# Federated Learning Based on Extremely Sparse Series Clinic Monitoring Data

LU Feng[1], GU Lin[1], TIAN Xuehua[1], SONG Cheng[1],
ZHOU Lun[2]

(1. School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China；
2. Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430074, China)

DOI: 10.12142/ZTECOM.202203004

Manuscript received: 2022-08-18

**Abstract:** Decentralized machine learning frameworks, e.g., federated learning, are emerging to facilitate learning with medical data under privacy protection. It is widely agreed that the establishment of an accurate and robust medical learning model requires a large number of continuous synchronous monitoring data of patients from various types of monitoring facilities. However, the clinic monitoring data are usually sparse and imbalanced with errors and time irregularity, leading to inaccurate risk prediction results. To address this issue, this paper designs a medical data resampling and balancing scheme for federated learning to eliminate model biases caused by sample imbalance and provide accurate disease risk prediction on multi-center medical data. Experimental results on a real-world clinical database MIMIC-IV demonstrate that the proposed method can improve AUC (the area under the receiver operating characteristic) from 50.1% to 62.8%, with a significant performance improvement of accuracy from 76.8% to 82.2%, compared to a vanilla federated learning artificial neural network (ANN). Moreover, we increase the model's tolerance for missing data from 20% to 50% compared with a stand-alone baseline model.

**Keywords:** federate learning; time-series electronic health records (EHRs); feature engineering; imbalance sample

**Citation** (IEEE Format): F. Lu, L. Gu, X. H. Tian, et al., "Federated learning based on extremely sparse series clinic monitoring data," *ZTE Communications*, vol. 20, no. 3, pp. 27–34, Sept. 2022. doi: 10.12142/ZTECOM.202203004.

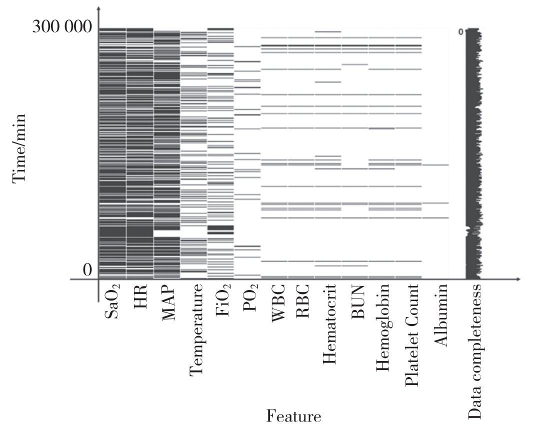
## 1 Introduction

With the increasing availability of electronic health records, artificial intelligence such as neural networks has been widely applied and explored to provide medical risk prediction[1–5]. Neural networks can predict the morbidity risk of patients in advance and find abnormalities in time. Application systems with prediction and early warning functions can be developed based on the neural networks. Predictive and early warning systems have been shown to improve patient outcomes by alerting surgeons to action in advance[6–9]. To train a general and accurate neural network, more medical data from multiple hospitals or medical institutes are desired. However, due to mandatory privacy practices and ethical constraints, hospitals or medical institutes cannot freely share patients' electronic medical records with each other. To this end, decentralized machine learning frameworks, e.g., federated learning[10], are proposed and designed to enable distributed learning with medical data privacy protection.

To establish an accurate and robust medical prediction model, many indicators obtained from different monitoring facilities with temporal integrity must be collected as training data. However, the electronic health record (EHR) system continuously enrolls sparse and error data in the clinic[11–13]. For example, there is a relatively large amount of missing data (Fig. 1) when we extract vital signs, laboratory measurements and other assessment data and combine them statistically according to the time in the Medical Information Mart for Intensive Care IV (MIMIC-IV)[14]dataset. This is partially due to the missing of data acquisition in the current equipment and the hand-filled omission of the nurse. Therefore, it is almost impossible to align data with the timeline to ensure the uniform frequency of patients' different indicators[15–16] (Fig. 1). The clinic monitoring data are incredibly sparse and imbalanced with error and time irregularity, so they cannot be used directly to train risk or disease prediction models[17]. Moreover, since the medical risk, such as septic shock, is critical, training local models using mass imbalance samples with clinic monitoring indicators is most challenging. Therefore, it is essential to propose a risk prediction model with a decentralized machine learning framework that can use these valuable but seriously sparse clinical data.

---

**This work is supported by Hubei Provincial Development and Reform Commission Program "Hubei Big Data Analysis Platform and Intelligent Service Project for Medical and Health".**



BUN: blood urea nitrogen
HR: heart rate
PO$_2$: partial pressure of oxygen
SaO$_2$: oxygen saturation
FiO$_2$: fraction of inspiratory oxygen
MAP: mean arterial pressure
RBC: red blood cell
WBC: white blood cell

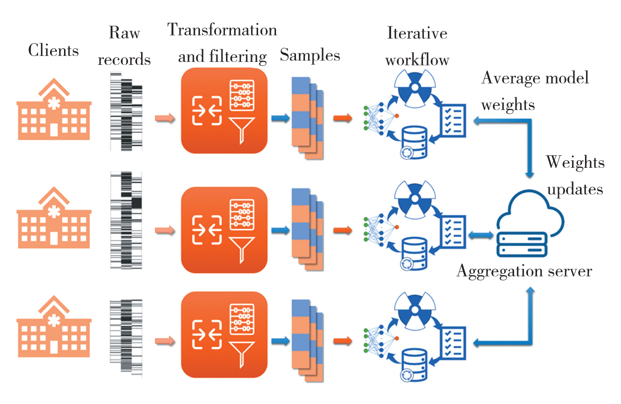
▲**Figure 1. Sparse data with time irregularity in the Medical Information Mart for Intensive Care IV (MIMIC-IV) dataset**

Unfortunately, most existing federated learning methods ignore the data sparsity and errors[18], hence the accuracy and reliability of the prediction models cannot be guaranteed. A multi-center dataset was used in the multi-center federated learning mortality prediction study, but it had different missing values in 11 indicators[19]. Although there are already machine learning frameworks dedicated to the problem of sparse data, this scheme is suitable for more traditional SVMs rather than neural networks with large-scale complex parameters. Solutions in a distributed environment lack validation[20]. Some researchers focus on compressing sparse data to improve the speed of the training phase[21]. For example, Google engineers can use federated learning to predict which emoji a user will likely choose based on sparse data and poorly balanced classes[22]. They tend to include as many participants as possible, but such a global model may not enable the prediction with good performance and reliable robustness[23]. That is to say, low-quality and untrusted participants need to be excluded from federated learning to improve the reliability of the prediction model.

In this paper, we propose a data transformation framework to transform sparse, error-prone and temporally irregular raw data from clients into more accurate patient records for federated learning. We first explore and analyze the MIMIC-IV dataset and obtain two findings about sample imbalance of medical data. One is the difference between the number of negative and positive samples. The other is the missing ratio of features, which results in the model tending to highlight them excessively. Based on these two findings we further improve the quality and reliability of the client and the final model through iteratively training the local models and comprehensively resampling and balancing feature missing rates. Then, we build a horizontal federated learning model of an artificial neural network (ANN) and apply the iterative feature balancing method based on SHapley Additive exPlanations (SHAP) to reduce model bias caused by different proportions of missing features. Finally, the performance of the transforming framework and final model is evaluated based on a real-world clinical database, MIMIC-IV.

## 2 Methodology

Fig. 2 briefly introduces our EHR transformation and iterative learning workflow based on federated learning architecture. Each client consists of transformation and iteration learning, interpretation and resampling. We transform raw clinic records into sample features in the first module, and then input them to a machine learning model such as an ANN and explain the model with SHAP[24].



▲**Figure 2. Overview of the proposed medical data resampling and balancing scheme**

Records with similar medical semantics on the clients are first merged, which are categorized into two types: text and numerical data. To better abstract the features of text indicators, we adopt one-hot encoding to transform the text into a Boolean value. Meanwhile, we remove some abnormal numerical data and encode some numerical data also into a Boolean value.

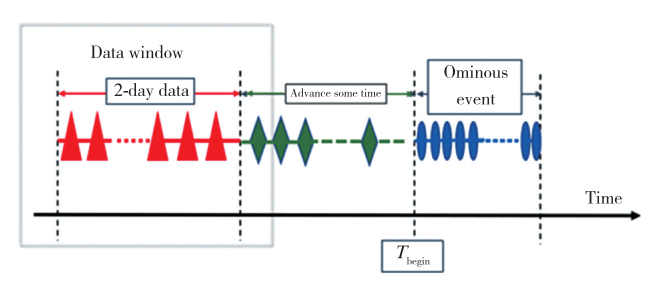
At this time, the encoded indicators are still sparse and have variable data acquisition frequency. It is again a challenge for our clients to train local models. We further transform the indicators using the statistic method, that is, calculating the statistic value of each indicator. The statistic transformation can help to solve the sparse and frequency alignment difficulties.

Finally, after all the patient records are transformed, we will get samples with statistical features on clients to start the iteration of training their local models and explaining the model with SHAP. Combined with the physician's experience and SHAP's results, we can further optimize the models through iteratively resampling the data.

## 2.1 Transformation and Data Filtering

As the analysis above, raw medical monitoring records in certain a time slot are sparse with a high missing rate. This tends to result in lengthy training procedures and unsatisfactory prediction performance. The EHR system records multiple indicators representing similar medical contexts by exploiting the raw time-series data. For example, the difference between left-hand systolic blood pressure and right-hand systolic blood pressure is less than 10 mmHg for most patients. This difference can always be ignored in the clinical judgment for shock disease. Therefore, in septic shock prediction, both left-hand and right-hand blood systolic pressures can represent patients' blood systolic pressure. Motivated by such facts, we first merge raw indicators with related or similar medical concepts into one. In this way, the missing rate of each indicator can be significantly reduced and the quality of data increases.

Here, we design a time window to collect a sequence of item data within a fixed period, as shown in Fig. 3. Because medical data are highly sparse, the acquisition frequencies of different items are uneven and there are NULL or duplicated values. Thus, it is challenging to design a model to predict a disease based on data with a high missing rate.



▲**Figure 3. Data window**

We calculate the timely-sequential data series of each item in the window from maximum, minimum, mean, and other statistical views. By recording the eight statistical views for each medical indicator, we get one training sample for this time window of a patient, and the sample's features are the derived views. In this way, the features' NULL value is reduced but not cleaned up. We calculate the eigenvalue miss rate of each sample. If the NULL values in a sample exceeds a certain threshold, e.g., 50%, we choose to eliminate the whole sample. Otherwise, we keep the sample as a record of the training dataset.

After the data transformation on distributed clients, these samples are used to train local prediction machine learning models for clients of a federated learning system. Here, we take an ANN as an example. The network consists of an input layer of 209 neurons and one hidden layer to fit the data on each client. Each iteration generates a set of SHAP values, an updated model, and a report of feature missing rates. The updates of the local model weights are uploaded to the server of our federated learning system and aggregated by averaging weighted by the dataset size of each client. Then every client gets the global model's weights as their new weights. To validate and improve the correctness of our final model, we take the medical basis, doctor experience and adjustment of the positive and negative sample distribution into consideration.

## 2.2 Iterative Learning and Imbalanced Data Resampling

First, for a client to train the local model, we have dataset *samples*, including the negative sample set $N$ and positive sample set $P$. We use $f$ to denote the element's initial feature set of *samples*, including medical indicators' statistics, such as feature $HR_{\max}$ (the maximum of the indicator item Heart Rate). We also use key-value pairs $\left(f, v\right)$ to represent features $f$ and the eigenvalue of $f$. As shown in Algorithm 1, we first input $\left(f, v\right)$ into the neural network to make the local train.

After inputting $\left(f, v\right)$ into the neural network, we get a model with weights fitted. We can use an AUC evaluation index to measure the advantages and disadvantages of the model. We also take SHAP (Kernel Model), a game-theoretic approach, to explain the output of the model and obtain the classical Shapley values $f_{\mathrm{SHAP}}$ with the highest ranking from SHAP (Kernel Model). SHAP values for each feature are first evaluated on clients. The summary of SHAP values is expected to be similar to the medical basics and physician's experience. For each feature $f_x$ in $f_{\mathrm{SHAP}}$, we use $m(\ )$, the miss rate calculation method deployed as a part of each client, to calculate the miss rate $m\left(P_{f_x}\right)$ and $m\left(N_{f_x}\right)$ based on $f_x$. When the difference between the positive and negative samples is larger than $\varepsilon$, that is, $\left|m\left(P_{f_x}\right)-m\left(N_{f_x}\right)\right|>\varepsilon$, the sample allocation is imbalanced. Then, we need to resample the imbalanced data based on the feature $f_x$.

If the missing rate of positive or negative samples is 100%, that is, $m\left(P_{f_x}\right) == 100\%$ or $m\left(P_{f_x}\right) == 100\%$, the values of the positive or negative samples of the feature $f_x$ are completely missing. Then, we use $delete\left(f_x\right)$ to delete the feature $f_x$ because it has no positive contribution to the prediction model. Otherwise, we use $Cmp\left(m\left(N_{f_x}\right), m\left(P_{f_x}\right)\right)$ to get the samples set $smp_{\max}$ with the largest miss rate and the samples set $smp_{\min}$ with the smallest miss rate based on the feature $f_x$, i. e., $\left(smp_{\max}, smp_{\min}\right) = Cmp\left(m\left(N_{f_x}\right), m\left(P_{f_x}\right)\right)$. After calculating the miss rate $m\left(s\right)$ of each single sample $s\left(s \in smp_{\max}\right)$ about all feature values, we use $sort\left(smp_{\max}\right)$ to sort $smp_{\max}$ according to $m\left(s\right)$. Next, we delete $s$ if its miss rate $m\left(s\right)$ is very high until $\left|m\left(smp_{\max}\right) - m\left(smp_{\max}\right)\right| \leqslant \varepsilon$. Now, the miss rates of the negative and positive samples based on the feature $f_x$ are balanced.

Then we get a new sample set $samples' = smp_{\max} \cup smp_{\max}$ with fewer features in set $f$. Based on the clinical experience, we obtained a new feature set $f'$. If $samples' \neq samples$ or $f' \neq f$, we must rebuild the model and balance the sample repeatedly. Otherwise, the model *NeuralNetworkModel*, the feature set $f'$ and AUC are returned and ready to be uploaded to the server.

**Algorithm 1.** Imbalanced Data Resampling Algorithm

**Require:** *samples* is the dataset which includes negative samples as $N$ and positive samples as $P$, as $samples = N \cup P$; $f$ is an element of *samples* and a set of the initial features for training model.

**Ensure:** $AUC, f', NeuralNetworkModel$;

```
initialization: $samples' = [\ ], f' = [\ ]$
while $samples' \neq samples \| f' \neq f$ do
  $(AUC, NeuralNetworkModel) = FitModel(f, v)$
  $f_{SHAP} = SHAP(NeuralNetworkModel)$
  for each $f_x \in f_{SHAP}$ do
    while $\left| m(P_{f_x}) - m(N_{f_x}) \right| > \varepsilon$ do
      if $m(P_{f_x}) == 100\% \| m(N_{f_x}) == 100\%$ then
        delete$(f_x)$
      else
        $(smp_{\max}, smp_{\min}) = Cmp(m(P_{f_x}), m(N_{f_x}))$
        sort$(smp_{\max})$according to $m(s|s \in smp_{\max})$
        delete $(s)$ until $\left| m(smp_{\max}) - m(smp_{\min}) \right| \leqslant \varepsilon$
      end if
    end while
  end for
  $sample = sample', f = f'$
  $sample' = smp_{\max} \cup smp_{\max}$
  $f' = doctorSelect(f)$
end while
Return: $AUC, f', NeuralNetworkModel$
```

# 3 Experiment and Evaluation

Our experiment is based on the closely related clinical indicators and a public real dataset in the intensive care unit, the MIMIC-IV clinical database. We compared the prediction by the final neural network of our federated learning system with other traditional standalone models for sepsis analysis after integrating data with the proposed method. The performance was comprehensively evaluated through multiple indicators. We explained our final model with the help of SHAP, adjusted the balance of the sample features, combined with the physician's experience, selected the eigenvalue, and tried to set it in line with the physician's clinical cognition. All our experiments were carried out on the same equipment. The computer was equipped with Intel Core i5 8400 CPU, 24 GB memory and GPU acceleration disabled. The training of our selected benchmark algorithm and model is implemented in Python 3.9 for Windows.

## 3.1 Task Background

Sepsis is defined as life-threatening organ dysfunction caused by a dysregulated host response to infection. Septic shock is a subset of sepsis with circulatory and cellular/metabolic dysfunction associated with a higher risk of mortality[25–26]. Sepsis and septic shock are major healthcare problems, affecting millions of people around the world each year and killing as many as one in four (and often more) people worldwide. It sometimes takes only about 24 hours to develop from sepsis to septic shock[27]. Some studies show that the delayed diagnosis and treatment of severe sepsis and septic shock within the first six hours of entering the ICU are closely related to the increased mortality and increased utilization of hospital resources[28–29]. It is a complex problem for clinicians to improve the ability of early clinical recognition, accurately evaluate the condition, and implement reasonable treatment strategies as soon as possible to improve the treatment effect and reduce mortality.

## 3.2 Dataset and Preprocessing

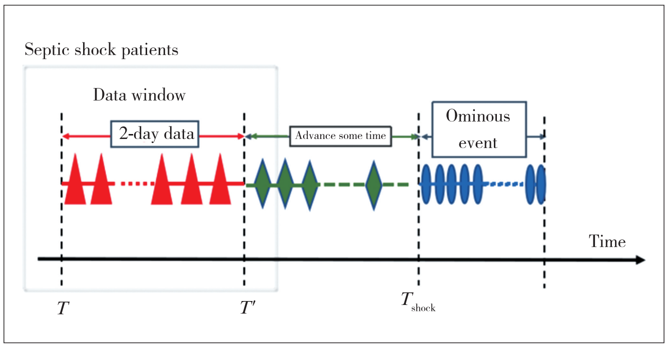
ICU patients are most prone to septic shock. Many patients are difficult to obtain long-term close monitoring in the early stage of shock. And most nurses record data manually. Therefore, the data are too sparse to predict the risk with the existing time series method. To validate our system, we carried out septic shock prediction experiments on MIMIC-IV that includes the data recorded by two different systems in the ICU of Beth Israel Deaconess Medical Center from 2012 to 2019.

More than 30 clinical detection indicators, such as demographics, vital signs, and laboratory values of 121 214 patients were collected, including 6 036 patients diagnosed with septic shock (positive samples) and 115 178 patients who were not diagnosed with any shock (negative samples). These required data include more than 30 monitor features generated from the patients in ICU, which were collected at different times. The time series data is required for subsequent data preprocessing.

All patient characteristics were aligned to the data at the same time, which simplifies the development and testing of the model. The data of each patient were saved to a CSV file. It contains the data of more than 30 index variables such as demographics, vital signs and laboratory results of the patients in ICU. We divided the data sample into four datasets, three of which were for clients in our federated learning system and one was for evaluating our final model.

We sorted the data according to the time of data collection, to facilitate the selection of the later data window and the marking of septic shock events.

As shown in Fig. 4, when a patient's mean arterial pressure (MAP) stays lower than 65 mmhg for at least five minutes, the starting time point of the five minutes is marked as the event of septic shock, which is recorded as $T_{\text{shock}}$. The first 30 minutes of a septic shock event are marked as $t'$ and the window start time before $t'$ is recorded as $T$. The size of the time window is $t' - t = 24$ h. The data in the time window is statistically used as mean



▲**Figure 4. Definition of positive sample**

values, max values and other forms, that is, statistical transformation. A piece of data generated after conversion is a sample data.

For the control experiment, we selected the patient records without any shock and the time series data were generated in the way we selected septic shock records. For these time series data, we selected the data window with the least two days to generate the data, and the sample generated by calculating the statistics is taken as the negative sample. We chose the two days with the least data because the patients of lower risk usually had fewer records in a hospital.

### 3.3 Performance Metrics

A typical objective function in multi-view classification is usually the internal measurement of a classification algorithm. However, the effectiveness and efficiency of this kind of objective function are poor. Instead, we employed multiple external validation metrics to evaluate whether the classification matches the specified label. Selected external indicators used in the following experiments were AUC (the area under the receiver operating characteristic), ACC (accuracy), Pre. (precision), Sens. (sensitivity), Spe. (specificity) and F1 score. AUC and ACC indicate the correctness of the classification; the higher the correctness is, the better. Usually they are the primary indicators. Pre. focuses only on the proportion of true positives among all results classified as positive. Sens. reflects the proportion of positive results detected in all true positive samples. Spe. represents the proportion of all true negative samples classified as negative samples. F1 comprehensively evaluates the recognition rate of positive samples and the overall accuracy.

### 3.4 Result Analysis

We comprehensively compared various performance indicators to determine the effectiveness and accuracy of the prediction. Additionally, we compared our experimental results to the baseline method and ablation studies.

1) Comparison to the baseline method

When the frequency of data collection is high and there is no missing data, the time series model can have higher accuracy, but for sparse data, its accuracy (ACC) decreases. Our design is characterized by data sparseness. Tables 1 and 2 show the comparison of the results of the time series model and our model on the dataset. Compared to the time series model, our final model has a 11% improvement in accuracy. The comparison of AUC, ACC, Rec., Pre., Sens., Spe. and F1 is following.

As can be seen from Table 1, the 5-minute, 10-minute and 15-minute values of AUC in advanced prediction are about 0.51, indicating that the authenticity of prediction is very low and there is no reference value. Most of the ICU clinical data are collected with low frequency and relatively sparse data, so it is not feasible to use the time series model for prediction.

▼**Table 1. LSTM prediction results**

| Time/min | AUC | ACC | Pre. | Sens. | Spe. | F1 |
|---|---|---|---|---|---|---|
| 15 | 0.502 | 0.990 | 0.044 | 0.013 | 0.998 | 0.020 |
| 10 | 0.512 | 0.991 | 0.085 | 0.026 | 0.998 | 0.040 |
| 5 | 0.506 | 0.993 | 0.229 | 0.013 | 0.999 | 0.025 |

ACC: accuracy
AUC: the area under the receiver operating characteristic
LSTM: Long Short-Term Memory
Pre.: precision
Sens.: sensitivity
Spe.: specificity

As shown in Table 2, we use the proposed method to perform transformation and statistical view of data and fit the data with our neural network on each client. The final model has an AUC of 62.8%, while the recall is 26.7% and precision is 87.8%. The AUC is 11% higher than that of Long Short-Term Memory (LSTM) which is not optimized for sparse data. It shows that our transformation and resampling for feature balance can help to improve the model predicting performance.

2) Ablation studies

Unlike other systems, we take SHAP into consideration to pick out the corresponding features and rebalance the proportions of their samples. Table 3 shows the comparison between the initial results and the adjusted results.

▼**Table 2. Prediction results of the proposed model**

| AUC | ACC | Rec. | Pre. | Sens. | Spe. | F1 |
|---|---|---|---|---|---|---|
| 0.628 | 0.822 | 0.267 | 0.878 | 0.989 | 0.267 | 0.409 |

ACC: accuracy
AUC: the area under the receiver operating characteristic
Pre.: precision
Rec.: recall
Sens.: sensitivity
Spe.: specificity

▼**Table 3. Comparison of ablation experiments**

| | AUC | ACC | Rec. | Pre. | Sens. | Spe. | F1 |
|---|---|---|---|---|---|---|---|
| Before Rebalancing | 0.501 | 0.768 | 0.003 | 0.333 | 0.998 | 0.003 | 0.005 |
| After Rebalancing | 0.628 | 0.822 | 0.267 | 0.878 | 0.989 | 0.267 | 0.409 |

ACC: accuracy
AUC: the area under the receiver operating characteristic
Pre.: precision
Rec.: recall
Sens.: sensitivity
Spe.: specificity

We evaluated the model with SHAP to check the factors behind these numbers. Aiming at the top 20 features of SHAP values that had a significant impact on the model, we checked the balance degree for these features in the sample. Because the imbalanced sample missing rate of certain features may bring a high impact, we balanced the positive and negative sample missing rates. After rounds of iterative experiments, with the adjusted sample balance degree, AUC, ACC and Precision are improved by 12.7%, 5.4% and 57.5%, respectively.

After several rounds of iterative adjustments, features were selected by balancing positive and negative samples, combining physician's experience and SHAP evaluation, thus our model has the accuracy of prediction, and the results are consistent with doctors' cognition. Fig. 5 shows SHAP evaluation results before and after iterations. Balanced features derive a more reasonable final model, which is not overly affected by just Platelet related features.

## 4 Conclusions

We propose a federated learning system optimized for sparse time series data with transformation and resampling. In this design, we merge multiple types of sparse time series data for each client and clean them sufficiently so that they are interpretable, design a view for each type of data, and perform statistical data processing, that is, filling in data and selection of data for local models. In this way, we get a batch of valid serialized data ready for clients to fit local artificial neural networks. Next, the network is fitted, the SHAP is used as the interpretation toolkit, and the SHAP ranking fed back is judged by the sample data balanced and physician selected. Due to the sparseness of this kind of data from hospitals and the implementation of the above methods, our federated learning clients depend on less the alignment of time series data on the timeline in the case of extremely poor data quality, which is an effective system to break the limit of high medical time series data missing ratio. The experiments were only conducted in a simple simulated experimental environment. For further verification and exploration, experiments in a real distributed environment and large-scale experiments are required. Due to the research goals, we did not additionally consider the scalability of the system in terms of transmission and induction, which may require more experimental and theoretical analysis.

## References

[1] DENNY J C, COLLINS F S. Precision medicine in 2030—seven ways to transform healthcare [J]. Cell, 2021, 184(6): 1415–1419. DOI: 10.1016/j.cell.2021.01.015

[2] RAGHUNATH S, ULLOA CERNA A E, JING L Y, et al. Prediction of mortality from 12-lead electrocardiogram voltage data using a deep neural network [J]. Nature medicine, 2020, 26(6): 886–891. DOI: 10.1038/s41591-020-0870-z

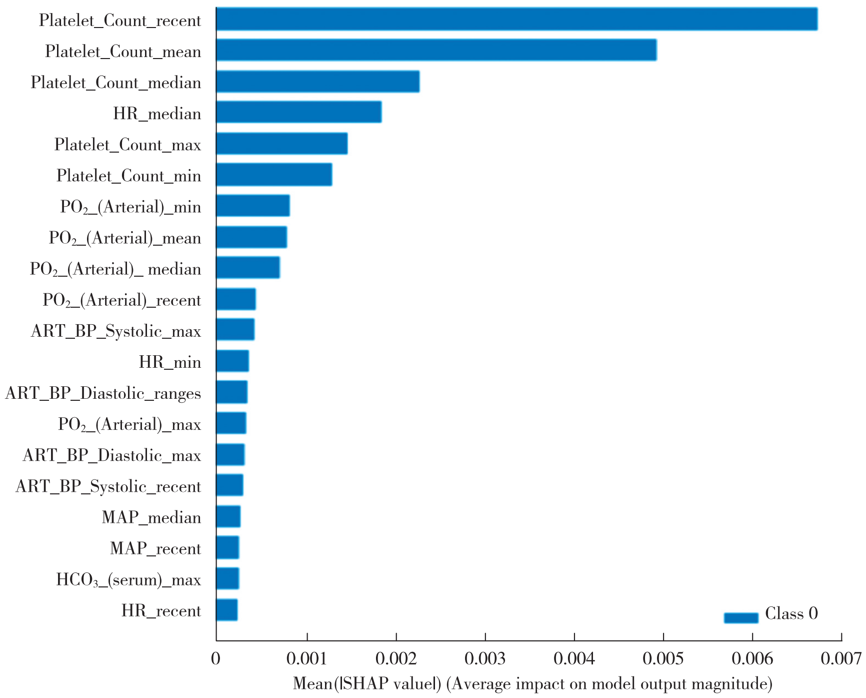
[3] GAO Y, CAI G Y, FANG W, et al. Machine learning based early warning system enables accurate mortality risk prediction for COVID-19 [J]. Nature communications, 2020, 11: 5033. DOI: 10.1038/s41467-020-18684-2

[4] TOMAŠEV N, GLOROT X, RAE J W, et al. A clinically applicable approach to continuous prediction of future acute kidney injury [J]. Nature, 2019, 572(7767): 116–119. DOI: 10.1038/s41586-019-1390-1

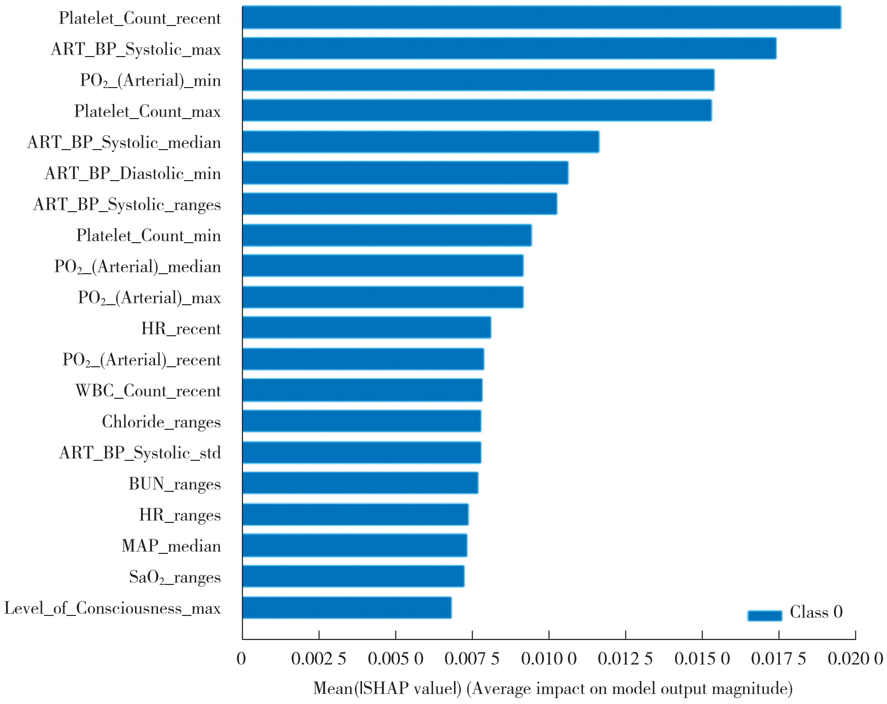
[5] LIANG H Y, TSUI B Y, NI H, et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence [J]. Nature medicine, 2019, 25(3): 433–438. DOI: 10.1038/s41591-018-0335-9

[6] KOCH M. Artificial intelligence is becoming natural [J]. Cell, 2018, 173(3): 531–533. DOI: 10.1016/j.cell.2018.04.007

[7] BACCHI S, TAN Y, OAKDEN-RAYNER L, et al. Machine learning in the prediction of medical inpatient length of stay [J]. Internal medicine journal, 2022, 52(2): 176–185



(a) Before iterations



(b) After iterations

ART_BP: arterial blood pressure
BUN: blood urea nitrogen HCO3 (bicarbonate)
HR: heart rate
MAP: mean arterial pressure
$PCO_2$: partial pressure of carbon dioxide
$PO_2$: partial pressure of oxygen
$SaO_2$: oxygen saturation
SHAP: Shapley Additive Explanation
WBC: white blood cell

▲**Figure 5. Comparison of SHAP values**

[8] VAROQUAUX G, CHEPLYGINA V. Machine learning for medical imaging: methodological failures and recommendations for the future [J]. NPJ digital medicine, 2022, 5(1): 1 – 8

[9] COUDRAY N, OCAMPO P S, SAKELLAROPOULOS T, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning [J]. Nature medicine, 2018, 24(10): 1559 – 1567. DOI: 10.1038/s41591-018-0177-5

[10] YANG Q, LIU Y, CHENG Y, et al. Synthesis lectures on artificial intelligence and machine learning: federated learning [M]. Berlin Heidelberg, Germany: Springer, 2019: 1 – 207. DOI: 10.2200/s00960ed2v01y201910aim043

[11] KELLY C J, KARTHIKESALINGAM A, SULEYMAN M, et al. Key challenges for delivering clinical impact with artificial intelligence [J]. BMC medicine, 2019, 17(1): 195. DOI: 10.1186/s12916-019-1426-2

[12] LI R W, CHEN Y, RITCHIE M D, et al. Electronic health records and polygenic risk scores for predicting disease risk [J]. Nature reviews genetics, 2020, 21(8): 493 – 502. DOI: 10.1038/s41576-020-0224-1

[13] SHILO S, ROSSMAN H, SEGAL E. Axes of a revolution: challenges and promises of big data in healthcare [J]. Nature medicine, 2020, 26(1): 29 – 38. DOI: 10.1038/s41591-019-0727-5

[14] JOHNSON A, BULGARELLI L, POLLARD T, et al. MIMIC-IV-ED [DB]. PhysioNet, 2021. DOI: 10.13026/as7t-c445

[15] LI J, YAN X S, CHAUDHARY D, et al. Imputation of missing values for electronic health record laboratory data [J]. NPJ digital medicine, 2021, 4(1): 147. DOI: 10.1038/s41746-021-00518-0

[16] WU J R, VODOVOTZ Y, ABDELHAMID S, et al. Multi-omic analysis in injured humans: patterns align with outcomes and treatment responses [J]. Cell reports medicine, 2021, 2(12): 100478. DOI: 10.1016/j.xcrm.2021.100478

[17] WEERAKODY P B, WONG K W, WANG G J, et al. A review of irregular time series data handling with gated recurrent neural networks [J]. Neurocomputing, 2021, 441: 161 – 178. DOI: 10.1016/j.neucom.2021.02.046

[18] KUMAR Y, SINGLA R. Federated learning systems for healthcare: perspective and recent progress [J]. Federated learning systems, 2021: 141 – 156

[19] VAID A, JALADANKI S K, XU J, et al. Federated learning of electronic health records improves mortality prediction in patients hospitalized with covid-19 [EB/OL]. (2020-08-11)[2021-11-21]. https://www.medrxiv.org/content/10.1101/2020.08.11.20172809v1

[20] SATTLER F, WIEDEMANN S, MULLER K R, et al. Robust and communication-efficient federated learning from non-i.i.d. data [J]. IEEE transactions on neural networks and learning systems, 2020, 31(9): 3400 – 3413. DOI: 10.1109/TNNLS.2019.2944481

[21] RAMASWAMY S, MATHEWS R, RAO K, et al. Federated learning for emoji prediction in a mobile keyboard [EB/OL]. (2019-06-11)[2021-11-21]. https://arxiv.org/abs/1906.04329

[22] KANG J W, XIONG Z H, NIYATO D, et al. Reliable federated learning for mobile networks [J]. IEEE wireless communications, 2020, 27(2): 72 – 80. DOI: 10.1109/MWC.001.1900119

[23] WEI R M, WANG J Y, SU M M, et al. Missing value imputation approach for mass spectrometry-based metabolomics data [J]. Scientific reports, 2018, 8(1): 663. DOI: 10.1038/s41598-017-19120-0

[24] LUNDBERG S, LEE S I. A unified approach to interpreting model predictions [EB/OL]. (2017-11-25)[2021-12-05]. https://arxiv.org/abs/1705.07874

[25] ANNANE D, BELLISSANT E, CAVAILLON J M. Septic shock [J]. The lancet, 2005, 365(9453): 63 – 78. DOI: 10.1016/S0140-6736(04)17667-8

[26] ASTIZ M E, RACKOW E C. Septic shock [J]. The lancet, 1998, 351(9114): 1501 – 1505. DOI: 10.1016/S0140-6736(98)01134-9

[27] RIVERS E P, MCINTYRE L, MORRO D C, et al. Early and innovative interventions for severe sepsis and septic shock: taking advantage of a window of opportunity [J]. CMAJ, 2005, 173(9): 1054 – 1065. DOI: 10.1503/cmaj.050632

## Biographies

**LU Feng** received her MS and PhD degrees in computer science from Huazhong University of Science and Technology, China in 1997 and 2006. She is currently an associate professor in School of Computer Science and Technology, Huazhong University of Science and Technology. Her current research interests include big data, artificial intelligence and distributed computing. She has authored two books and over 20 papers in refereed journals and conferences in these areas. She is a member of CCF and a senior member of the first Session of Chinese Hospital Association, Health Data Application and Management Committee. She was the recipient of three teaching achievement and curriculum development awards.

**GU Lin** (lingu@hust.edu.cn) received her MS and PhD degrees in computer science from University of Aizu, Fukushima, Japan in 2011 and 2015. She is currently an associate professor in School of Computer Science and Technology, Huazhong University of Science and Technology, China. Her current research interests include serverless computing, network function virtualization, cloud computing, software-defined networking, and data center networking. She has authored two books and over 40 papers in refereed journals and conferences in these areas. She is a member of IEEE and a senior number of CCF.

**TIAN Xuehua** received her MS degree from the School of Computer Science and Technology, Huazhong University of Science and Technology, China in 2022. She works on medical data mining and machine learning. She mainly focuses on working with sparse time series data.

**SONG Cheng** received his bachelor's degree from the School of Computer Science and Artificial Intelligence, Wuhan University of Technology, China in 2021. He majored in software engineering. He is working on data mining and machine learning for an MS degree at Huazhong University of Science and Technology, China.

**ZHOU Lun** received his MS Degree in Tongji Medical College, Huazhong University of Science and Technology, China in 2003. He is an associate professor of geriatrics at Tongji Hospital. His research interests include the pathogenesis of congenital heart disease and early warning of severe diseases in the elderly. He has published more than 20 papers in refereed journals such as PNAS and JMCC.