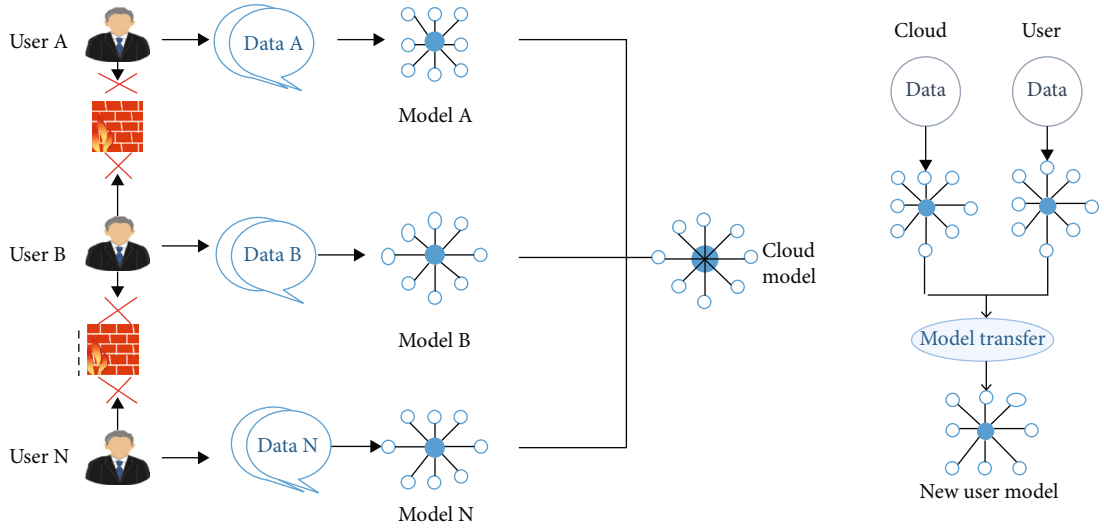


Figure 1: Overview of FT-IoMT Health framework.

The transfer learning mechanism is composed of the following three major processes: (1) feature extraction process (obtained from the original network), (2) feature-based learning process, and (3) supervised classification process. The first step is to perform data tracking on the original network to extract features based on the statistical calculation of network traffic. In the second step, a feature-based transfer learning algorithm is used to learn the new feature representation from both the source data and the target data, and the new representation will be fed to the general basic classifier.

The data detection is modeled as a binary classification problem, i.e., the data state is classified as malicious or normal. Assume a source training instance $S = \{X \mid x_i\}, X \in R^m$ with label $L_S = y_i$, and target data $T = \{Z \mid z_i\}, Z \in R^n$, where X and Z are both users' data extracted from the network. X and Z come from different distributions $P_S(X) \neq P_T(X)$, X and Z have different dimensions, $R^m \neq R^n$. Our goal is to accurately predict the label on T.

The method is to apply new public latent space through spectrum transformation, in which the distribution of malicious examples is similar, but the distance between discriminatory ones is still very different. The ultimate purpose is to learn a new representation of the original data and target data in the k-dimensional latent semantic space, namely, $V_S \in R^k, V_T \in R^k$, so that it can use V_S and V_T instead of the original S and T better against malicious data sort. Its key purpose is given in Figure 2, because in the new projected public latent space (Figure 2(c)), the distribution of malicious A and malicious B are indistinguishable, even though they are in their original 2D and 3D spaces.

The following discusses how to search the public latent subspace. The optimal subspace is described in the following.

3.3.1. Optimization. Based on the given source data S and target data T, find the best projection of S and T on the best subspace V_S and V_T on the basis of the optimization goals given below:

$$\min_{V_S, V_T} L(V_S, S) + L(V_T, T) + \gamma \bullet D(V_S, V_T), \tag{4}$$

where $L(*, *)$ is a distortion function used to evaluate the difference between the original data and the projection data and $D(V_S, V_T)$ indicates the projection difference between the source data and the target data. γ is a trade-off parameter used to control the resemblance between two datasets.

Therefore, the first two components of (4) can assure that the projection data is as consistent as possible with the original data structure. Define $L(*, *)$ as follows:

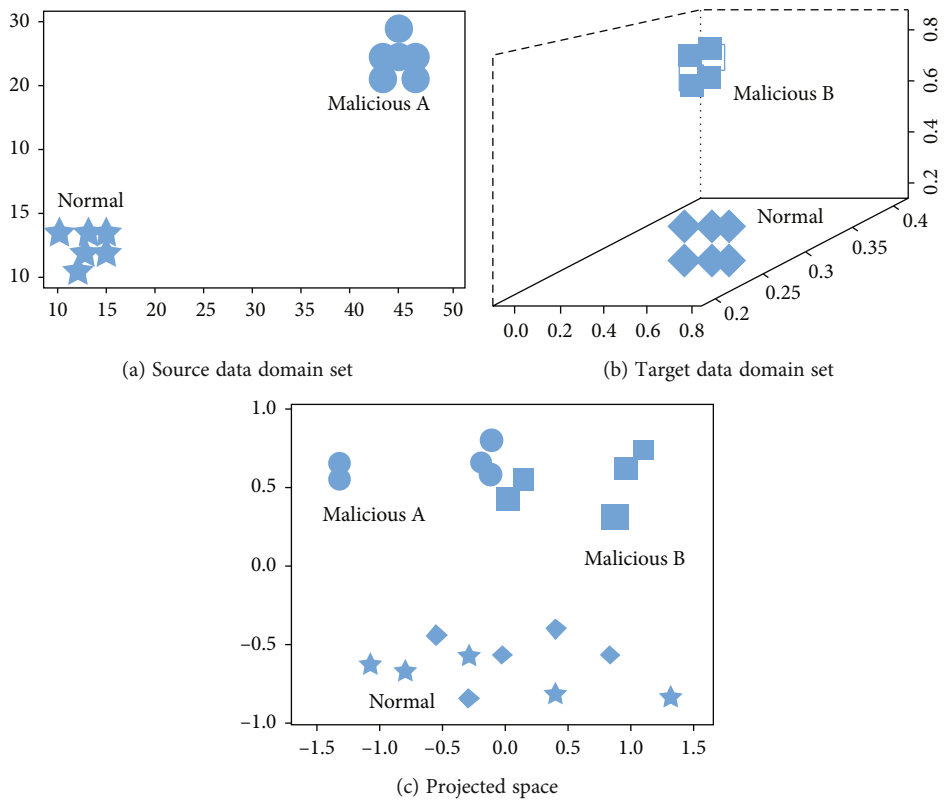
$$L(S, V_S) = \|S - P_S * V_S\|, L(T, V_T) = \|T - P_T * V_T\|, \tag{5}$$

where V_S and V_T are realized via a linear transformation with linear mapping matrices expressed as $R_S \in R^{k \times m}$ and $P_T \in R^{k \times n}$ to the source data and target data. $\|X\|^2$ indicates the Frobenius norm, which is also denoted as the matrix trace norm. In another point of view, $P_S' \in R^{k \times m}$ and $P_T' \in R^{k \times n}$ project the original data S and T into a k-dimensional space, in which the projected data are equivalent $(L(S, V_S) = \|SP_S' - V_S\|^2)$. But it can produce trivial solutions $P_S = 0, V_S = 0$. Therefore, Equation (5) will be applied. It is regarded as matrix factorization, which is a well-known advantageous tool for extracting latent subspaces while maintaining the original data structure.

According to $L(*, *)$ to define $\Delta(V_S, V_T)$ as

$$\Delta(V_S, V_T) = L(V_S, V_T), \tag{6}$$

which represents the difference between the projection target data and the source data. Therefore, based on the minimized



(a) Source data domain set

(b) Target data domain set

(c) Projected space

FIGURE 2: Overview of the proposed feature space transformation form.

difference function (6), the source data and target data constraints of the projection are similar.

Substituting (5) and (6) into (4), the following optimization goals to minimize with respect to V_S, V_T, P_S, and P_T are as follows:

$$\min_{V_S'V_S=I,\, V_T'V_T=I} \nabla(V_S, V_T, P_S, P_T) = \min_{V_S'V_S=I,\, V_T'V_T} = I\,\|S - V_S P_S\|^2 + \|T - V_T P_T\|^2 + \gamma\bullet\left(\|V_T - V_S\|^2\right). \tag{7}$$

Therefore, the loss function of the user model can be calculated by the following formula:

$$\underset{\theta^s}{\operatorname{argmin}} L_s = \sum_{i=1}^{N} L(y_i, f(x_i) + \sum_{i=1}^{N_s} L(y_i^s, f_s(x_i^s)) + \|S - V_S P_S\|^2 + \|T - V_T P_T\|^2 + \gamma\bullet\left(\|V_T - V_S\|^2\right). \tag{8}$$

The learning process of FT-IoMT Health is given in Algorithm 1. The framework will work continuously with newly emerging user data. When faced with new user data, FT-IoMT Health can simultaneously update the user model and the network-based cloud model. Thus, the longer the user spends data, the more personalize the model. In addition to transfer learning, other common methods (e.g., incremental learning) are also implanted in FT-IoMT Health for personalized settings.

4. Experiments

4.1. Datasets. We employ a public human action recognition dataset named UCI smartwatch. The dataset involves 6 actions gathered from 35 users who use smartwatch around their wristband. 10 accelerometer and gyroscope data channels are gathered at a constant rate of 50 Hz. There exist 10,300 cases. To construct the subject status in FT-IoMT Health, five relevant topic features (content IDS 31-35) are extracted from them, and they are regarded as independent users, who will not share data because of privacy security. The data of the remaining 30 users is used to train the cloud model. Then, the goal is to use the cloud model and all 5 independent objects to improve the accuracy of the activity recognition of these 5 objects without compromising privacy. Consider it is a simplification of the framework in Figure 2, where there are 5 users.

For the feature transfer learning used in the construction of the personalized model, we mainly analyze from the network data detection. The network functions that contains can be summarized into three groups: here, we focus on studying the traffic data features, which are generally extracted by flow analysis tools, and content features, which need to deal with grouping content.

4.2. Specific Implementation Steps. Both the server and the user side use CNN for training and testing. The cyber is

1: **Input:** $T, S, \gamma, k, \{\mathscr{D}_1, \mathscr{D}_2, \cdots, \mathscr{D}_N\}$, learning rate α, steps = 500
2: **Output:** f_u, V_S, V_T
3: Construct an initial cloud model f with common datasets applying Equation (1)
4: Distribute f to all users
5: Train user model by Equation (2)
6: All user models are updated to the server through homomorphic multiparty encryption. Perform aggregation on the model employing Equation (3). Then, the server treats the aggregation model as the updated cloud model $\bar{f}$.
7: Distribute $\bar{f}$ to all users and then execute transfer learning on each user to obtain their model f_u
Applying Equation (8)
8: **while** optimized function Equation (7) not converge **do**
9: Update V_T by gradient descent with $V_T = V_T - \alpha(\partial\nabla/\partial V_T)$
10: Update V_S by gradient descent with $V_S = V_S - \alpha(\partial\nabla/\partial V_S)$
11: Update P_T by gradient descent with $P_T = P_T - \alpha(\partial\nabla/\partial P_T)$
12: Update P_S by gradient descent with $P_S = P_S - \alpha(\partial\nabla/\partial P_S)$
13: step++
14: **end**
15: Repeat the above process for new user data constantly appearing

Algorithm 1: The learning process of FT-IoMT Health.

consisted of the following 2 convolutional layers, 2 pooling layers and 3 fully connected layers, which employ a 1×9 convolution size. It is optimized using small batch stochastic gradient descent (SGD). In the training process, 80% of the training data is used for model training, and the remaining 20% is used for assessment. Set user $B = 5$ and fixed. When the batch size is 64 and the training period is set to 80, the learning rate α is set to 0.01. Model network data detection as a binary classification issue to differentiate malicious traffic from normal one.

To effectively assess the transfer learning method, source and target datasets will be generated as follows. To assess the performance of the transfer learning method in detecting unknown model variants in the cloud, so as to construct personal model, the problem is regarded as a detection that only exists in the target network but is not visible in the source network. Suppose there is one data detection in the source and another detection in the target. Therefore, the distribution of detection feature value between the source and the target is different. Therefore, three datasets are reconstructed, each of which includes a series of randomly chosen normal cases and a set of detections from one category. Here, one of the datasets is set as the target, and the other dataset is set as the source. Therefore, there are mainly the following three detection tasks: Seen⟶Unseen (i.e., source Seen data for training, target Unseen data (new network) for testing), Seen⟶Detection, and Detection⟶Unseen. It is presumed that the feature space between the source and target is the identical. The accuracy of user u is computed by the following formula: $\mathscr{A}_u = |X : X \in \mathscr{D}_u \wedge \tilde{y}(X) = y(X)|/|X : X \in \mathscr{D}_u|$, where $y(X)$ and $\tilde{y}(X)$ define the true and predicted labels on X, respectively. Perform federated learning according to homomorphic encryption. During the transfer learning period, all convolution and pooling layers in the network are frozen, and only the parameters of the fully connected layer are updated using SGD. To verify the validity of FT-IoMT Health, its performance was compared with conventional deep learning (DL). In traditional deep learning, we only use the primary server model and other conventional machine learning modes to record each the performance of each subject. The hyperparameters used in all comparison methods are adjusted by cross-validation. To achieve a fair study, all experiments were performed 5 times to record the average accuracy. Table 1 shows the performance comparison between the detection technologies proposed based on FT-IoMT and the benchmark method. Table 2 shows the accuracy of activity classification for each topic. Figure 3 indicates the ROC curve. Figure 4 compares FT-IoMT with other transfer learning methods. Figure 5 shows the results of extending FT-IoMT through other transfer learning methods.

4.3. Evaluation. FT-IoMT achieves the best classification accuracy for all users. From the outcomes in Tables 1 and 2, it can be concluded that FT-IoMT Health has importantly enhanced performance in all examples. Compared with DL, it slightly increases the average result by 5.6%. Mainly due to the fact that federated learning can be used indirectly for more info from distributed data model to train better and applying transfer learning, the model can be more personalized for each user's features. Compared with traditional methods such as KNN, SVM, and RF, FT-IoMT Health also significantly enhances the recognition outcomes. Overall, it proves the validity of the FT-IoMT Health mechanism. For activity recognition, the results also show that deep learning methods (DL and TL-IoMT) attain better outcomes than conventional modus.

It is controlled by the representation capabilities of deep neural networks, while conventional modus depend on manual feature learning. Deep learning also has another advantage of enabling the online update model to be incrementally updated without retraining, while conventional modus need further incremental algorithms. The performance is very valuable in model reuse and federated transfer learning. In view of the unseen new network data detection

Table 1: Classification accuracy of the test objective.

Subject	KNN	SVM	RF	DL	FT-IoMT Health
P_1	82.6	80.8	86.7	93.4	97.6
P_2	87.4	95.7	94.6	94.1	97.8
P_3	91.8	96.8	87.5	92.6	99.7
P_4	84.5	94.8	90.3	94.8	98.9
P_5	91.3	97.9	92.1	91.9	99.8
AVG	87.5	93.2	90.2	93.2	99.1

Table 2: Accuracy of unprediction network detection.

Datasets	Method	SVM	KNN	RF
Seen→Unseen	No-TL	0.51	0.52	0.54
	TL-IoMT	0.82	0.81	0.80
Seen→Detection	No-TL	0.76	0.75	0.65
	TL-IoMT	0.87	0.82	0.80
Detection→Unseen	No-TL	0.50	0.52	0.53
	TL-IoMT	0.84	0.82	0.81

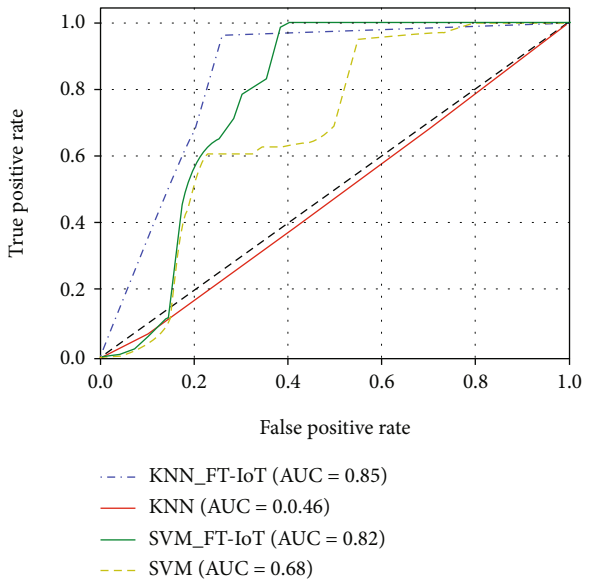


Figure 3: ROC curve on Seen⟶Unseen.

environment, we will compare the performance of FT-IoMT Health with common basic classifiers, instead of using the transfer learning method for the three detection tasks. We chose random forest (RF), SVM, and KNN as common basic classifiers. From the ROC curve illustrated in Figure 3, it will be seen that FT-IoMT Health has improved the detection rate compared to the baseline. Comparison of IoT-based transfer learning methods: we have used other feature-based transfer learning methods (such as HeMap [22] and CORrelation ALignment (CORAL) [23]) to evaluate FT-IoMT network data detection tasks. From the outcomes

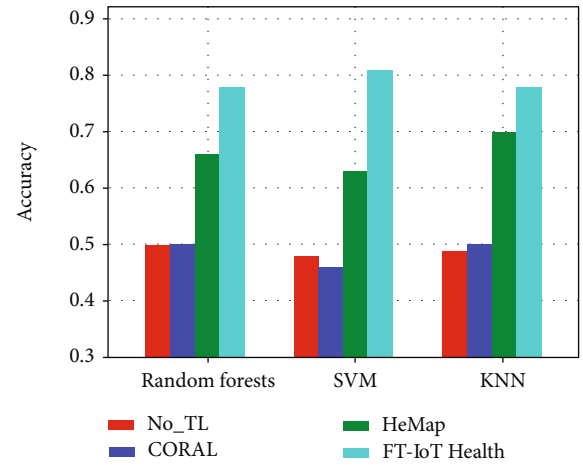


Figure 4: Performance comparison of feature-based transfer learning on Seen→Unseen.

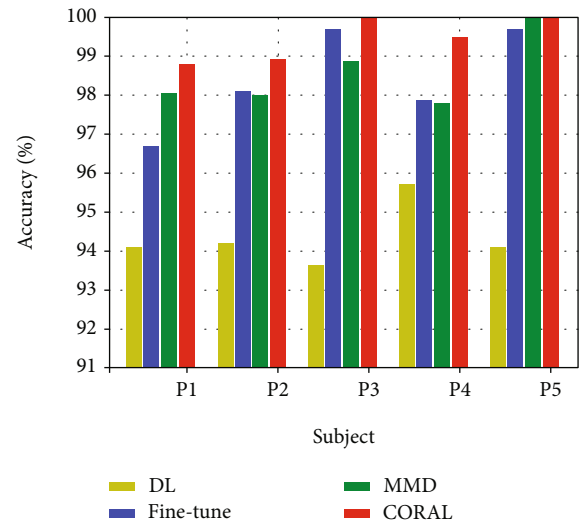


Figure 5: Extending FT-IoMT with other transfer learning approaches.

illustrated as Figure 4, it can be achieved that the performance of FT-IoMT is better than other feature-based methods in all classifiers for network data detection tasks. There exist two adjustable parameters, the similarity confidence parameter γ and the size of the new feature space k, which will be set manually or automatically by experiential research. There are methods for automatically determining the best parameters, for example, by calculating the similarity degree between the source and the target data to determine the similarity confidence parameter γ. In this work, a small labeled dataset (300 labeled) is used in the test set to search the best parameters.

For the use of other transfer learning methods to expand FT-IoMT Health, using different transfer learning methods to analyze the scalability of FT-IoMT Health, it uses two methods to compare its performance: (1) fine-tune, by only fine-tune the network on each subject, it will not significantly reduce the distribution difference between sets; (2)

Table 3: Classification accuracy of every subject in arm swing and postural normal tremor.

Subject	KNN	SVM	RF	DL	FT-IoMT Health	Upper bound
Arm swing						
P_1	37.2	41.5	45.1	50.2	74.7	87.9
P_2	45.3	47.6	49.2	58.5	91.2	99.4
P_3	56.2	55.7	53.4	63.6	86.4	87.5
P_4	63.8	62.0	56.7	69.4	94.6	100
P_5	84.9	73.6	66.6	71.3	85.7	88.4
AVG	57.5	56.0	54.2	62.6	86.5	92.6
Postural tremor						
P_1	51.3	46.5	58.3	46.2	84.3	86.2
P_2	52.5	58.7	56.9	60.4	75.8	85.9
P_3	64.8	54.1	56.1	58.7	68.6	75.6
P_4	58.6	59.2	52.7	62.8	71.4	86.8
P_5	65.2	53.6	52.0	59.2	69.1	76.4
AVG	58.4	54.5	55.2	57.4	73.8	82.4

MMD (Maximum Mean Difference) is used for transfer, and MMD loss is used instead of alignment loss. The comparison outcome is shown in Figure 5. It will be seen from the figure that in addition to alignment loss, FT-IoMT Health can also attain desirable outcomes through fine-tuning or MMD.

The outcomes of transfer learning are greatly better than no transfer in average accuracy. It shows that the transfer learning process of FT-IoMT Health is very valid and scalable. Thus, FT-IoMT Health is universal and will be expanded in many fields by merging other transfer learning algorithms. In addition, other encryption algorithms can also be used to extend the federated learning, which may be a future research direction.

5. Application of Assistance in Diagnosis and Treatment of Neurological Diseases

Parkinson's disease is generally a neurological disease characterized by some motor symptoms, so biosensors can be used in IoMT to help diagnose [24]. In addition, patient data is also a privacy-sensitive problem and must be resolved through federated learning. Therefore, FT-IoMT Health is applied to assist in diagnosis and treatment of Parkinson's disease and is arranged in hospitals. After training the user model on the user side, the patient downloads it to the biosensor and connects to the network to update it during the next access. This allows users to detect and obtain real-time feedback on their own, so as to more easily obtain disease status.

Based on this, a biosensing application was developed to collect the patient's acceleration and gyroscope signals at a frequency of 80 Hz for symptom testing. The symptom condition test is designed in the following states: arm swing, balance, walking, postural normal tremor, and resting tremor. For each test set, each symptom is divided into five levels from normal to severe. The treating doctor evaluated the collected symptoms. We collected sensor data from 150 patients aged 18 to 85 years. In the following evaluation process, the test data of arm swing and postural normal tremor are evaluated, and two categories with quite sufficient data are chosen as references.

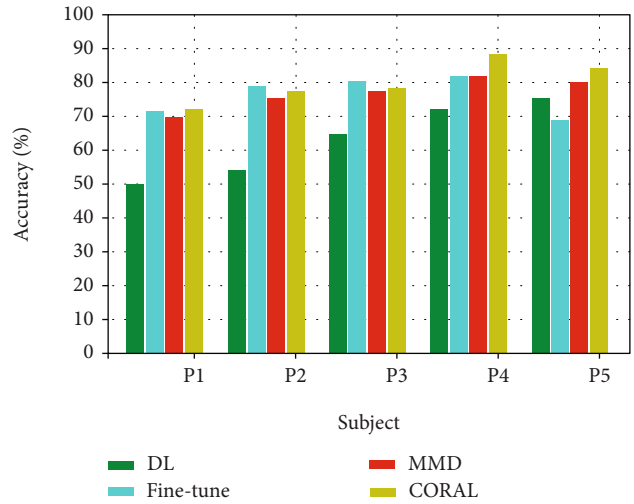


Figure 6: Extending on arm swing test.

Evaluate the classification accuracy of the collected dataset. The data is gathered from three hospitals, 80% of each hospital is randomly chosen as the public dataset, the remaining 20% are randomly selected as 5 users, and $K = 5$.

Table 3 shows the comparison results. In addition, the proposed method gives the result of the ideal scheme. Due to all the data is preserved in one location, it is easier to view the upper bound of the model performance. From the outcomes, it will be seen that FT-IoMT Health has achieved the best classification accuracy, which obviously exceeds the best comparison means, and has narrowed the gap with the perfect case. It is fully proved that using federated

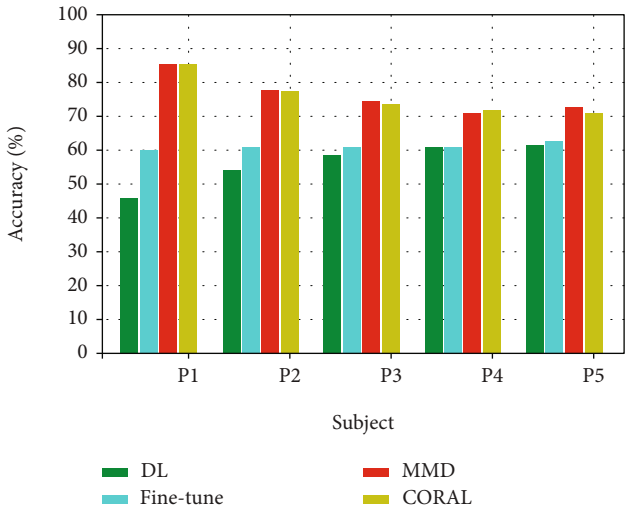


Figure 7: Extending on normal posture tremor test.

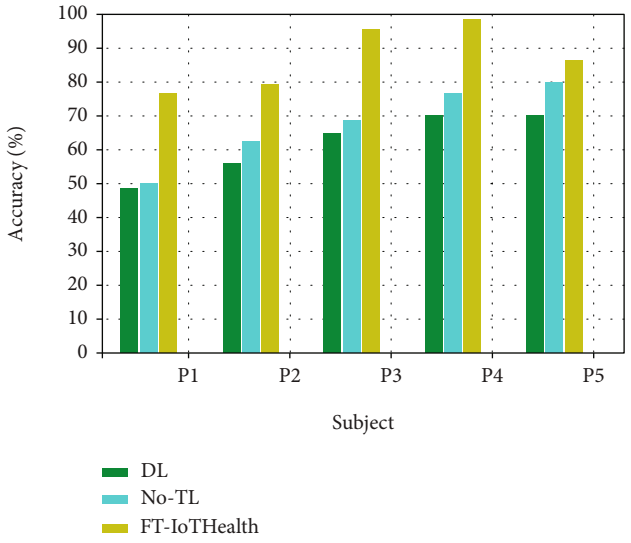


Figure 8: Ablation analysis on arm swing test.

transfer learning technology, the FT-IoMT Health mechanism can achieve effective symptom classification in practical applications.

Consistent with the experimental setup mentioned above, Figures 6 and 7 show the scalability results of the arm swing and normal posture tremor test data, respectively. It can be shown that in most cases, FT-IoMT Health can achieve satisfactory results using fine-tuning or MMD, which also shows that FT-IoMT Health and other transfer learning algorithms are as effective and scalable in practical applications.

For the performance of the given model, we further study ablation analysis (also called sensitivity analysis) to evaluate the two components of joint learning and transfer learning. We apply No-TL to mean an average model without personalize transfer learning. The outcomes are

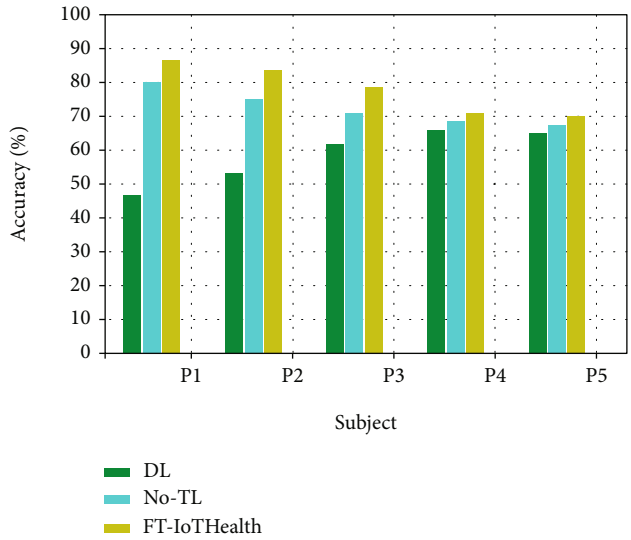


Figure 9: Ablation analysis on normal posture tremor test.

indicated in Figures 8 and 9. It can be seen from the results that both federated learning and transfer learning have made significant achievements to the performance of FT-IoMT Health. Comparing No-TL with DL, it can be seen that the model with federated conditions will increase the classification accuracy, which shows the effectiveness of federated learning. By further comparing No-TL with our federated transfer learning mechanism FT-IoMT Health, it can be seen that integrated with transfer learning technology, each user model will attain better performance in classification. The reasons are as follows. (1) Using federated learning, the server can indirectly aggregate more communication from multiple users to obtain a more general network cloud model. (2) Using transfer learning, users will obtain a more personalized user data model based on the cyber cloud model.

6. Conclusion

In the paper, we propose FT-IoMT Health, which is a federated transfer learning mechanism based on IoMT healthcare. FT-IoMT Health aggregates data from different network users without affecting privacy and security and realizes the user's relatively personalized model learning through transfer learning. The key is feature-based transfer learning technology to overcome various detection methods that lead to variants in network performance. Experiments and applications have verified the validity and accuracy of the mechanism compared to other benchmark methods. Meanwhile, the experimental outcomes also indicate that the transfer learning method enhances the performance of detecting unseen new network malicious data compared with the baseline and proves that FT-IoMT Health can support the detection of new data in different feature spaces. In the future, we will plan to expand FT-IoMT Health through incremental learning to achieve a more personalized, flexible, and efficient healthcare system.

Data Availability

All the data is available in the paper.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors gratefully acknowledge the support from the Shandong National Science Foundation of China (Grant No. ZR202103040468).

References

[1] L. Lyu, C. Chen, J. Yan, F. Lin, and X. Guan, "State estimation oriented wireless transmission for ubiquitous monitoring in industrial cyberphysical systems," *IEEE Transactions on Emerging Topics in Computing*, vol. 7, no. 99, pp. 12–23, 2016.

[2] S. Kurt, H. U. Yildiz, M. Yigit, B. Tavli, and V. C. Gungor, "Packet size optimization in wireless sensor networks for smart grid applications," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 3, pp. 2392–2401, 2017.

[3] Q. Chi, H. Yan, C. Zhang, Z. Pang, and L. D. Xu, "A reconfigurable smart sensor interface for industrial WSN in IoT environment," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 2, pp. 1417–1425, 2014.

[4] R. Zhu, X. Zhang, X. Liu, W. Shu, T. Mao, and B. Jalaian, "ERDT: energy-efficient reliable decision transmission for intelligent cooperative spectrum sensing in industrial iot," *IEEE Access*, vol. 3, pp. 2366–2378, 2015.

[5] B. Holfeld, D. Wieruch, T. Wirth et al., "Wireless communication for factory automation: an opportunity for LTE and 5g systems," *IEEE Communications Magazine*, vol. 54, no. 6, pp. 36–43, 2016.

[6] H. Shariatmadari, R. Ratasuk, S. Iraji et al., "Machine-type communications: current status and future perspectives toward 5g systems," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 10–17, 2015.

[7] A. J. Ward, G. Pirkl, P. Hevesi, and P. Lukowicz, *Towards recognising collaborative activities using multiple on-body sensors*, UbiComp Adjunct, 2016.

[8] D. Bekerman, B. Shapira, L. Rokach, and A. Bar, "Unknown malware detection using network traffic classification," in *2015 IEEE Conference on Communications and Network Security (CNS)*, pp. 134–142, Florence, Italy, 2015.

[9] K. Bartos, M. Sofka, and V. Franc, "Optimized invariant representation of network traffic for detecting unseen malware variants," in *25th {USENIX} Security Symposium ({USENIX} Security 16)*, pp. 807–822, USA: USENIX Association, 2016.

[10] N. Inkster, *China's Cyber Power*, Routledge, 2016.

[11] M. Goddard, "The eu general data protection regulation (gdpr): European regulation that has a global impact," *International Journal of Market Research*, vol. 59, no. 6, pp. 703–705, 2017.

[12] 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.

[13] R. L. Rivest, L. Adleman, and M. L. Dertouzos, "On data banks and privacy homomorphisms," in *Foundations of Secure Computation*, pp. 169–179, Academia Press, 1978.

[14] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[15] K. D. Feuz and D. J. Cook, "Transfer learning across feature-rich heterogeneous feature spaces via feature-space remapping (fsr)," *Acm Transactions on Intelligent Systems & Technology*, vol. 6, no. 1, pp. 1–27, 2015.

[16] R. Perdisci, W. Lee, and N. Feamster, "Behavioral clustering of httpbased malware and signature generation using malicious network traces," in *ser. NSDI'10*, p. 26, USENIX Association, USA, 2010.

[17] A. Valdes and D. Zamboni, *Recent Advances in Intrusion Detection*, Springer, Berlin Heidelberg, 2006.

[18] J. Konen, H. B. Mcmahan, D. Ramage, and P. Richtarik, "Federated optimization: distributed machine learning for on-device intelligence," 2016, http://arxiv.org/abs/1610.02527.

[19] P. Huang, G. Wang, and S. Qin, "Boosting for transfer learning from multiple data sources," *Pattern Recognition Letters*, vol. 33, no. 5, pp. 568–579, 2012.

[20] X. Qin, Y. Chen, J. Wang, and C. Yu, "Cross-dataset activity recognition via adaptive spatial-temporal transfer learning," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 4, pp. 1–25, 2019.

[21] H. B. Mcmahan, E. Moore, D. Ramage, S. Hampson, and B. Arcas, *Communication-efficient learning of deep networks from decentralized data*, arXiv e-prints, 2016.

[22] X. Shi, Q. Liu, W. Fan, P. S. Yu, and R. Zhu, "Transfer learning on heterogenous feature spaces via spectral transformation," in *Proceeding of the IEEE International Conference on Data Mining*, pp. 1049–1054, Sydney, NSW, Australia, 2010.

[23] B. Sun, J. Feng, and K. Saenko, "Return of frustratingly easy domain adaptation," 2015, http://arxiv.org/abs/1511.05547.

[24] Y. Chen, X. Yang, B. Chen, C. Miao, and H. Yu, "Pdassist: objective and quantified symptom assessment of Parkinson's disease via smartphone," in *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Kansas City, MO, USA, 2017.