Information Sciences 614 (2022) 138–152

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

An improved federated learning approach enhanced internet of health things framework for private decentralized distributed data

Chenxi Huang [a], Gengchen Xu [a], Sirui Chen [b], Wen Zhou [c,*], Eddie Y.K. Ng [d], Victor Hugo C. de Albuquerque [e,*]

[a] School of Informatics, Xiamen University, Xiamen 361005, China
[b] School of Software Engineering, Tongji University, Shanghai 201804, China
[c] School of Computer and Information, Anhui Normal University, Wuhu 241000, China
[d] School of Mechanical and Aerospace Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore
[e] Department of Teleinformatics Engineering, Federal University of Ceará, Fortaleza, Fortaleza, CE, Brazil

ARTICLE INFO

Article history:
Received 5 September 2021
Received in revised form 26 August 2022
Accepted 3 October 2022
Available online 7 October 2022

Keywords:
Federal learning
Machine learning
Internet of things
Multi-center privacy protection

ABSTRACT

With the privacy protection increasingly being concerned, Data centralization often heavily causes a big risk of privacy protection, gradually, there is a prevailing trend to enhance the security performance by means of data decentralization, above all, for health care internet of things (IoT) data. Meanwhile, Federated learning has obvious privacy advantages compared to data center training on protecting privacy data. For this reason, a novel framework based on federated learning is presented in this paper, which is suitable for private and decentralized data sets, such as big data in healthy Internet of Things. Specifically, the main work of the puts forward framework includes: (1) Multi-center data collection of healthy Internet of Things. (2) healthy data analysis of Internet of Things. (3) privacy protection method for data of healthy Internet of Things. Finally, related experiments show that the proposed method is feasible, and compared with the traditional methods, it has significantly improved the performance in Quality of Service (QoS) and IoUs indicator.

1. Introduction

In the complicated political situation, network security often symbolizes the security of national situation, which includes system security, application security, data security and so on. As a significant part of network security, data security records users' personal information. In the era of Internet of Everything, the popularity of portable devices has led to a surge in related data. On one hand, it promotes the development of machine learning. It trains the model to correctly predict the samples by the aid of these massive data. On the other hand, data desensitization, information tampering and other events often occur during this process. Whether the data used is legal or not and whether it violates people's privacy is something we need to consider. In view of the above challenges, this paper raises a new policy-making method based on a federated learn-

* Corresponding authors.
E-mail addresses: supermonkeyxi@xmu.edu.cn (C. Huang), zhouwen327@gmail.com (W. Zhou), mykng@ntu.edu.sg (E.Y.K. Ng), victor.albuquerque@ieee.org (Victor Hugo C. de Albuquerque).

https://doi.org/10.1016/j.ins.2022.10.011
0020-0255/

ing framework to solve the problems that need to be settled urgently in the IoT, such as the security of derived data, data storage, privacy leakage and so on.

Taking health big data analysis as an example, when devices are used to extract physiological characteristics from patients, devices and networks are connected to form the IoT. One key point of successful construction is how to store the data. In the past, the traditional methods used centralized management, which often required large servers. In addition, many standby devices were needed to store the data to realize functions such as data migration and recovery. As time goes by, the system performance of query and other operations will gradually decrease and the related costs will usually increase. Concerned about the above problems, we hope to adopt a new distributed way to store data. Combined with historical experience, Zhang et al. [1] explained the advantages and disadvantages of the two modes in the article about XML storage model. In contrast with the monotonous centralized management, it was limited by capacity and efficiency, and distributed system could support subscribers' big data storage requirements more effectively. Under the background of multi-node access in the IoT, it can prevent an error from damaging the usability of the whole database, which not only effectively ensures the security of data storage, but also reduces the processing time of concurrent requests.

The next step is to face the credibility of data use, known as data security. When training machine learning, whether the data is accurate or not will affect the recall rate of the model in the future. Data anomalies and errors may be caused by equipment errors, storage failures and even malicious attacks. In the IoT, the network connection between the computing center and the terminal equipment is not secure, for example, the data transmitted via HTTP will be intercepted or even modified.

Another key is that privacy needs to be protected. Whether these participants in the system can be trusted is a complicated problem, for example, if the transmission of network data will leak data to unrelated participants. Along with the improvement of data theft technology in the network, there are more and more incidents of data theft, and countless users' information is stolen, so people pay more and more attention to privacy, which is also an important focus in this paper. There are still many places related with privacy to protect in the IoT. First, it is sensitive to obtain data privacy from equipment terminals. Before analyzing data, physiological characteristics such as the height and weight of patients will be recorded, which may reveal their identity. Secondly, data may be exposed to others at a certain node in the system during transmission. Finally, there are still various malicious attacks during data storage. Many people have put forward some new data encryption methods. For instance, Deng et al. [2] came up with a scheme of encrypting data based on identity and introducing a fine-grained mechanism for privacy protection. Although this method avoids allowing data encryption, it restricts a receiver from sharing data. In contrast, our framework can deal with multi-user condition by using distributed.

Considering the above challenges, the traditional big data processing methods can no longer meet the requirements of consumers, not only the cloud center can not give a good answer, but also the mobile terminal is often attacked in the IoT system. In order to change this situation, we put forward a new data learning framework. The main contributions include three aspects:

(1) Multi-center distributed big data collection and storage framework: It stores the collected data in fragments to form an efficient and permanent data preservation mechanism. Encrypting and distributing data through distributed network can store data more safely and transparently and achieve better performance at a lower price. This distributed storage framework will become the safest storage mode currently.
(2) Multi-center iterative learning model: It's based on federated learning. Taking the privacy of data into account based on the existing basic technology of artificial intelligence, the shortest time of reducing errors can be quickly achieved under the constraints of existing system resources by multi-center and multi-objective.
(3) Federal learning: We combine it with a data privacy protection learning strategy to help deal with more data tags. As a result, it performs better quality of service.

The remainder of this article is structured as follows: Section II introduces the work on big data management, machine learning and federated learning. Section III describes the work we have done and the differences between the proposed framework and other methods from three aspects. Section IV makes corresponding comparative experiments on the service quality and running performance of the method and other models, and finds that it is more suitable for private and decentralized distributed data sets. Section V summarizes the full-text work and summarizes the shortcomings, and expounds that the algorithm in different scenarios may be improved appropriately for further research and improvement.

2. Related work

To face the overhead caused by data movement and the problem of storing semi-structured and unstructured data, Zhang et al. [3] put forward the distributed storage of spatial big data based on NoSQL database in 2019, and adopted PostgreSQL, MongoDB and other in-memory databases to solve the difficulties of logical structure, which lifted the indexing speed. In 2019, the audit scheme of data integrity was put forward [4], and static and dynamic audits were conducted simultaneously based on fuzzy identity. In terms of hardware, Lin [5] and others adopted the hybrid storage system of hard disk and solid state disk, and considered the data death in the process, and divided the data into three types: active data, invalid data and dying data, which reduced the response time of the migration scheme. Kasu et al. [6] adopted the bottom storage layout of

each endpoint in order to solve the data blocking problem, and applied the Fourier transform mechanism method to the space overhead, which significantly reduced the recovery time. In addition, a new secure data search method [7] was proposed, which enabled the weak trust assumption on the edge server to ensure the confidentiality and data security sharing, so as to effectively reduce the Internet of things devices and reduce the overhead. Wei et al. [8] revealed a new selection method of adaptive coding distributed storage system, which could allow for the data access characteristics of data without considering parity blocks, but this method might increase the overhead of I/O and network. In 2020, someone designed a multi-agent-based cloud storage multi-copy data integrity inspection scheme [9], used the bilinear mapping means to construct the key generation process, and used multi-branch authentication tree to perform multi-copy and data signature to realize authentication, signature and verification of multiple copies. At the same time, Tsou and others [10] applied a secure functional query method, which could deal with untrusted storage servers and adapted it in combination with sequence preserving encryption to protect data privacy and the correctness of query results. Moreover, the data could be searched more efficiently due to the preserved order relationship of encrypted data points.

In the past, many models were sophisticated to deal with big data and had their own limitations, such as the quantitative system capacity. For this reason, Ye et al. [11] once raised a solution to optimize federated learning based on edge computing, which used the average depth network of iterative models to cache content actively. Besides, Khanal et al. [12] made use of active caching in layered FL to cross focus on themselves, so as to predict local content and expand system capacity. In the meantime, constraint learning was used to determine the input weight and deviation based on the difference of samples between classes [13], and voting selection was introduced to improve the accuracy. Li et al. [14] used biological password to protect data, and put forward a new security analysis framework that integrated the selected biological characteristics and decision rules. Eventually, it got rid of the limitations of the original entropy in measurement. Besides, others [15] had found that time and space could be captured as signals by joint deep learning prediction and interpolation of network signals, which reliably protects data privacy. In addition, in order to improve throughput, Jeon et al. [16] also devised a comprehensive solution of distributed processing connection in micro batch mode so that Network broadband capacity would change with time. This method could dynamically process data and improve broadband utilization. The difference was that Moreno and others [17] had found another way to think of new infrastructure in the blockchain to protect data. When other requests tried to change the access strategy, they would warn the requester. Compared with other algorithms, the improvement was to make adjustments in hardware.

Distributed federated learning solutions jointly trained deep learning modeled on combined data, but a lot of communication overhead was required during training int the past decade. Therefore, coefficient ternary compression was proposed by Sattler et al. [18] to cope with their learning environment. The other was the machine learning model based on edge computing in IoT in 2020, in which a federated learning framework [19] with limited delay deadline was designed to avoid excessive training delay, so as to calculate the dynamic client selection with maximum utility. Aminifar [20] found that the integration of extreme randomized tree algorithm in distributed privacy protection could ensure the scalability of the framework, reduce overhead and improve efficiency properly. Hu et al. [21] also presented a differential private federated learning method, which firstly selected the sampling joint average, and then strictly converged the analysis, and verified its practicability and protection for different loss functions. Coincidentally, someone also designed a distributed design method which depends on the local model to increase the computational power [22]. On account of combing with ADMM algorithm, it was still conservative. Last year, someone described a comprehensive overview of the blockchain protocol [23]. More importantly, the incentive engineering mechanism of the protocol for distributed data safely protects the negligence or malicious behavior of the central node, which proved that the unique distributed processing could well support the performance of the service. Ikeda et al. [24] put forward a distributed control algorithm, which considers first-order or second-order integrators, so that it can be obtained by efficient numerical optimization, which can require less network and computing resources. Ultimately, the piecewise federated learning raised in 2020 adapted to various data of large-scale distributed networks [25] based on the characteristics of automatic architecture transformation of periodic evaluation. However, it did not pay attention to the different types of data. Lee et al. [26] designed to enable the incentive management mechanism in the joint data, and extended the Bayesian game in the scene to maximize the benefits of all participants. Similarly, some people had demonstrated a decentralized framework based on blockchain [27], and designed a series of algorithms combined with smart contracts, which could provide more efficient, secure and privacy-aware functions than the original ones. Similarly, Zang et al. [28] fused FL with online unloading algorithm to improve the accuracy of data unloading, and proposed an adaptive adjustment method to improve the learning rate.

Among the methods and strategies of privacy protection, the commonly used methods of traditional data encryption included using public key encryption, and the authorization center was responsible for generating and managing public key certificates for each participant [29]. In order to reduce the overhead, in 2019, a data integrity verification scheme based on a short signature algorithm was presented on the IoT, which could protect data privacy by applying random masking technology [30] without extra overhead in obtaining data. In addition, in 2020, it was presented to adapt a new protocol for privacy protection linear regression (PPLR) [31]. The PPLR protocol based on secret sharing and homomorphic encryption was completed in the process of initialization, aggregation and regression. The client could go offline after submitting the data, which reduced the amount of computation. Or, a game model was designed to protect privacy based on information theory. For example, in the same year, Wu et al. [32] added randomness to real data to protect privacy information, and formulated the minimax privacy problem as a zero-sum game for two people. There was also a classification and construction model of privacy protection decision tree based on differential privacy protection mechanism [33], which selected the avail-

able quality function with low sensitivity to decision and improved the privacy budget allocation method to effectively resist malicious attacks relying on background knowledge. Apart from this, in order to protect the security of image data, Li [34] embedded image segmentation, privacy association separation and sub image allocation into image input. A solution to prevent privacy disclosure was proposed. Although the effect was effective and feasible, it was not intelligent and efficient. And in 2021, someone [35] gave a solution to the privacy protection of the Internet of things, which uses quantum cryptography to establish the key based on the lattice for privacy protection, which is more effective than the traditional digital signature. Proposed privacy protection methods were invoked by many people, for example, according to the privacy model established by domestic mainstream browsers [36], the information leaked by owners' privacy was collected according to different categories. In the same year, the scheme designed by Wang et al. [37] had different privacy for content in fog computing environment, and the applied index number and Laplace mechanisms were used to ensure the difference. Emara et al. [38] proposed two strategies to support the analysis center of distributed data, which could separate the storage and analysis of data and achieve better data security and privacy protection. Bayerlein [39] introduced multi-agent reinforcement learning, combined with experience replay and convolution processing information, designed decentralized partial reference Markov decision, and then balanced the efficiency and security of data collection. In the same year, the location reorganization mechanism was adopted to meet the needs of privacy protection [40]. Compared with other methods, it can have better service effectiveness. Gomez Barrero et al. [41] used the combination of fixed length and sub sampling variable length descriptors to ensure privacy and security. Compared with the general method of online signature, it cost very little in computing overhead, but there are still defects in the processing of variable length data.

Recently, Xu et al [42] proposed a privacy-protected and efficient attribute-based access control (EPABAC) scheme to prevent the privacy leakage of access subject in the decision-making process of ABAC by introducing a novel hash-based binary search tree. Dourado et al. [43] proposed an Internet of Things (IoT) framework for the classification of stroke from CT images applying Convolutional Neural Networks (CNN) in order to identifying a healthy brain, an ischemic stroke or a hemorrhagic stroke. Dourado et al. [44] propsoed a new online approach based on deep learning tools according to the concept of transfer learning to generate a computational intelligence framework for use with the Internet of Health Things (IoHT) devices. Xu et al [4546] proposed a decentralized arbitrable remote data auditing scheme for network storage services based on blockchain techniques and a blockchain-based deduplicatable data auditing mechanism. Xu et al [47] proposed an approach to detect the malicious domain name by extracting and analyzing the features using deep neural network. Han et al. [48] proposed a clustering model for medical applications (CMMA) for cluster head selection to provide effective communication for IoMT based applications.

The above-mentioned multi-center colle ction and storage framework is used for privacy protection and analysis of IoT health data. We have explored a computing solution based on distributed federated learning, which effectively supports data private transmission processing and solves the trouble of the low data reliability and delay in traditional methods. For one thing, the system capacity of the traditional model doesn't have expansibility, which is difficult to deal with the requirements of dynamic increase. For another thing, it is difficult to deal with the storage and processing of all kinds of data. In contrast to the methods proposed by others, it has more excellence in dealing with different data, and can well deal with the defect of less types of identification by traditional methods. Meanwhile, it has good universality and intelligence, and has better accuracy than similar methods in the face of incomplete data. In the next section, we will elaborate on the work we have done in detail.

3. Proposed framework

In this section, we expound how to solve the problem of data storage in the IoT environment. The related tasks will be introduced first. Then we formally defined the solution model.

3.1. Big data acquisition framework and training model

IoT is divided into network layer, application layer and sensing layer. The final goal is to establish an IoT Big Data Health (IoTBDH). After sensor devices in the sensing layer collect data from patients, threads transmit the obtained data to storage nodes through the network. The data collection of wireless sensor networks depends on the sensor nodes deployed in the event area.

The collection framework of big data can be modeled as an interactive model, in which the terminal device will decide the task of data, instead of the cloud processor. The basic principle includes the following aspects. (1) The computing resources of the central server for processing distributed nodes are too expensive. (2) The calculation of the edge handler can make the server only pay attention to its own rights and interests. (3) The edge processing does not need to change the original topology and structure. (4) The edge processor is closer to the user terminal device, speeding up the data transmission and processing speed and reducing the delay time.

The proposed system (as shown in Fig. 1) mainly includes participants, local servers and central servers. First, the global cloud processor will initialize and copy the global model to each local node. That's because the local server lacks data at the beginning, some data will be randomly generated in the early stage. Moreover, the terminal sensing equipment will collect the required information from owners, and the local ones simply train and aggregate the data to form new data. In this pro-

Fig. 1. The Overview of Proposed Framework.

cess, the data is simply processed (like compression), thus reducing the amount of transmitted data. Federal learning is added to protect the privacy of data (which will be described in detail later) during the entire operation.

In IoTBDH system, terminal equipment is not only responsible for collecting sensing data, but also for some sensing calculations. Related handlers are divided into central servers and edge ones, and the algorithm of program allocation can be employed to process the collected raw data in iterative calculation. The trained model and the processed data to the cloud processor will be sent by the local service, and the global handler will compare the obtained models to reduce the loss function. Afterwards, the trained global model will be transmitted to each local equipment through the network, and then this process will be iterated continuously, which will motivate the terminal and global equipment to update the global model continuously according to the collected data, so that the loss value will become smaller and smaller.

Optimization of collection methods includes a multi-center distributed big data collection method, in which we can configure the number of central servers. Compared with centralized communication to collect data, this method can effectively reduce the energy consumption caused by frequent data communication, make the network begin to decouple, and stimulate the data to produce aggregation characteristics. Finally, according to the above-mentioned reasonable scheduling mechanism, the system will ensure the rationality of data transmission at each node, and will not lead to an increase or a steep drop in transmission volume.

3.2. Learning strategies for privacy protection

3.2.1. Data mapping

To ensure that the size of the collected data is within a certain range, which is convenient for data verification and security protection, we introduce a formula to map the data to a certain range. We assume that the generating $f(x)$ is:

$$f(x) = \begin{cases} \frac{x^{\alpha}-1}{\alpha} & if\ \alpha \neq 0 \\ \log(x) & if\ \alpha = 0 \end{cases} \tag{1}$$

where α is used to represent the control constant of the range size, and discuss the cases where α takes different quantities:

When α is 0, the logarithm of X data will be taken;

When α is not 0, first take the α power of X and then subtract 1, and then divide it by α.

In this way, we can limit the value of $\times$ to a specific range $[q_{min}, q_{max}]$.

3.2.2. Normalization processing

After controlling the range of the data, we need to normalize the data for subsequent model training and solve the problem of the original dimension inconsistency. We define a formula that.

$$\hat{q}_{ij} = \frac{f(q_{ij}) - f(q_{\min})}{f(q_{\max}) - f(q_{\min})} \tag{2}$$

where i represents the ith user, j represents the index order of time slice, and q_{ij} represents the collected data value of the ith in the jth time slice. q_{max} and q_{min} are the range of extremum obtained in the previous step. The original data is linearly changed by using min–max standardization, and the number is changed into a decimal between [0,1]. Because the communication and computing resources of the system are limited, it is necessary to normalize the data to improve the iteration speed.

In addition, we also need to do the same operation on the predicted value learned by the local machine. We suppose that the managed result is defined in (3):

$$\hat{p}_{ij} = \frac{1}{1 + e^{-p_{ij}}} \tag{3}$$

Where, obviously, the term $e^{-p_{ij}}$ is always greater than 0, thus ensuring that the normalized predicted value $\hat{p}_{ij}$ is maintained between (0,1). Compared with the original method, the formula we put forward can be calculated without worrying about the change of max and min values due to the emergence of new data, so that the process of finding the optimal solution will become smoother and it will be easier to converge to the optimal solution correctly.

3.2.3. Set loss function

$$\ell_i = \frac{1}{2}\sum_i I_{ij}(\hat{q}_{ij} - \hat{p}_{ij})^2 + \frac{\lambda}{2}\left(||U_i||^2 + ||V^i||^2\right) \tag{4}$$

Here, the term q represents the actual value matrix size, whereas, the term p represents the corresponding prediction matrix, and q_{ij} and p_{ij} respectively represent the elements of the actual value and the predicted value in the matrix. Said we set the learning rate, we will adjust according to the result to adjust the convergence speed. U stands for client data, and v stands for server data. Indeed, the term U_i, V^i represent local server data labeled I and local data of service obtained from global device respectively, and I_{ij} represents identity matrix. Finally, the loss function is proportional to the difference between the predicted value and the actual value. Equation (5) was employed to combine with Equation (4) for local model training in a group from the parameter l.

$$\ell = \sum_i \ell_i \tag{5}$$

which determines the minimization function of the sum of predicted values, sum the user data on each time slice, then converge to find the minimum value according to the value of each iteration, and finally make the loss function reach balance.

3.2.4. Average the data

$$V = \frac{1}{m}\sum_i V^i \tag{6}$$

where m represents the number of global servers configured by us, and v is the average of the calculated values of global servers. To count each transmission model of each local server node, we average the global model data generated by combining the models to eliminate the randomness of one iteration. Merging data to alleviate random interference in the field of big data, and simple averaging also improves the robustness of the system.

3.2.5. Iterative process of training process

$$U_i \leftarrow U_i - \eta\left((\hat{q}_{ij} - \hat{p}_{ij})\hat{p}'_{ij}V^i_j + \lambda U_i\right) \tag{7}$$

where λ is a penalty factor multiplied by local data. It combines the difference between the normalized actual value and the predicts value to iterate the data U_i of the local facility continuously.

$$V^i_j \leftarrow V^i_j - \eta\left((\hat{q}_{ij} - \hat{p}_{ij})\hat{p}'_{ij}U_i + \lambda V^i_j\right) \tag{8}$$

where V_i is the global generated data which iterates λ continuously. Update data to balance after continuous training iteration. In the above two formulas, we encourage local equipment and cloud devices to constantly accept new data to update their status. Based on the above model, the local computer that receives the most data will also play a greater role in updating.

Furthermore, according to Equation (3), we can get the term$\hat{p}'_{ij}$. Whereas, considering some viruses and malicious objects attacks, sometimes, abnormal situations may occur in the predicted data, such as the possibility of data prediction error and data overload, so we define Equation (9) and (10) to predict the data locally.

$$\hat{p}'_{ij} \approx \frac{e^{p_{ij}}}{(e^{p_{ij}}+1)^2} \tag{9}$$

$$p_{ij} = U_i^T V_j^i \tag{10}$$

Where the term U_i is the i^{th} client data and the term V_j^i is the cloud or server data updated by the i^{th} client on the j^{th} timestamp. Besides, the prediction data will be combined with the results of local and global equipment to increase the robustness of the framework. Because the model may not have enough data in the early stage, the model cannot obtain enough distinguishing features for generalization, or the data quality is too low and the categories are unbalanced. Therefore, the combination of exponential distribution and local cloud is adopted for the predicted values in the two formulas. The exponential distribution changes the probability mode of the original statistics, and the greater the difference of data prediction, the smaller the probability. The local and cloud together give the prediction value, that is, having local data can more effectively meet the prediction, and the cloud can avoid over-fitting, considering the possibility of other categories of data that do not appear.

Subsequence, the proposed algorithm can be shown as following.

Algorithm 1: Prediction and analysis of data privacy protection
Input: parameters of network training.
Output: model U, V
  Foreach u_i do
  Local copy of global server data $V_i \leftarrow V$; (assigned to each edge server)
In itialize Ui with random values
Repeat (iteration, loop)
    $p_{ij} \leftarrow U_i^T V_j^i$
    Calculation U_i
    Calculation V_j^i
  Until converge;
  End
  According to the Equation (6), the term V is obtained.
End

The whole algorithm flow can be summarized as outputting the local model U and cloud global model V according to input parameters, network settings and other factors. First, V will be copied to every node that should have U, and uncertain value will be added randomly. Because the prediction function of the system is to update its own model every time, the predicted values are also obtained according to U and V, and then the models of the two types of servers are updated according to the predicted values. This process is repeated in every iteration until the end of learning. The final v can be used directly under similar circumstances. The whole framework can be summarized as a mixed decimal nonlinear statistical problem, in which two kinds of servers interact constantly to realize the task of providing incomplete data prediction services.

Indeed, numerous factors are put into effect on machine learning based on privacy data in distribution environments, such as untrusted third parties, malicious servers. Especially, there are many unknown complex situations, for instance, untrusted third parties, malicious servers. Undoubtedly, these often bring various new challenges and difficulties to federation learning. In this case, traditional training settings often are posed great troubles. To efficiently handle these above-mentioned issues, a differential privacy approach [49] is adopted in our proposed framework to train several related models based on IoT data. Besides, the privacy data training procedure is shown as following Fig. 2.

Nonetheless, to guarantee the robustness and privacy abilities of distributed data, it is common method that add noise to protect privacy, and take a step in the opposite direction of the average noisy gradients. Therefore, clip gradient and add noise is available to enhance privacy during training between clients and servers, clearly, it is hard to effectively achieve that local privacy data are directly harnessed by the attacker in a global untrusted server. Specially, the clip gradient $\bar{g}$ is shown in equation (11).

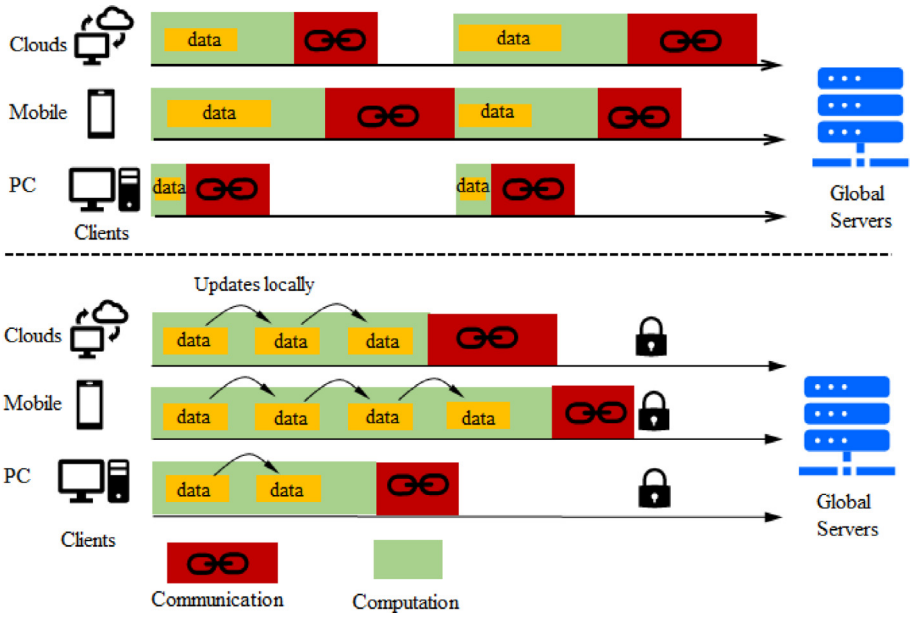

Fig. 2. Privacy data training procedure based on federation learning, In the upper half part, the mini-batch data update and parameters communication between clients and servers, is conducted alternatively, to some extent, there probably exists the risk to data attacking or leaking. In the lower half part, the data is divided into several data pieces, it is continuous that the data parameters are updated locally.

$$\bar{g}_t(d_i) \leftarrow \frac{g_t(d_i)}{\max(1, \frac{\|g_t(d_i)\|_2}{C})} \tag{11}$$

Where, the term di is the privacy data saved in i-th client. Moreover, the threshold C is constant value, it ensures that if$\|g(d_i)\|_2 < C$,then $g_t(d_i)$ is preserved, whereas, it gets scaled down to be of norm C, to stay within a constant privacy budget, here,$C = \sqrt{N}$,the parameter N is the number of training examples.

Furthermore, to improve the confidence of privacy data in decentralized distributed environments, in every gradient update process, add noise would efficiently enhance the privacy and dependence of decentralized data. Therefore, the approximated gradient is gotten to output the model. Specifically, the notation of the approximated gradients $\hat{g}$ in the t epoch can be shown in equation (12).

$$\hat{g}_t \leftarrow \frac{1}{L}(\sum_i \bar{g}_t(d_i) + N(0, \sigma^2 C^2 I)) \tag{12}$$

Where, the term L is the epochs size of the training decentralized dataset, additionally, the term I represents an identity matrix, and the variable σ is the variance of the normal distribution.

$$\theta_{t+1} \leftarrow \theta_t - \eta_t \hat{g}_t \tag{13}$$

Where, the term θ_t, η_t are represented the training parameters and learning rate in the t timestamp, respectively. In addition, the parameter η_t is the learning rate in the t epoch. In this paper, while the learning rate in the range of [0.01,0.07], accuracy indicator is stable.

In the next section, we will discuss the experimental results, add comparative scores, and get the characteristics of the influence of federal learning on the framework.

4. Experiments

We conduct experiments to evaluate the quality of service based on federal learning privacy protection, and choose different independent variables to observe the changes of indicators.

4.1. Experimental environment

This experiment is founded on the local-remote cloud model, and the code is mainly implemented by python. We apply Intel i7 CPU whose operating system is Windows 10 as the basic configuration. The details are as follows: memory 8 GB, main hard disk 240 GB, and graphics card NVIDIA GeForce GTX 1050. In order to validate the feasibility of proposed frame-

work, image classification task is conducted. Initially, we complete MNIST digit recognition task. Besides, the second data set we tested is cifar 100, which has 100 classes, each class contains 600 images (32×32 color images). 500 training images and 100 test images consist of them. The whole data set is divided into super classes such as biology, home and so on. Assorted data can better verify the universality of the raised approach. This published data set provides us with complete data, which is convenient for subsequent tag occlusion processing. A wide variety of pictures have better versatility compared with other data sets.

Furthermore, a real example on health care dataset is to achieve object segmentation, more concretely, the task is to finish the Stent segmentation based on OCT images (the part of the dataset is provided as following website), it is used to further validate the superiority of proposed method.

For the sake of adding some uncertainty to the experiment, we randomly delete some labels from the data to verify the adaptability of the model. Through the analysis of experimental results, the effectiveness of privacy protection algorithm based on federal learning is verified.

4.2. Results

We first test the working effect of the model raised in the previous section, and then use different models (MLP and classic Convolutional Neural Network (CNN)) as the basic framework for comparative experiments. The iteration diagram of the measurement system is as follows (Fig. 3).

The vertical axis represents the loss value obtained by Formula 4 and Formula 5 set in the previous section, that is, the variable L, which indicates the changing trend with the iteration times n. Starting from the input parameters of the system, we can see that at the beginning of time, the loss value does not change much, and the less iteration times do not make the framework much optimized. We speculate that it may be due to the characteristics brought by the model structure itself, the lack of data submitted at the initial stage does not make the algorithm fit the existing data effectively, or it may be too regularized, and it is also a common situation that the loss decline itself is not obvious at the beginning of training time. After 2w times, we can see that loss has an obvious downward trend. Although sometimes the loss value will increase instead, it will generally decrease until it gradually approaches 0 after 12w. Later, we did a comparative experiment on the data set, and used MLP and CNN as the framework to evaluate the test (Table 1).

The accuracy rate of both is over 90 %, and CNN's performance is even better, which is 5 percentage points higher than MLP. We also tested the performance of both in the random and same situation (Table 2).

Experimental results show that CNN's service quality and accuracy are higher than MLP in random processing and the same index. In the random case, the accuracy of both has significantly improved. We surmise that the under-fitting of Equal may be due to the fact that these trained frameworks are based on all data, so the characteristics obtained by the model cannot be well targeted at a small number of data, but the randomly selected data can represent all data better and work better. Finally, we observe the change of loss value in these experiments (solid line indicates MLP, dotted line indicates CNN).

With the increase of iteration times, the loss values of MLP and CNN are decreasing (as shown in Fig. 4), and the loss of MLP is higher under the same iteration times. It means that if MLP requires to reach the same value as MLP, it will take more time and computation. The following figure also shows the change in the correct rate of the two under the environment of missing data labels (Fig. 5).

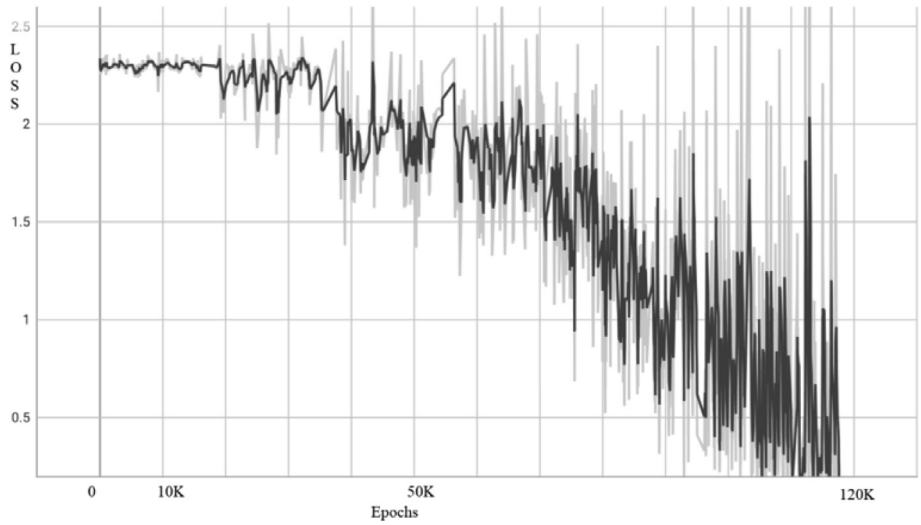

Fig. 3. Iteration Loss Based on FL Privacy Protection Framework.

Table 1
Average accuracy of MLP-CNN method.

Model	Test Acc
MLP	93.32 %
CNN	98.55 %

Table 2
IID: Identity ID of Distribution of Data Amongst Users.

Model	Random	Equal
MLP	89.53 %	72.37 %
CNN	96.65 %	74.77 %

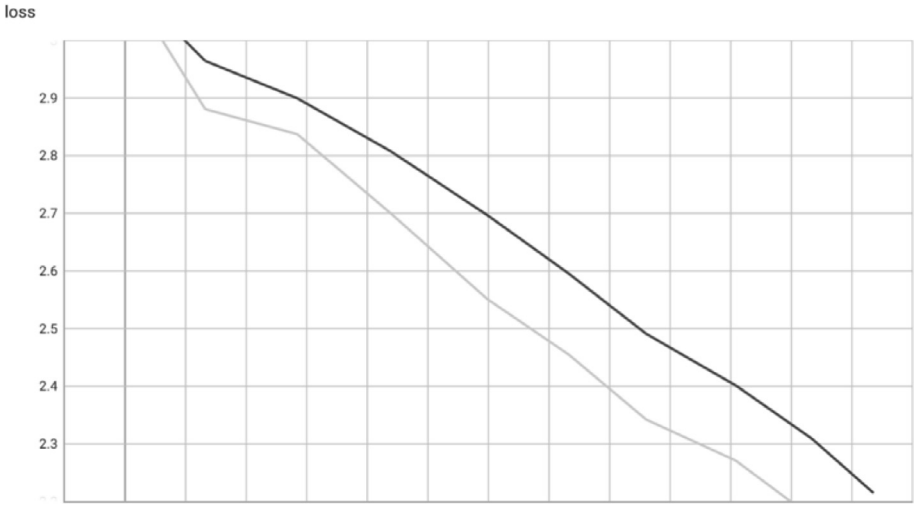

Fig. 4. MLP-CNN Privacy Incident Loss.

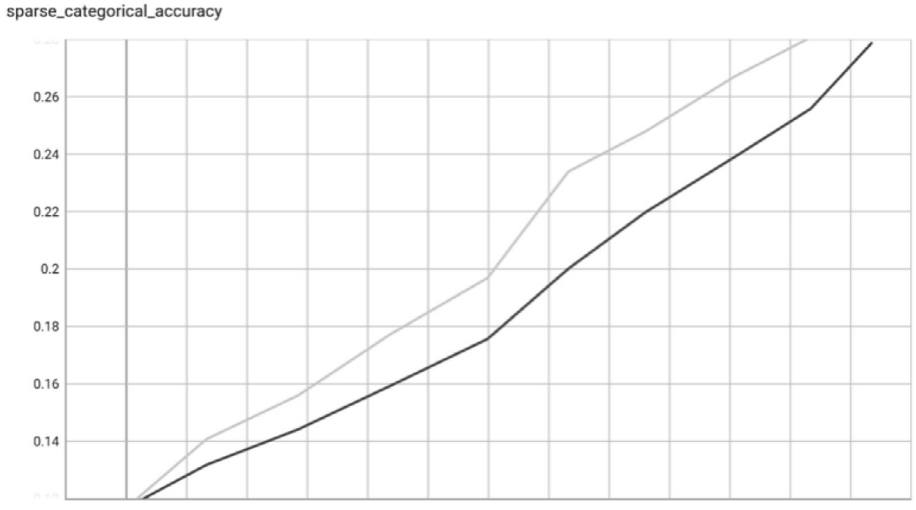

Fig. 5. Sparse Superclass Classification Accuracy.

At first, we can see that the discrimination rates of both are equally low, but with the increase of times, the accuracy rate of both for the lack of integrity data is getting higher and higher, and it is obvious that CNN is faster than MLP. Comparing the results of the original integrity data experiment, we can see that the correct rate has obviously decreased under the same iteration times. This may be caused by sparse data and low data relevance. Because data from different participants are stored in the big health care system, individual differences or the willingness to hide their own labels will lead to missing data in varying degrees.

4.3. Comparison

For the purposes of comparison, we define the absolute error of service $\triangle$, which can be used as the quality index of different approaches of service.

$$\Delta = \frac{1}{N}\sum_{i,j}|q_{ij} - p_{ij}| \tag{14}$$

where the term q_{ij}, p_{ij} indicate the actual value and the predicted value in the i epoch and the j client, respectively. Additionally, the term N is the number of samples, which is divided by it to take the average value, because fewer experiments may produce larger errors.

In reality, the data to be processed won't get all the labels due to the influence of policies, personal wishes or other factors, we choose the degree of data evacuation as an independent variable to conduct a comparative experiment on whether to join federated learning, which also has good practical significance and social background (Table 3).

We can see that the privacy protection method based on FL (federated learning) demands more response time. We hold that compared with the original framework, new steps are added, handy for being compared with the original method. Although it wastes more time to calculate interactively, and the time consumption inevitably increases, it's still worth it, and the throughput and errors have been improved in the follow-up (Table 4).

To observe the adaptability of the algorithm under the background of big data, we analyzed and compared the processing capabilities of the two algorithms at the same time. Due to the evacuation degree that the more tags a piece of data has, the more time it takes, with the increase of the evacuation degree, the throughput gradually decreases, which is a normal phenomenon. In the case of the same evacuation degree, it is obvious that the throughput has been partially improved by adding FL, which verifies the superiority of federated learning in the context of big data (Table 5).

According to the experimental results, under the same background, the error is related to the degree of data evacuation. The more tags a piece of data has, the more reliable the predicted value is. The addition of federated learning also makes the loss value smaller, and the predicted value is obtained more effectively, with smaller absolute errors and smaller $\triangle$. Under the background of more and more data generation, our proposed algorithm can work better in a higher security level environment.

Subsequently, we conducted results on accuracy for different noise levels. The result is as the following figure (Fig. 6).

As shown in Fig. 4, it is not hard to conclude that add noise does not decrease the accuracy, instead, it improves the final accuracy. Specifically, in the beginning, to some extent, more noise would lower the accuracy regardless of the training stage or test stage. However, with the training or test epoch continuing, the accuracy would gradually improve, finally, better accuracy results can be gained. In following table, the framework setting comparison results are shown, in this table, we assume, the local epoch is 10, global round is 20, learning rate is 0.01, SGD optimizer is adopted in our framework, besides, frequently-used independent and identically distributed (I.I.D.) of data is evacuated in our framework (Table 6)

In Fig. 5, the result based on I.I.D data is more stable, the network parameters would be less greatly updated in the training epoch. Nonetheless, in our framework, it is still effectively handled sorts of different data, regardless of I.I.D or Non-I.I.D. Clearly, it can obtain a reasonable result. (Fig. 7).

In order to further validate the superiority of proposed method, we conduct relate experiment on the OCT image segmentation for detecting Bioresorbable Vascular Scaffolds (BVS) task, the OCI images from real health care IoT data (as aforementioned in section 4.1). More specifically, these data is distributed in ten different hospital client, there are different number of OCI image, and there is center server to use as global center. We finish the related experiment circumstance setting and design. Besides, Zhou et al. [49] proposed a segmentation method based on U -shape network, in essence, it heavily relied on larger scale of data learning, however, such health care data often are senstive and refer to person privacy information. while data desensitization will bring a series of problems such as data analysis delay, low data reliability and low data availability. For this reason, Federated learning method can strengthen the application value of deep learning on this field, mean-

Table 3
Evacuation Degree-Response Time Score.

Method	Service quality	Data evacuation degree		
		10 %	20 %	30 %
Privacy protection method	Response time	0.501	0.478	0.411
Privacy Protection Method Based on Federated Learning		0.538	0.481	0.451

Table 4
Evacuation Degree-Throughput Score.

Method	Service quality	Data evacuation degree		
		10 %	20 %	30 %
Privacy protection method	Throughput	17.222	15.889	14.033
Privacy Protection Method Based on Federated Learning		17.982	15.908	14.121

Table 5
Evacuation Degree-Service Quality Score.

Method	Service quality	Data evacuation degree		
		10 %	20 %	30 %
Privacy protection method	Absolute error	13.13 %	9.71 %	8.68 %
Privacy Protection Method Based on Federated Learning	Absolute error	12.11 %	8.86 %	8.06 %

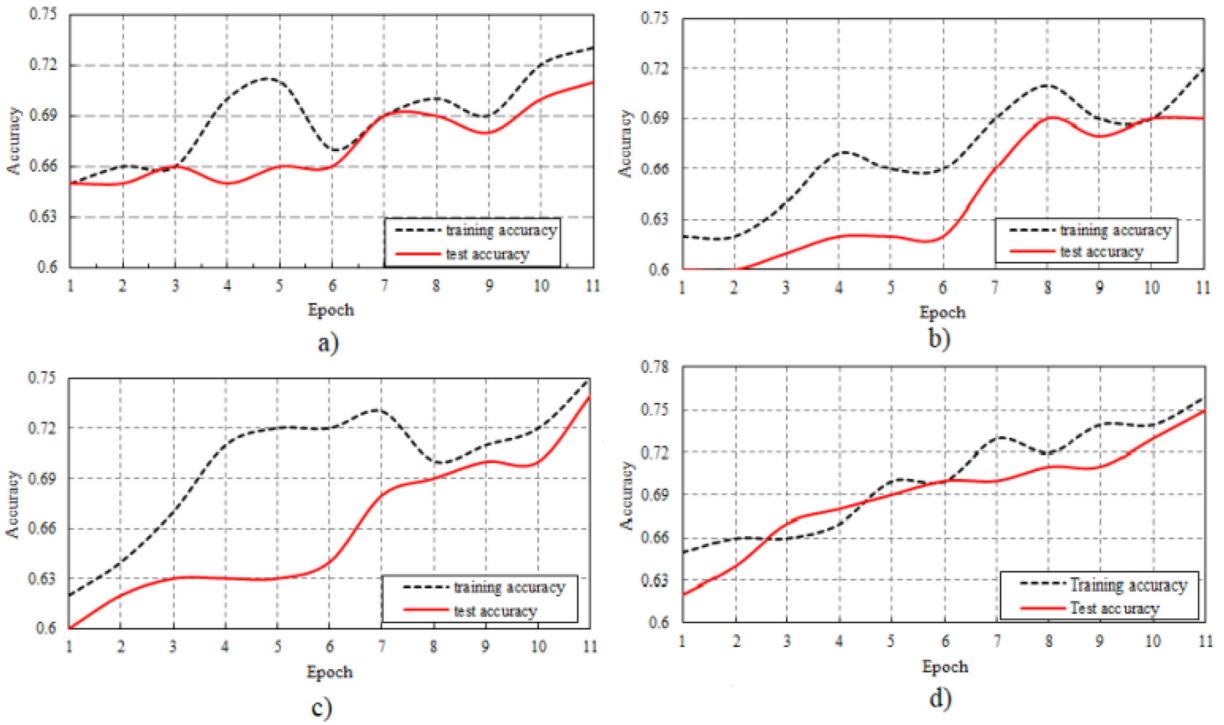

Fig. 6. Comparison results on accuracy for different noise levels, in a),the term σ is 1,and$\sigma = 2$,$\sigma = 3$,$\sigma = 4$ in b),c),d) respectively. Noted that the number of epochs is 110.

Table 6
Comparison between I.I.D AND Non-I.I.D DATA.

Parameter	Avg Training Accuracy	Test Accuracy
I.I. D	56.2 %	44.06 %
Non-I.I.D	44.1 %	45.79 %

while, to some extent, avoid the risk of privacy leaking. Furthermore, IoUs indicator is used to measure the superiority ours and Zhou et al. [49]. The result is as following Fig. 8.

As Fig. 8, our average IoUs of segmentation result and benchmark is better, in particular, the standard deviation is lower, it illustrated that ours is stable and superior than the others, meanwhile the privacy protection can be sufficiently completed and guaranteed.

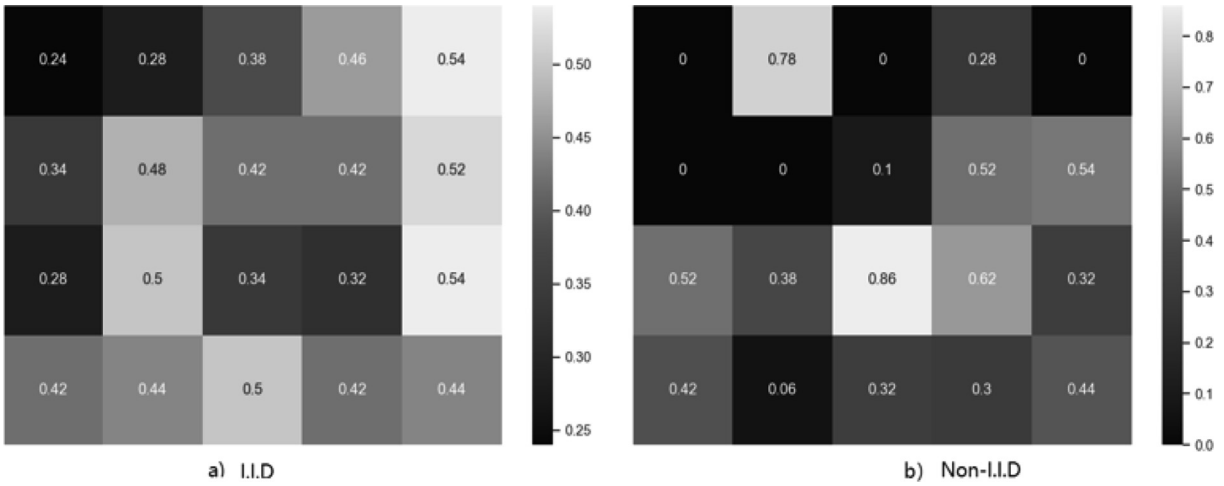

Fig. 7. Training accuracy comparison between I.I.D and Non-I.I. D.

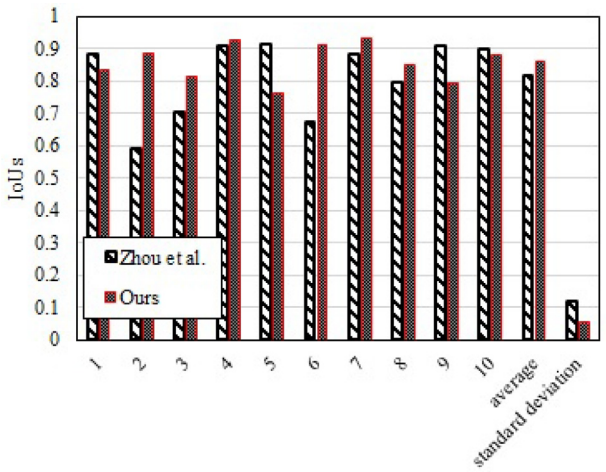

Fig. 8. Comparison on BVS segmentation based on IoUs criterion.

4.4. Limitation

Although the data set employed in this experiment contains various classes, the number of experiment times is relatively small, and the characteristics of it are not adapted completely. Moreover, the network structure of data is not complicated enough. There are not plenty of variables to be tested, and more calculations are demanded in the future. Simultaneously, fewer iterations also means that it takes a lot of time instead of being capable to response immediately. Moreover, what is applicable to this model is the data without labels in privacy. It may not perform better in other situations. As well as the accuracy of measurement indicators, it lacks the addition of time factors. The quality of service should not only be simply judged the error, but also contains the accuracy of data judgment per unit time.

Our work mainly focuses on the comparison of experiments. According to our experimental results, we can see that it has high reliability. CNN is obviously superior to MLP in common, random and absent situations, and has better adaptability. With the quality of service as an index, FL can also deal with large traffic more efficiently. In addition, the experimental results also show that with the increase of data evacuation degree, the classification accuracy gradually increases and the algorithm tends to be stable.

Besides, our method only consider classification and segmentation on image data, for language, video etc. other type data, we don't complete related experiment. As known to all, data privacy protection becoming increasingly important an widespread many different fields. Besides, our basic method for training a model with parameters by minimizing the empirical loss function still following these existed methods, such as differential privacy, in the future, a better and fast convergence

method should be presented to target to health care IoT data, so as to real-time finish related learning task and protect privacy of health care data.

5. Conclusion

In the work of privacy protection method founded on Federated Learning (FL), we realize the framework of interaction between local and cloud servers. We present a novel iteration learning method between global server data and client data, so as to fully consider the privacy of health care data. Besides, added noised data are fully design to enhance the robustness of proposed framework. In the follow-up experiment, Based on famous mechine learning dataset MNIST and Cifar, we evaluate and compare the service quality of MLP and CNN, optimize the overhead, and come up with the possible additional consumption of FL for response time. Correspondingly, FL not only has higher accuracy in the face of incomplete data, but also has greater throughput.Furthermore, in order to validate the widespread of proposed method, we conduct OCT image BVS segmentation task to enhance the privacy performance and robust learning capacity, the data are collected on real health care IoT data, and related experiment environment is designed. The final result shown our method is superiority and more stable, it further shows that the feasibility of our method.

At present, the detection results of the system are satisfactory, but there are still some shortcomings in the work done this moment. For instance, the granularity of data protection is too average, and it is difficult to have excellent work in a scene with a large structure size span. And the universality of the framework lacks more experimental data to prove it. Besides, the learning task is too little, only refered to classification and segementation, in fact, there are more tasks need to attempt, such as recognition and analysis of health care data. What's more, the convergence speed of proposed method is a bit of slow, and it is hard to suitable for real-time task, in this way, it is hard to real apply into actual scenario to conduct relate task.

In the future, we will further try more scenarios of algorithm work, hoping to try to build a new audit scheme that can face dynamic conditions, such as subdividing the picture types on the data set, instead of counting the living and non-living species together. On the one hand, we will pay attention to the dynamic scheme of data emphasis of the algorithm under different circumstances. On the other hand, we strongly suggest that the manufacturer of terminal equipment can formulate a detailed definition of privacy to warn users of hidden risks. Moreover, more learning task need to attempt.

CRediT authorship contribution statement

Chenxi Huang: Conceptualization, Methodology, Software. **Gengchen Xu:** Data curation, Writing – original draft. **Sirui Chen:** Visualization, Investigation. **Wen Zhou:** Writing – review & editing. **Eddie Y.K. Ng:** Software, Validation. **Victor Hugo C. de Albuquerque:** Supervision.

Data availability

Data will be made available on request.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] X.L. Zhang, H. Ding, Y.S. Tan, et al, Centralized and Distributed Storage Model Based on Object-oriented XML, Jisuanji Gongcheng/Computer Engineering 33 (15) (2007) 58–60.
[2] H. Deng, Z. Qin, L. Sha, H. Yin, A Flexible Privacy-Preserving Data Sharing Scheme in Cloud-Assisted IoT, IEEE Internet Things J. 7 (12) (2020) 102–103.
[3] D. Zhang, Y. Wang, Z. Liu, S. Dai, Improving NoSQL Storage Schema Based on Z-Curve for Spatial Vector Data, IEEE Access 7 (2019) 78817–78819.
[4] C. Zhao, X. Li, J. Li, F. Wang, H. Fang, Fuzzy Identity-Based Dynamic Auditing of Big Data on Cloud Storage, IEEE Access 7 (2019) 160459–160471.
[5] M. Lin, R. Chen, J. Xiong, X. Li, Z. Yao, Efficient Sequential Data Migration Scheme Considering Dying Data for HDD/SSD Hybrid Storage Systems, IEEE Access 5 (2017) 23366–23373.
[6] P. Kasu, T. Kim, J.-h. Um, K. Park, S. Atchley, Y. Kim, FTLADS: Object-Logging Based Fault-Tolerant Big Data Transfer System Using Layout Aware Data Scheduling, IEEE Access 7 (2019) 37448–37449.
[7] Y. Tao, P. Xu, H. Jin, Secure Data Sharing and Search for Cloud-Edge-Collaborative Storage, IEEE Access 8 (2020) 15963–15972.
[8] B. Wei, L.-M. Xiao, W. Wei, Y. Song, B.-Y. Zhou, A New Adaptive Coding Selection Method for Distributed Storage Systems, IEEE Access 6 (2018) 13350–13357.
[9] C. Wang, X. Di, Research on Integrity Check Method of Cloud Storage Multi-Copy Data Based on Multi-Agent, IEEE Access 8 (2020) 17170–17178.
[10] Y.-T. Tsou, Y.-L. Hu, Y. Huang, S.-Y. Kuo, SFTopk: Secure Functional Top-k Query via Untrusted Data Storage, IEEE Access 3 (2015) 2875–2890.
[11] Y. Ye, S. Li, F. Liu, Y. Tang, W. Hu, EdgeFed: Optimized Federated Learning Based on Edge Computing, IEEE Access 8 (2020) 209191–209193.
[12] S. Khanal, K. Thar, E.-N. Huh, Route-Based Proactive Content Caching Using Self-Attention in Hierarchical Federated Learning, IEEE Access 10 (2022) 29514–29527.
[13] M. Mengcan, C. Xiaofang, X. Yongfang, Constrained voting extreme learning machine and its application, J. Syst. Eng. Electron. 32 (1) (2021) 209–219.
[14] C. Li, J. Hu, J. Pieprzyk, W. Susilo, A New Biocryptosystem-Oriented Security Analysis Framework and Implementation of Multibiometric Cryptosystems Based on Decision Level Fusion, IEEE Trans. Inf. Forensics Secur. 10 (6) (2015) 1193–1206.

[15] G. Lewenfus, W.A. Martins, S. Chatzinotas, B. Ottersten, Joint Forecasting and Interpolation of Time-Varying Graph Signals Using Deep Learning, IEEE Trans. Signal Inf. Process. Networks 6 (2020) 761–773.
[16] Y. Jeon, K. Lee, H. Kim, Distributed Join Processing Between Streaming and Stored Big Data Under the Micro-Batch Model, IEEE Access 7 (2019) 34583–34598.
[17] R.T. Moreno, J. García-Rodríguez, J.B. Bernabé, A. Skarmeta, A Trusted Approach for Decentralised and Privacy-Preserving Identity Management, IEEE Access 9 (2021) 105788–105804.
[18] F. Sattler, S. Wiedemann, K.-R. Müller, W. Samek, Robust and Communication-Efficient Federated Learning From Non-i.i.d, Data, IEEE Ttransactions on Neural Networks and Learning Systems, Sept. 31 (9) (2020) 3400–3413.
[19] S. Zhai, X. Jin, L. Wei, H. Luo, M. Cao, Dynamic Federated Learning for GMEC With Time-Varying Wireless Link, IEEE Access 9 (Jan. 2021) 10400–10412.
[20] A. Aminifar, M. Shokri, F. Rabbi, V.K.I. Pun, Y. Lamo, Extremely Randomized Trees With Privacy Preservation for Distributed Structured Health Data, IEEE Access 10 (2022) 6010–6027.
[21] R. Hu, Y. Guo, Y. Gong, Concentrated Differentially Private Federated Learning With Performance Analysis, IEEE Open Journal of the Computer Society 2 (2021) 276–289.
[22] Y. Zheng, M. Kamgarpour, A. Sootla, A. Papachristodoulou, Distributed Design for Decentralized Control Using Chordal Decomposition and ADMM, IEEE Trans. Control Network Syst. 7 (2) (2020) 614–626.
[23] D. Xenakis, A. Tsiota, C.-T. Koulis, C. Xenakis, N. Passas, Contract-Less Mobile Data Access Beyond 5G: Fully-Decentralized, High-Throughput and Anonymous Asset Trading Over the Blockchain, IEEE Access 9 (2021) 73963–74016.
[24] T. Ikeda, M. Nagahara, K. Kashima, Maximum Hands-Off Distributed Control for Consensus of Multiagent Systems with Sampled-Data State Observation, IEEE Trans. Control Network Syst. 6 (2) (2019) 852–862.
[25] Y. Sun, H. Esaki, H. Ochiai, Adaptive Intrusion Detection in the Networking of Large-Scale LANs With Segmented Federated Learning, IEEE Access 2 (Dec. 2020) 102–112.
[26] Hua Deng, Zheng Qin, Member, Letian Sha, and Hui Yin, A Flexible Privacy-Preserving Data Sharing Scheme in Cloud-Assisted IoT, *IEEE Internet of Things Journal*, vol. 7, no. 12, pp. 106-112 , Dec. 2020.
[27] L. Jiang, X. Zhang, BCOSN: A Blockchain-Based Decentralized Online Social Network, IEEE Trans. Comput. Social Syst. 6 (6) (2019) 1454–1466.
[28] L. Zang, X. Zhang, B. Guo, Federated Deep Reinforcement Learning for Online Task Offloading and Resource Allocation in WPC-MEC Networks, IEEE Access 10 (2022) 9856–9867.
[29] Hongliang, Ying Yuan, Yuling Chen, Yaxing Zha, Wanying Xi, Bin Jia, and Yang Xin, A Secure and Efficient Data Integrity Verification Scheme for Cloud-IoT Based on Short Signature, *IEEE Access*, vol. 7, pp. 90036-90044, Jun. 2019.
[30] G. Qiu, X. Gui, Y. Zhao, Privacy-Preserving Linear Regression on Distributed Data by Homomorphic Encryption and Data Masking, IEEE Access 8 (2020) 107601–107613.
[31] W.u. Ningbo, C. Peng, K. Niu, A Privacy-Preserving Game Model for Local Differential Privacy by Using Information-Theoretic Approach, IEEE Access 8 (2020) 216741–216751.
[32] L. Zhang, Y. Liu, R. Wang, X. Fu, Q. Lin, Y. Liu, R. Wang, F. Xiong, Q. Lin, Efficient privacy-preserving classification construction model with differential privacy technology, J. Syst. Eng. Electron. 28 (1) (2017) 170–178.
[33] Y. Zhao, L. Yang, Z. Li, L. He, Y. Zhang, P. Model, Detect Privacy Leakage for Chinese Browser Extensions, IEEE Access 9 (2020) 44502–44513.
[34] F. Li, C. Shang, K. Liu, A. Pang, S. Huang, PPM: Privacy Protection Method for Outsourcing Data Entry, IEEE Access 7 (2019) 29745–29753.
[35] L. Malina et al, Post-Quantum Era Privacy Protection for Intelligent Infrastructures, IEEE Access 9 (2021) 36038–36077.
[36] Q. Wang, D. Chen, N. Zhang, Z. Ding, Z. Qin, PCP: A Privacy-Preserving Content-Based Publish-Subscribe Scheme With Differential Privacy in Fog Computing, IEEE Access 5 (2017) 17962–17974.
[37] H. B. McMahan, D. Ramage, K. Talwar, and L. Zhang. Learning differentially private recurrent language models, *International Conference on Learning Representations*, 2018.
[38] T.Z. Emara, J.Z. Huang, Distributed Data Strategies to Support Large-Scale Data Analysis Across Geo-Distributed Data Centers, IEEE Access 8 (2020) 178526–178538.
[39] H. Bayerlein, M. Theile, M. Caccamo, D. Gesbert, Multi-UAV Path Planning for Wireless Data Harvesting With Deep Reinforcement Learning, IEEE Open Journal of the Communications Society 2 (2021) 1171–1187.
[40] Y. Wang et al, LRM: A Location Recombination Mechanism for Achieving Trajectory k -Anonymity Privacy Protection, IEEE Access 7 (2019) 182886–182905.
[41] M. Gomez-Barrero, J. Galbally, A. Morales, J. Fierrez, Privacy-Preserving Comparison of Variable-Length Data With Application to Biometric Template Protection, IEEE Access 5 (2017) 8606–8619.
[42] C.M.J.M. Dourado Jr, S.P.P. da Silva, R.V.M. da Nóbrega, A.C. da S. Barros, P.P.R. Filho, V.H.C. de Albuquerque, Deep learning IoT system for online stroke detection in skull computed tomography images, Comput. Netw. 152 (2019) 25–39.
[43] Dourado, Cmjm , et al., An Open IoHT-based Deep Learning Framework for Online Medical Image Recognition, *IEEE Journal on Selected Areas in Communications*, PP.99(2020):1-1.
[44] Y. Xu, J. Ren, Y. Zhang, C. Zhang, B. Shen, Y. Zhang, Blockchain Empowered Arbitrable Data Auditing Scheme for Network Storage as a Service, IEEE Trans. Serv. Comput. 13 (2) (2020) 289–300.
[45] Y. Xu, C. Zhang, G. Wang, Z. Qin, Q. Zeng, A Blockchain-enabled Deduplicatable Data Auditing Mechanism for Network Storage Services, IEEE Trans. Emerging Top. Comput. 9 (3) (2021) 1421–1432.
[46] Y. Xu, X. Yan, Y. Wu, Y. Hu, W. Liang, J. Zhang, Hierarchical Bidirectional RNN for Safety-enhanced B5G Heterogeneous Networks, IEEE Trans. Network Sci. Eng. 8 (4) (2021) 2946–2957.
[47] Y. Xu, Q. Zeng, G. Wang, C. Zhang, J. Ren, Y. Zhang, An Efficient Privacy-enhanced Attribute-based Access Control Mechanism, Concurrency Computat Pract Exper 32 (5) (2020) 1–10.
[48] T. Han, L. Zhang, S. Pirbhulal, W. Wu, V.H.C. de Albuquerque, A novel cluster head selection technique for edge-computing based IoMT systems, Comput. Netw. 158 (2019) 114–122.
[49] W. Zhou, F. Chen, Y. Zong, et al, Automatic detection approach for bioresorbable vascular scaffolds using a U-shaped convolutional neural network, IEEE Access 7 (2019) 94424–94430.