Predicting the Prognosis of Stroke Patients Based on Personalized Federated Learning

Jie Yang[1], Haoyu Xie[2], Lianfen Huang[2], Zhibin Gao[3], Shaowei Shen[2]*

[1] *National Institute for Data Science in Health and Medicine, Xiamen University, China*
[2] *School of Information, Xiamen University, China*
[3] *Navigation Institute, Jimei University, China*
leoyang@stu.xmu.edu.cn, hyxie@stu.xmu.edu.cn, lfhuang@xmu.edu.cn,
gaozhibin@jmu.edu.cn, 23320231154450@stu.xmu.edu.cn

Abstract

In recent years, the incidence of stroke has been increasing and showing a trend of younger people. Based on the distributed stroke risk assessment modeling scenario, this paper solves the problems of insufficient data and difficult data sharing in medical institutions through federated learning. Considering the features of structured data, the proposed algorithm takes the non-neural network model as the base model, and combines bagging and gradient boosting algorithms to achieve model updating and aggregation. This paper also proposes the model pruning method to realize the personalization of each participant's model and reduces the data transmission cost of the algorithms by separating the weight matrix of the model and the model parameters. Experiments show that the proposed method greatly outperforms existing baseline approaches according to the predictive results, and the accuracy of the personalized model and the global model in the International Stroke Trial (IST) dataset reaches 78.20% and 76.85%, respectively, which has broader application scenarios.

Keywords: Stroke, Non-IID data, Personalized federated learning, Ensemble learning, Machine learning

1 Introduction

Stroke is an acute cerebrovascular disease, which has a high fatality rate and disability rate [1]. In China, the annual death toll caused by stroke is about 1.6 million, accounting for about 1.57% of the total population, which means that stroke has surpassed heart disease as the main cause of death [2]. Furthermore, according to the American Heart Association (AHA), 15 to 30 percent of stroke survivors will face permanent disability. In terms of the great harm of stroke to life and health, it's of great significance to identify the risk of stroke in non-patients and to predict the prognosis of patients after treatment.

At present, medical institutions often have the problem of insufficient data and difficult data sharing when applying big data algorithms, and general neural network is difficult to utilize its advantages in small-scale structured data due to its own features and interpretability issue. Therefore, we propose a personalized federated ensemble learning algorithm in this paper. Our primary contributions can be summarized into three-fold: (1) Aiming at the shortcomings of general neural network in structured data, traditional machine learning models are integrated into federated learning framework. On this basis, the model is updated by integration algorithm. (2) At the same time, in order to avoid the influence of Non-Independent and Identically Distributed (non- IID) data, the proposed algorithm simplifies the structure of the global model through model pruning on the local data of each participant, so as to obtain the local model with personalized structure. (3) The proposed algorithm can reduce the communication cost of the above steps by separating the weight matrix of the model and the model parameters, and makes the algorithm have a wider application scenario.

Experiments are conducted on the global model and the personalized model on the open non-IID dataset. Compared with some existing methods, experiments show the model trained by the proposed algorithm has better predictive performance. It's concluded that the proposed federated learning algorithm has certain advantages in the modeling of structured data among medical institutions, which provides a new scheme to solve the difficult problem of data sharing in stroke risk assessment.

2 Related Works

With technological innovations and progressions, lots of researchers focus on the applications of machine learning algorithms on risk prediction and prognosis of stroke and related cardiovascular diseases, and further use ensemble learning to combine multiple classifiers, which can result in a strong classifier that performs better than any single model [3].

Weng et al. [4] compared RF and GBDT with non-ensemble models including ANN, and the research results proved that ensemble learning significantly improved the accuracy of cardiovascular and cerebrovascular risk prediction. In [5], Heo et al. collected data on 2604 stroke patients and recorded their recovery after 3 months, using mRS scores as a standard to measure the long-term prognosis

*Corresponding Author: Zhibin Gao; E-mail: gaozhibin@jmu.edu.cn
DOI: https://doi.org/10.70003/160792642024112506002

of patients. Based on this data, three modeling algorithms, ANN, RF and LR, were used to make long-term prognosis, and achieved better prediction effect in comparison with ASTRAL prognosis score widely used clinically [6]. Al-Mekhlafi et al. [7] used health records and MRI images for early diagnosis of hemorrhagic stroke, and RFE + RF and AlexNet + SVM algorithms were used to train the models on structured data and image data respectively, and get excellent performance. Reddy et al. [8] designed three different attribute evaluators to select their respective optimal attribute subsets before model training, and three different attribute subsets were then used to train different machine learning models in the model training stage, and the accuracy of heart disease risk prediction was improved after the integration of these models. Jakhar et al. [9] applied the Stacking method to tumor diagnosis, and several machine learning models were used as the base classifier, including RF, Adaboost and GBDT. In 2023, Wang et al. tried to mine the predictive factors related to the survival after one year of stroke from clinical data, and used multiple logistic regression to screen clinical feature variables in the study [10]. Several machine learning methods, including LR and ensemble learning, were used to build death prediction models, among which XGBoost [11] based prediction models were shown to help physicians make more accurate clinical decisions.

The full performance of the model often needs high quality data support. At the same time, some problems caused by data, such as privacy, gradually emerged [12]. In the medical field, the protection of patient privacy is an important principle for research. Federated learning avoids direct private data transmission by sharing model parameters. In order to alleviate the non-IID feature of data from all participants, personalized federated learning has become the focus of federated learning research. The research focus of personalized federated learning is how to update the whole model in the case of non-IID data and how to assign a personalized model that is more suitable for local data distribution to each participant [13].

In 2020, Li et al. proposed FedProx. In order to solve the problem of drift global optimization caused by non-IID data, FedProx set the regularization term of local optimization for each participant to limit the adverse impact of local optimization on the global model [14]. In the same year, Wang et al. proposed FedNova. Based on FedAvg [15], FedNova weighted the local model parameters according to the amount of data of each participant before each epoch of model aggregation, so as to balance the influence of all participants on the global model [16]. Gradient-boosted trees is considered to be the optimal solution for structured data modeling in most cases [17]. In 2021, Wang et al. [18] proposed a personalized federated learning algorithm based on GBDT: Gradient boosting forest (GBF). In the case of non-IID data, GBF outperforms traditional encryption schemes in terms of privacy protection and algorithm performance. In 2022, Gao et al. replaced the tree-based communication scheme with a completely gradient-based communication scheme, and compressed the intermediate information to obtain good model performance and communication cost while combining GBDT and federated learning [19]. In 2020, Wu et al. [15] proposed FedHome, the first personalized federated learning framework in medical scenarios, which provides a personalized health detection model for different families. In 2021, Thwal et al. applied personalized federated learning in clinical decision support systems to realize large-scale clinical data mining while ensuring patient privacy security and overcoming the risk of cyber attacks [20]. In 2023, Gu et al. proposed a federated transfer learning algorithm for efficient communication, which can combine healthcare data from multiple sources to train risk prediction models for the target population, addressing important challenges including population heterogeneity and data sharing difficulties [21].

3 Proposed Approach

In this paper, we propose a personalized federated ensemble learning framework and solve the problems existing in stroke federated learning through the following points: Firstly, referring to the existing federated GBDT algorithm [18], this paper proposes a model aggregation algorithm based on ensemble learning. All participants can employ different machine learning algorithms beyond neural networks to adapt to structured data features and obtain better performing predictive models. Secondly, inspired by the existing model structure simplification scheme, this paper realizes the ensemble model personalization by adjusting the different weights of a single classifier. The non-IID data problem in stroke federated learning scenario can be effectively solved on the basis of the above aggregation algorithm. Finally, in order to achieve the above goals, parameters will be uploaded and downloaded twice for each epoch in the algorithm. By maintaining a weight matrix locally to reduce communication costs, participants ensure wider applicability of the algorithm.

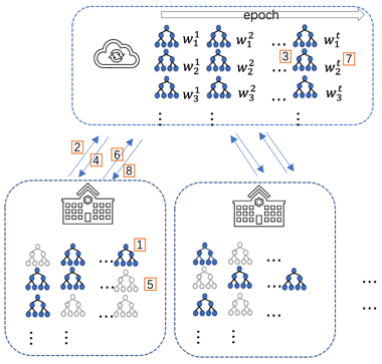

Figure 1. Personalized federated learning algorithm framework for stroke

The algorithm framework is shown in Figure 1. After global initialization, each epoch of the federated learning algorithm will contain the following operations: (1) First upload: Participants verify the current global model on local data, use the residual of the global model as a response variable to train the new model, and upload to the server. The operation corresponds to steps 1 and 2 in Figure 1. (2)

First download: The server uses Bagging to integrate the models uploaded by all participants, and uses the global proportion of the amount of data owned by each participant as the model weight. After that, each participant receives the current epoch of the ensemble model sent by the server. The operation corresponds to steps 3 and 4 in Figure 1. (3) Second upload: each participant verifies the fitting effect of the received ensemble model on the local data and prunes the received ensemble model according to the error. The pruning results will be updated to the local personalized models and uploaded to the server at the same time. The operation corresponds to steps 5 and 6 in Figure 1. (4) Second download: The server aggregates the model pruning results uploaded by all participants, and updates the model weights in the global model based on this. The server then shares the updated model weights with the participants and integrates the model generated by this epoch of Bagging into the global model using gradient boosting. The operation corresponds to steps 7 and 8 in Figure 1.

Finally, each participant synchronously updates the global weight matrix based on the received weights to end the current epoch. Based on the above operations, the proposed algorithm will be described in details in the following three subsections.

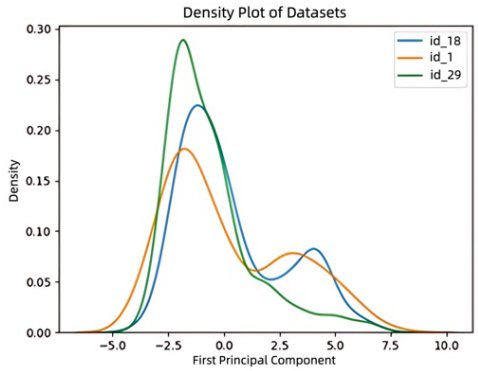

Figure 2. Principal component density curves of data in each region after feature dimension reduction

3.1 Model Aggregation Method

We propose a horizontal federated learning framework consisting of one server and N medical institutions (participants). The kth participant whose sample size is n_k has the local dataset $D_k = \{(x_{k,i}, y_{k,i})\}$, $i \in [1, n_k]$, and the loss function of the global model is defined as follows:

$$L_0 = \frac{1}{N}\sum_{k=1}^{N} n_k L_k \tag{1}$$

where

$$L_k = \frac{1}{n_k}\sum_{\{x_{k,j}, y_{k,j}\} \in D_k} l(f_0(x_{k,j}), y_{k,j}) \tag{2}$$

Where $f_0(x_{k,j})$ is the output of the global model when the input feature vector is $x_{k,j}$, and the loss function l quantifies the error between the predicted value and the ground true label.

In each global epoch, participants use gradient boosting to train a new machine learning model locally. Firstly, each participant verifies the global model of the previous epoch on the local data. In the t th epoch, for the sample $\{x_i, y_i\}$, the prediction result of the global model is:

$$\hat{y}_i^{t-1} = \sum_{j=1}^{t-1}\sum_{k=1}^{N} w_k^j f_k^j(x_i) \tag{3}$$

Where w_k^j is the weight of the model trained by the kth th participant in the j th global epoch in the current global model. $f_k^j(x_i)$ is the model output when input feature vector is x_i. If $t = 0$, the initialization is performed, that is, $\hat{y} = 0$.

Therefore, in the tth epoch, the optimization objective of the global model is:

$$\min_{w_k^t, f_k^t} L^t, L^t = \Sigma_{k=1}^{N}(\Sigma_{i=1}^{n_k} l(y_i, \hat{y}_i^{t-1} + \Sigma_{k=1}^{K} w_k^t f_k^t(x_i)) + \Omega(f_k^t)) \tag{4}$$

Where $\Omega(f_k^t)$ is regularization term to avoid overfitting of the model.

The global model in the federated learning framework is aggregated from the local model of each participant, and the loss of the global model is divided among the participants and optimized by new models trained by participants. Since the training data is not shared, the participants will use the output $f_k^t(x_i)$ of the local model instead of the output $\Sigma_{k=1}^{K} w_k^t f_k^t(x_i)$ of the global model of the current epoch. Therefore, based on the above global optimization objectives, the loss function of the model trained by each participant in the algorithm in the tth epoch is:

$$L_k^t = \Sigma_{i=1}^{n_k} l\left(y_i, \hat{y}_i^{t-1} + f_k^t(x_i)\right) + \Omega(f_k^t)) \tag{5}$$

The above loss function can be understood as that, in each epoch of model training process, each participant in federated learning trains a new model to fit the residuals of the global model on local data.

In the actual situation, each medical institution has different amounts of data, and the trained model is more globally representative for participants with more samples, while participants with less data are prone to overfitting during model training, resulting in degraded global model performance. Therefore, we first weight each participant's model according to their data volume to balance the differences in the distribution of data volume to a certain extent. The weights of the model trained by each participant after the first weighting are:

$$w_k^* = \sum_{k=1}^{m} \frac{n_k}{\sum_{k'=1}^{m} n_{k'}} \tag{6}$$

After participants complete the training of the local models, the models trained in this epoch are uploaded to the server, and the server needs to carry out weighted integration of these models. For the input data feature vector x, the output of the global model is as follows:

$$f_0^t(x) = f_0^{t-1}(x) + \sum_{k=1}^{m} \frac{n_k}{\sum_{k'=1}^{m} n_{k'}} f_k^j(x) \tag{7}$$

In the server, new models uploaded for each epoch will be integrated by Weighted Bagging for global gradient boosting.

3.2 Personalized Model Method Based on Classifier Pruning

Model weighting based on the amount of data can mitigate the non-IID impact to some extent, but non-IID features are not only reflected in the difference in the amount of data, and each participant needs to prune the model to obtain a personalized model that is more suitable for the local data distribution. This section realizes the ensemble model personalization in the federated learning framework by adding an extra model upload and download step in each epoch.

Specifically, participants prune models based on their performance on local data, and eliminate models that perform poorly. As with most federated learning algorithms, the server shares the results of the model aggregation, and each participant calculates the residual of the global model on the local data before this epoch and uses the mean square error to measure the performance of models in fitting the residuals in current epoch, which is defined as follows:

$$e_{k,k'}^t = \frac{1}{n_{k'}} \sum_{i=1}^{n_{k'}} (y_i - \hat{y}_i^{t-1} - f_k^t(x_i))^2 \tag{8}$$

Where $e_{k,k'}^t$ is the fitting error of the model trained by participant k on the dataset of participant k' in epoch t. Each model will be sorted according to the magnitude of the error. By setting the global parameter p, each participant prunes the ensemble model of the current epoch and deletes $[(1-p) * k']$ models with large errors, where [] is the rounding function. The result of the personalization selection is uploaded to the server in the form of a binary vector, which can be defined as follows:

$$v_k^t = [v_{k,1}^t, v_{k,2}^t, \ldots, v_{k,N}^t]^T \tag{9}$$

Where $v_{k,1}^t, v_{k,2}^t, ..., v_{k,N}^t = 0/1$, and $\|V_k\|_1 = k - [(1-p) * k']$. In the personalized model of participant k, the model weights of participant k' of current epoch is:

$$w_{k,k'}^t = \frac{v_{k,k'}^t w_{k'}^*}{\sum_{k'=1}^{N} v_{k,k'}^t w_{k'}^*} \tag{10}$$

Model pruning results of all participants on the server are aggregated to obtain the global model selection result:

$$v^t = \left[v_1^t, v_2^t, \ldots, v_N^t \right]^T \tag{11}$$

Where $v_k^t = \Sigma_{k'=1}^{N} v_{k',k}^t$

The server needs to standardize the selection result of all participants to generate new weights of the models of current epoch. In the weighted global model, the model weight of participant k of current epoch can be defined as follows:

$$w_k^t = \frac{v_k^t w_k^*}{\sum_{k'=1}^{N} v_{k'}^t w_{k'}^*} \tag{12}$$

At the same time, new weights will be shared among the participants at the end of current epoch, completing the synchronization of the global model. By adding weight aggregation for personalized models, the global model can synchronize the model information of each participant in time and obtain better predictive performance. Participants then start model updating for a new epoch on this basis until the scheduled epoch is completed.

3.3 Communication Cost Optimization

Each participant will maintain and update both the global model and the local personalized model at the same time. Considering that the parameters of a single model are not changed in the global iteration, The algorithm shares the weight matrix globally instead of the model, avoiding the cost loss caused by repeated transmission of model parameters.

The specific approach includes: in the initialization process, there are $N*T$ decision tree models in total, the server initializes the global weight matrix M_0 of $N*T$, and each participant initializes two $N*T$ weight matrices $M_{0,k}$ and M_k locally. After the first upload and receive process of each epoch, the server and each participant use gradient boosting to integrate the model trained by each participant in the current epoch, and then in the model selection process of the participant, each participant updates the binary vector of the selected result in the local weight matrix, instead of modifying the structure of the ensemble model. Similarly, in the second receive process, the server only needs to share weight vectors of the current epoch of the ensemble model, and each participant update received weight vectors to the weight matrix of global model which is maintained locally. Related algorithm is given in the Algorithm 1.

Algorithm 1. Personalized federated learning algorithm for stroke

Input: server S, participant $C = \{c_1, c_2, \ldots, c_N\}$, dataset $D_1, D_2, \ldots, D_N$, Global epoch T, and model selection rate p.
Output: global model f_0 , personalized model $\{f_1, f_2, \ldots, f_N\}$
1 global initialization: the server collects data of each participant $\{n_1, n_2, \ldots, n_N\}$, and initializes the model set F_0 and the weight matrix M_0. The participant initializes the model set F_k , the global weight matrix $M_{0,k}$, and the personalized weight matrix M_k ($k \in [1, N]$). The server sends data weight matrix $[w_1^*, w_2^*, ..., w_N^*]^T$ to each participant.
2 For each global epoch t ($t \in [1, T]$), the loop calls **3** through **6**.
3 Each participant calculates the global model loss L^t on the local data according to formula (4) and L_k^t on the local data according to formula (5), creates a new model f_k^t on the local data, and uploads f_k^t.
4 The server adds the received model to F_0 , updates the global model f^t according to formula (6) and formula (7), and sends the current epoch of model $\{ f_1^t, f_2^t, \ldots, f_N^t \}$.
5 Each participant adds $\{ f_1^t, f_2^t, \ldots, f_N^t \}$ to F_k , calculates the fitting error $\{ e_{k,0}^t, e_{k,1}^t, \ldots, e_{k,N}^t \}$ according to formula (8), sorts $\{ f_1^t, f_2^t, \ldots, f_N^t \}$ according to the fitting error, obtains the model selection vector v_k^t according to formula (9), calculates the personalized weight by formula (10), adds it to M_k , and uploads v_k^t to the server.
6 The server calculates the weight $[w_1^t, w_2^t, \ldots, w_N^t]^T$ of the model of the current epoch according to formula (4) and sends it to all participants, and updates it to M_0 . Each participant completes synchronization of $M_{0,k}$ and M_0 after downloading data.
7 The model in server F_0 matches the weights in M_0 , and obtains f_0 according to formula (14). Each participant combines F_k and M_k and gets f_k according to formula (15).

After all epochs are completed, the weight matrix of the global model in the server is defined as follows:

$$M_0 = [m_{kt}]_{N*T} \tag{13}$$

Where $m_{kt} = w_k^t$ ($k \in [1, N]$), $t \in [1, T]$, and the output of the global model for the feature vector x is:

$$f_0(x) = \sum_{t=1}^{T} \sum_{k=1}^{N} m_{kt} f_k^t(x) \tag{14}$$

Similarly, participant k updates the personalized weight matrix M_k according to $w_{k,k'}^t$, and the output of the personalized model for the feature vector x is as follows:

$$f_k(x) = \sum_{t=1}^{T} \sum_{k'=1}^{N} w_{k,k'}^t f_{k'}^t(x) \tag{15}$$

At the beginning of the next epoch, participants can complete synchronization with the global model on the server by combining global weights and saved models.

4 Experimental Results and Analysis

Experiments mainly verify the effectiveness of the above algorithms by simulating the stroke federated learning scenario among medical institutions on IST dataset [22]. For the existing federated aggregation algorithm, personalized federated algorithm, and federated learning algorithm based on traditional machine learning models, the experiments replicates these representative methods and compares them with proposed algorithms, and experimental results present that the proposed algorithms can achieve better performance than other baseline algorithms.

4.1 Datasets

IST was a large prospective trial conducted by the University of Edinburgh in Scotland from 1992 to 1996, which eventually builds a database containing stroke prognosis data from medical institutions. Data are collected at three key time points in the outcome of stroke patients: at the start of the trial (the admission stage), half a month in hospital (the discharge stage), and six months later (the follow-up stage). The dataset contains information on 19,435 patients diagnosed with acute ischemic stroke, with no missing baseline data and less than 1% missing follow-up data. This high-quality large dataset supports the exploration of the clinical course, long-term prognosis, and association of risk factors for stroke. It is worth mentioning that the data are collected from medical institutions in different regions and labeled manually. In this paper, samples of different regions are used as data for various participants in federated learning, and the federated learning framework is constructed on this basis.

4.2 Non–Independent and Identically Distributed Validation

One focus of personalized federated learning research is to reduce the impact of non-IID data on model aggregation. Therefore, in order to ensure the validity of the experiments, this section requires the validation of non-IID features among data of different regions.

First, one preliminary validation is carried out by data visualization. Since the medical dataset has a high feature dimension, it's difficult to directly visualize the distribution of sample points, so we first reduce the dimensionality of the dataset and visualize non-IID features of the data by comparing data feature density curves after dimensionality reduction. As shown in Figure 2, after normalizing each feature, Principal Component Analysis (PCA) is used to reduce the dimensionality of the data, and then some data are randomly selected to visualize their density curve of the first principal component. It's worth noting that both normalization and PCA operations are performed on global data, which means that the transformation and linear combination processes of different datasets are the same. On this basis, there is a significant difference in the density distribution of the first principal component of randomly selected data.

On the basis of visualization, the experiment further quantifies the non-independent and identical distribution

among the data according to indicators. Jensen-Shannon Divergence (JS Divergence) is often used to measure the difference between two distributions [23]. For the two distributions P_i and P_j , JS divergence is calculated as:

$$D_{JS}(P_i \| P_j) = \frac{1}{2} D_{KL}(P_i \| \frac{1}{2}(P_i + P_j)) + \frac{1}{2} D_{KL}(P_j \| \frac{1}{2}(P_i + P_j)) \quad \textbf{(16)}$$

Where D_{KL} is KL divergence [24], and taking the distribution difference on the label as an example, the KL divergence can be calculated as:

$$D_{KL}(P_i \| \frac{1}{2}(P_i + P_j)) = \Sigma_{y \in Y} P_i(y) \log \left(\frac{P_i(y)}{\frac{1}{2}(P_i + P_j)(y)} \right) \quad \textbf{(17)}$$

Where $P_i(y)$ represents the probability that the sample is labeled as y on the dataset with region number i. Therefore, the greater the JS divergence, the greater the difference in the distribution of the data, and usually the JS divergence is less than 0.1, which means that the data comes from the same distribution.

In this section, the data of some regions are selected, and their JS divergence on the label distribution is calculated in pairs, and the heat map is plotted as shown in Figure 3, and JS divergence of the labels between the data in most regions is higher than 0.2, which means that there are significant distribution differences in the data in various regions.

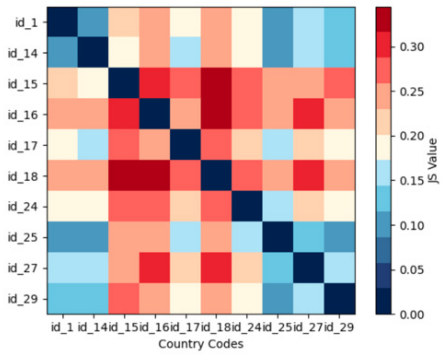

Figure 3. Heat map of the JS values of the data label of different regions

4.3 Global Model Comparison Experiment

According to two research focuses of personalized federated learning, evaluating algorithm performance should start from the following two aspects: the fitting of the global model to the global data and the prediction performance of the personalized models on the local data of the participants, so as to verify the superiority of the algorithm in model aggregation and personalization.

In this section, experiments are conducted to compare and verify the personalized federated ensemble learning algorithm. Based on the features of structured data, In the baseline methods based on neural network adopted in this section, a fully connected network with three hidden layers is selected as the base model. The proposed method and the baseline methods based on the traditional machine learning all use the decision tree model as the base model for iterative training.

In the experiments of this section, the control group algorithms reproduced on the IST dataset includes: (a) FedAvg [15], a classical federated learning algorithm. (b) FedProx [14], an aggregation algorithm that restricts the update of the participant models and introduces regularization to reduce the distance between the local and the global models, so that the optimization direction of the global model doesn't have a large deviation in the training process. (c) FedNova [16], a federated learning algorithm for weighted aggregation of model parameters of all participants, sets higher weights for local models with more steps by setting a globally uniform batch size. (d) GBF-Fed [18], a federated learning algorithm based on GBDT, uses CART decision tree as a training algorithm for each participant.

In the experiment, the data of each participant is split to obtain the test data, that is, the test data of the global model is obtained through regional stratified sampling. In order to ensure the reliability of the experimental results, the experiment sets different numbers of participants, performs multiple sampling under the same number of participants, generates different regional combinations to conduct multiple experiments, and compares the average results of the test results. Four experiments with the number of participants of 5, 10, 15, and 20 are set respectively. Each experiment carries out 100 times of global model sharing, and the number of local iterations in each sharing epoch is 10. Table 1 shows the accuracy rate and AUC value of each algorithm with different number of participants.

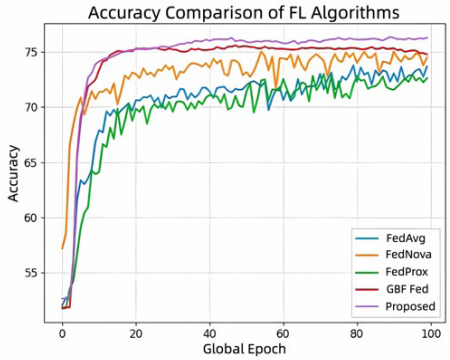

Figure 4. Accuracy curve of each algorithm in a single experiment

Figure 4 shows the change curves of global model accuracy of each federated learning with global epochs in a single experiment when the number of participants is 10. Due to the excellent performance of decision tree models on structured data, the two methods using the decision tree as the base model (GBF-Fed and Proposed) have faster convergence speed and higher final accuracy. Compared with the other three methods, the accuracy of these two methods fluctuates less during the training process, because the methods based on neural network directly update the parameters of the current global models, while these two methods continuously integrate participants' models for

model aggregation and integrate them into the current global models, and the model parameters of the previous epoch are not changed in the subsequent epochs, so the accuracy curve changes more stable.

Figure 5 shows the accuracy changes of algorithms with different number of participants. When the number of participants is small, the global data amount is small, and the model performance is difficult to give full play to. Therefore, when the number of participants increases to 10, the accuracy and AUC values of each algorithm are higher than the control group with the number of participants of 5. When the amount of participants are 15 or 20, the total amount of global data is sufficient, but the accuracy of all federated learning algorithms starts to decrease, except for the centrailized method. The accuracy of the centralized method increases as the number of participants increases and eventually becomes saturated. The reason for adding the centralized method in Figure 5 is to demonstrate that the decline in the performance of federated learning algorithms after more than 10 participants is mainly affected by non-IID feature of each participant, excluding interference from other factors. Moreover, the accuracy of the three algorithms based on neural networks (FedAvg, FedProx and FedNova) has decreased more significantly than that of the other two algorithms (GBF-Fed and Proposed). It can also be seen in Table 1 that the classification accuracy of the proposed method is better than those of baseline methods when AUC values are close, because the decision tree model constantly splits nodes, which is more suitable for binary classification tasks than the neutral network.

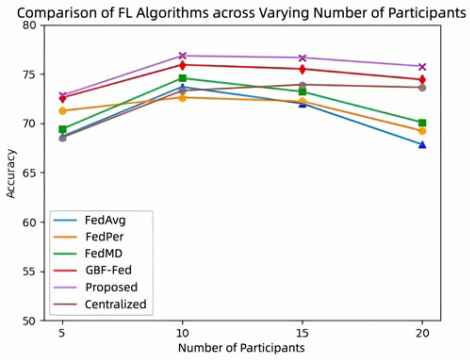

Figure 5. Accuracy of algorithms with different number of participants

Table 1. Global model comparison results. The values show accuracy (ACC%) and area under curve (AUC)

No. of participants	FedAvg		FedProx		FedNova		GBF-Fed		Proposed	
	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC
5	68.65	0.7691	71.26	0.7500	69.40	0.7664	72.58	0.7632	**72.81**	0.7651
10	73.69	0.8197	72.62	0.8167	74.57	0.8195	75.93	0.8205	**76.85**	0.8219
15	71.99	0.808	72.22	0.8113	73.20	0.8117	75.51	0.8152	**76.66**	0.8230
20	67.84	0.8131	69.23	0.8153	70.08	0.8158	74.43	0.8204	**75.78**	0.8206

4.4 Personalized Model Comparison Experiment

In this section, the prediction results of local test samples are calculated based on all participants' personalized models, and the evaluation indicators of the models were compared.

The control algorithms used in this section mainly include: (a) FedAvg [15], a classical federated learning algorithm. (b) Federated transfer learning algorithm FedPer [25], whose personalized model consists of a base layer and a personalized layer. (c) Federated distillation algorithm FedMD [26], whose local model is generated by model distillation of the global model after the participants receive the global model, that is, the global model can be regarded as the teacher model of all the participant models. (d) GBDT-based federated learning algorithm GBF-Fed, where each participant has a local personalized GBDT model.

Due to the different amount of data of each participant, the calculated indicator will be linearly weighted by the amount of data as the final result, which can be regarded as the prediction values obtained by using different personalized models to predict the corresponding regional data in the global validation data. Table 2 shows the weighted accuracy and AUC value of the personalized models obtained by different algorithms with different numbers of participants.

As shown in Figure 6, except for FedAvg, other personalized algorithms don't show the decrease in accuracy caused by non-IID data when the number of participants exceeds 10, which means that these algorithms can avoid non-IID influence to a certain extent. When the number of participants changes, the accuracy of the proposed method changes by up to 1.21%, which is much smaller than that of other algorithms. Compared with GBF-Fed algorithm which also uses the decision tree, the proposed algorithm uses local data for model pruning to avoid global update that may cause performance degradation when the number of participants is small, and can selectively aggregate the global model to improve the prediction effect on local data.

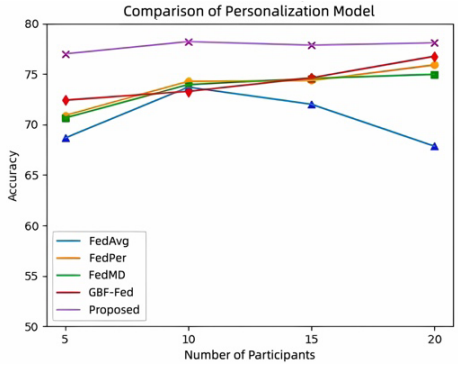

Figure 6. Accuracy of each algorithm's personalized model with different number of participants

In order to avoid the large amount of data in a certain region and the indicator cannot reflect the application effect of algorithms on different regional data, the average accuracy and AUC value of the personalized models are supplemented by Table 3.

According to Table 2 and Table 3, proposed personalized algorithm can obtain better performance than the baseline algorithms under calculation methods of different indicators, which verifies the superiority of the proposed algorithm in federated learning modeling of structured data. Especially when the number of participants is small, the performance of the decision tree-based approaches are much higher than those of other three neural network-based algorithms, because the algorithm of traditional machine models generally requires less training data than the neural network. Therefore, combining the traditional machine learning model with the federated learning algorithm can avoid the deficiency of the traditional federated learning algorithm which is highly dependent on the amount of global data, and better solve the problem of insufficient data in medical institutions.

Table 2. Comparison results of data volume weighted indicators of personalized models. The values show accuracy (ACC%) and area under curve (AUC)

No. of participants	FedAvg		FedPer		FedMD		GBF-Fed		Proposed	
	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC
5	68.65	0.7691	70.89	0.7873	70.64	0.7958	72.40	0.8238	**76.99**	0.8543
10	73.69	0.8197	74.26	0.8090	73.94	0.8149	73.27	0.8289	**78.20**	0.8373
15	71.99	0.808	74.34	0.8067	74.56	0.8168	74.61	0.8201	**77.85**	0.8330
20	67.84	0.8131	75.89	0.8156	74.96	0.8204	76.75	0.8228	**78.08**	0.8352

Table 3. Comparison results of average indicators of personalized models. The values show accuracy (ACC%) and area under curve (AUC)

No. of participants	FedAvg		FedPer		FedMD		GBF-Fed		Proposed	
	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC
5	70.07	0.7627	70.05	0.7622	72.56	0.7956	79.46	0.8249	**80.43**	0.8705
10	74.86	0.8269	74.31	0.8206	75.26	0.8347	77.47	0.8570	**78.99**	0.8532
15	74.30	0.8287	72.04	0.7794	75.85	0.8429	76.96	0.8490	**77.98**	0.8578
20	76.56	0.8161	72.63	0.7464	76.48	0.8382	78.20	0.8626	**79.68**	0.8782

5 Conclusion

Based on the research direction of stroke risk assessment, we propose a personalized federated ensemble learning algorithm to solve the problem of insufficient data and difficult data sharing in medical institutions. Considering the features of structured data, we combine the non-neural network model with federated learning, train new models of all participants in each global epoch by gradient boosting algorithm, and use the bagging algorithm on the training results to update the global model. Also, a model pruning method is proposed for model personalization. In addition, each participant maintains and updates both the global model and the local personalized model to reduce the communication costs of the preceding steps. Finally, we conduct comparative experiments on the global model and the personalized model respectively on the public non-IID dataset. Compared with some existing methods, the model trained by the proposed algorithm has better predictive performance, and it's concluded that the proposed federated learning algorithm has advantages in the modeling of structured data among medical institutions. At the same time, the use of nonneural network model as the base model avoids the shortcomings of the federative learning algorithm based on general neural network in terms of the interpretibility and the amount of training data, and is more suitable for the actual needs of stroke risk assessment.

Acknowledgments

The work presented in this paper was partially supported by 2024 National Natural Science Foundation of China (Grant number 62371406), Key Science and technology Project of Fujian Province (Grant number 2020H6001) and Xiamen Major science and Technology Project (Grant number 3502Z20221026).

References

[1] Global status report on noncommunicable diseases 2014, *World Health Organization*, October, 2014.

[2] R. A. Grysiewicz, K. Thomas, D. K. Pandey, Epidemiology of ischemic and hemorrhagic stroke: incidence, prevalence, mortality, and risk factors, *Neurologic clinics*, Vol. 26, No. 4, pp. 871-895, November, 2008.

[3] H. Han, W. Liu, The coming era of artificial intelligence in biological data science, *BMC bioinformatics*, Vol. 20, No. Suppl. 22, pp. 1-2, December, 2019.

[4] S. F. Weng, J. Reps, J. Kai, J. M. Garibaldi, N. Qureshi, Can machine-learning improve cardiovascular risk prediction using routine clinical data?, *PloS one*, Vol. 12, No. 4, Article No. e0174944, April, 2017.

[5] J. N. Heo, J. G. Yoon, H. Park, Y. D. Kim, H. S. Nam, J. H. Heo, Machine learning–based model for prediction

of outcomes in acute stroke, *Stroke*, Vol. 50, No. 5, pp. 1263-1265, May, 2019.
[6] G. Ntaios, M. Faouzi, J. Ferrari, W. Lang, K. Vemmos, P. Michel, An integer-based score to predict functional outcome in acute ischemic stroke: the ASTRAL score, *Neurology*, Vol. 78, No. 24, pp. 1916-1922, June, 2012.
[7] Z. G. Al-Mekhlafi, E. M. Senan, T. H. Rassem, B. A. Mohammed, N. M. Makbol, A. A. Alanazi, T. S. Almurayziq, F. A. Ghaleb, Deep learning and machine learning for early detection of stroke and haemorrhage, *Computers, Materials and Continua*, Vol. 72, No. 1, pp. 775-796, February, 2022.
[8] K. V. V. Reddy, I. Elamvazuthi, A. A. Aziz, S. Paramasivam, H. N. Chua, S. Pranavanand, Heart disease risk prediction using machine learning classifiers with attribute evaluators, *Applied Sciences*, Vol. 11, No. 18, Article No. 8352, September, 2021.
[9] A. K. Jakhar, A. Gupta, M. Singh, SELF: a stacked-based ensemble learning framework for breast cancer classification, *Evolutionary Intelligence*, Vol. 17, No. 3, pp. 1341-1356, January, June, 2024.
[10] K. Wang, L. Gu, W. Liu, C. Xu, C. Yin, H. Liu, L. Rong, W. Li, X. E. Wei, The predictors of death within 1 year in acute ischemic stroke patients based on machine learning, *Frontiers in Neurology*, Vol. 14, Article No. 1092534, February, 2023.
[11] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, San Francisco, California, USA, 2016, pp. 785-794.
[12] V. S. Cheng, P. C. Hung, Health insurance portability and accountability act (hippa) compliant access control model for web services, *International Journal of Healthcare Information Systems and Informatics (IJHISI)*, Vol. 1, No. 1, pp. 22-39, January-March, 2006.
[13] Q. Wu, K. He, X. Chen, Personalized federated learning for intelligent IoT applications: A cloud-edge based framework, *IEEE Open Journal of the Computer Society*, Vol. 1, pp. 35-44, May, 2020.
[14] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, V. Smith, Federated optimization in heterogeneous networks, *Proceedings of Machine learning and systems*, Austin, TX, USA, 2020, pp. 429-450.
[15] Q. Wu, X. Chen, Z. Zhou, J. Zhang, Fedhome: Cloud-edge based personalized federated learning for in-home health monitoring, *IEEE Transactions on Mobile Computing*, Vol. 21, No. 8, pp. 2818-2832, August, 2022.
[16] J. Wang, Q. Liu, H. Liang, G. Joshi, H. V. Poor, Tackling the objective inconsistency problem in heterogeneous federated optimization, *Advances in neural information processing systems 33 (NeurIPS 2020)*, Virtual, 2020, pp. 7611-7623.
[17] C. N. Ochotorena, C. A. Yap, E. Dadios, E. Sybingco, Robust stock trading using fuzzy decision trees, *2012 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr)*, New York, NY, USA, 2012, pp. 1–8, March, 2012.
[18] F. Wang, J. Ou, H. Lv, Gradient boosting forest: a two-stage ensemble method enabling federated learning of gbdts, *International Conference on Neural Information Processing*, Sanur, Bali, Indonesia, 2021, pp. 75-86.
[19] S. Gao, X. Li, Z. Shi, P. Liu, C. Li, Towards fair and decentralized federated learning system for gradient boosting decision trees, *Security and Communication Networks*, Vol. 2022, No. 1, Article No. 4202084, August, 2022.
[20] C. M. Thwal, K. Thar, Y. L. Tun, C. S. Hong, Attention on personalized clinical decision support system: Federated learning approach, *2021 IEEE International Conference on Big Data and Smart Computing (BigComp)*, Jeju Island, Korea (South), 2021, pp. 141-147.
[21] T. Gu, P. H. Lee, R. Duan, COMMUTE: communication-efficient transfer learning for multi-site risk prediction, *Journal of Biomedical Informatics*, Vol. 137, Article No. 104243, January, 2023.
[22] P. A. Sandercock, M. Niewada, A. Członkowska, The international stroke trial database, *Trials*, Vol. 12, Article No. 101, April, 2011.
[23] B. Fuglede, F. Topsoe, Jensen-Shannon divergence and Hilbert space embedding, *International symposium onInformation theory*, Chicago, IL, USA, 2004, pp. 1-6.
[24] S. Kullback, Information theory and statistics, *Courier Corporation*, 1997.
[25] A. Li, J. Sun, X. Zeng, M. Zhang, H. Li, Y. Chen, Fedmask: Joint computation and communication-efficient personalized federated learning via heterogeneous masking, *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*, Coimbra Portugal, 2021, pp. 42-55.
[26] P. Tsankova, G. Momcheva, Sentiment detection with fedmd: Federated learning via model distillation, *Information Systems and Grid Technologies*, pp. 236-247, 2020.

Biographies

Jie Yang received his M.S. degree in Telecommunications from University of New South Wales, Sydney, Australia, in 2017. He was a lecturer and head of Teaching-research Section of Cloud Computing Technology and Application in Xiamen Institute of Software Techn-ology. He is currently a Ph.D student in Xiamen University, China. His research interests include natural language processing and computer vision, etc.

Haoyu Xie received his M.S. degree in Telecommunications from Xiamen University, Xiamen, China, in 2023. His research interests include federated learning, data mining, etc.

Lianfen Huang received her B.S. degree in Radio Physics in 1984 and PhD. degree in Communication Engineering in 2008 from Xiamen University, Xiamen, Fujian, China. She is now a professor and Doctoral Supervisor in Dept. of Information and Communication Engineering and National Institute for Data Science in Health and Medicine, Xiamen University. She was a visiting scholar in Tsinghua University in 1997 and the Chinese University of Hong Kong in 2012. She is now the key laboratory director of Digital Fujian on IoT Communication, Architecture and Security Technology. Her current research interests include IoT and Wireless of Embed System, Communication Network and Signal Processing.

Zhibin Gao received the B.S. degree in communication engineering, the M.S. degree in radio physics, and the Ph.D. degree in communication engineering from Xiamen University, Xiamen, China, in 2003, 2006, and 2011, respectively. He is a professor of Navigation Institute, Jimei University, Xiamen, China. Previously, he worked with Xiamen University, Xiamen, China, as a senior engineer of Communication Engineering and he was a visiting scholar with University of Washington from 2016 to 2017. His research interests include Internet of Vehicles, marine communications, wireless network resource management, and signal processing.

Shaowei Shen received his B.S. degree from Fuzhou University, Fuzhou, China. He is currently a graduate student in Xiamen University, Xiamen, China. His research interests include Intelligent Healthcare Big Data, Smart Medicine, Artificial Intelligence, etc.