A Federated Learning Paradigm for Heart Sound Classification

Wanyong Qiu[1], Kun Qian[*1], Zhihua Wang[2], Yi Chang[3], Zhihao Bao[1], Bin Hu[*1], Björn W. Schuller[3,4], and Yoshiharu Yamamoto[5]

[1]Laboratory of Brain Health Engineering (BHE), School of Medical Technology, Beijing Institute of Technology, Beijing, China; [2]School of Mechatronic Engineering, China University of Mining and Technology, Xuzhou, China; [3]GLAM–Group on Language, Audio & Music, Imperial College London, UK; [4]Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Germany; [5]Educational Physiology Laboratory, Graduate School of Education, The University of Tokyo, Japan.

Correspondence to: {qian, bh}@bit.edu.cn

北京理工大學
BEIJING INSTITUTE OF TECHNOLOGY

Cardiovascular diseases (CVDs) have been ranked as the leading cause for deaths. The early diagnosis of CVDs is a crucial task in the medical practice. A plethora of efforts were given to the automated auscultation of heart sound, which leverages the power of computer audition to develop a cheap, non-invasive method that can be used at any time and anywhere for measuring the status of the heart. Nevertheless, previous works ignore an important factor, namely, the privacy of the user data. On the one hand, learnt models are always hungry for bigger data. On the other hand, it can be difficult to protect personal private information when collecting such large amount of data. In this dilemma, we propose a federated learning (FL) framework for the heart sound classification task. To the best of our knowledge, this is the first time to introduce FL to this field. We conducted multiple experiments, analysed the impact of data distribution across collaborative institutions on model quality and learning patterns, and verified the feasibility and effectiveness of FL based on real data. Non- independent identically distributed (Non- IID) data and model quality can be effectively improved by adding a strategy of globally sharing data.

Introduction

Healthcare data cannot be shared, considering patient privacy and related laws and regulations. As a result, data resides in various institutions forming data silos, such as hospitals, home devices, and smartphones. Federated learning (FL) retains the ownership of data among institutions and conducts collaborative modelling, which can effectively solve data silos and protect privacy. Healthcare data such as medical records and disease symptoms are highly sensitive and tightly managed, and the centralised collection of clinical data from isolated medical centres and hospitals is a challenge. FL can solve this problem. We apply FL to an island system in which each data island (e. g., a medical institution) communicates with a central server, but does not share anything between them.

Federated Learning for Heart Sound

Healthcare institutions face issues such as data decentralisation and privacy protection, which motivate us to explore the potential and value of FL in heart sound analysis.

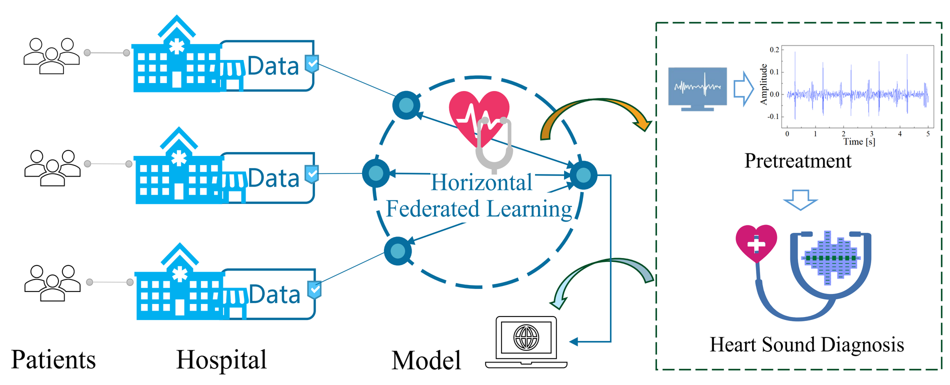


Fig. 1. A horizontal federated learning paradigm for heart sound analysis. As opposed to ECG (Electrocardiograph), which describes only the physiological fluctuations of the heart, PCG (Phonocardiogram) can distinguish between different pathological cases. The figure shown that several medical institutions collaborate in modelling through horizontal FL without sharing any private data. Due to the limitation of the database, this study focuses on the analysis of heart sound abnormality detection under FL, which contains three main parts: 1) The PCG signal was segmented into basic heart sound segments; 2) the one-dimensional waveform is transformed into a two-dimensional spectrogram using continuous wavelet transform (cwt); 3) the performance of FL is validated and evaluated on real heart sound data.

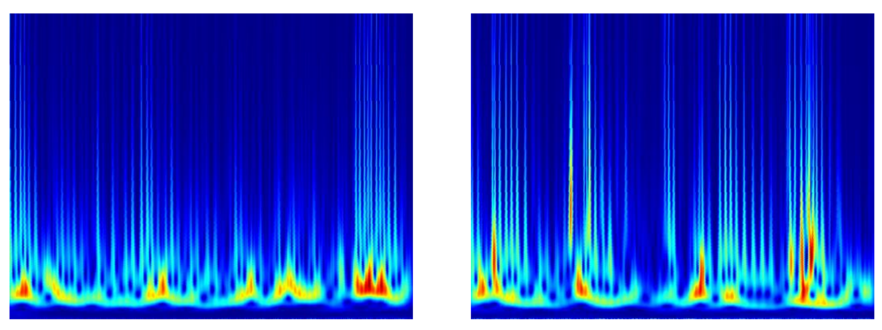
Materials and Methods

Database. As shown in Table 1, Non-IID heart sound data is mainly reflected in label distribution skew and quantity skew. The datasets (D_A-D_F) contains 3 240 records from 764 subjects/patients, all records are divided into two categories according to expert labels: Normal and Abnormal. Normal records are from healthy subjects, and abnormal records are from patients with typical heart valve defects and coronary heart disease.

Database	Recordings	Proportion of recordings(%)		Durations(s)
		Abnormal	Normal	Minimum
D_A (MIT)	409 (12.62)	28.61	71.39	9.27
D_B (AAD)	490 (15.12)	21.22	78.77	5.31
D_C (AUTH)	31 (0.95)	77.74	22.58	9.65
D_D (UHA)	55 (1.69)	50.91	49.09	6.61
D_E (DLUT)	2 141 (66.08)	5.51	91.45	8.06
D_F (SUA)	114 (3.52)	29.82	70.17	29.38
Total	3 240	20.52	79.47	

Table 1. Description of Physionet heart sound database. MIT: The Massachusetts Institute of Technology heart sounds database; AAD: Aalborg University heart sounds database; AUTH: The Aristotle University of Thessaloniki heart sounds database; UHA: The University of Haute Alsace heart sounds database; DLUT: The Dalian University of Technology heart sounds database; SUA: The Shiraz University adult heart sounds database.

Image Representations. Basic heart sounds (S1, systolic, S2, diastolic) were analysed according to PCG signals. Referring to previous work, instead of using all the information of a record for experimental analysis, we chose to process and analyse 5-second heart sound segments. The reasons are as follows: 1) The heart sound segment contains the complete basic heart sound; 2) abnormal heart sounds can be fully judged within 5 seconds; 3) overfitting can be reduced.



Abnormal (f0027.wav) Normal (f0072.wav)

Fig. 2. The scalogram images are extracted from the abnormal / normal heart sounds using cwt. According to previous work, wavelet analysis bears great potential to present the differences in the time-frequency characteristics of normal and abnormal heart sounds. In this work, we implement the continuous wavelet transform (cwt) sound-image transformation in Matlab. The sampling frequency is 2 kHz. The wavelet base function is $cga8$, the image generation function is $imagesc$, and the wavelet coefficient value of the heart sound is obtained by the cwt function. The axis and edge markers are removed by the $imwrite$ function. Fig. 2 shows the cwt spectrograms of an abnormal heart sound sample (f0027.wav) and a normal heart sound sample (f0072.wav) from the D_F database.

FL Algorithm Model. FL can be divided into two types: Cross-device and Cross-silo. Cross-device is mainly oriented towards IoT devices, (e. g., cell phones). In contrast, the horizontal FL in this paper consists of a server and several clients (e. g., medical institutions), which belong to the Cross-silo category. Therefore, it can be assumed that there is no communication problem, and only the differences in data distribution among institutions need to be addressed.

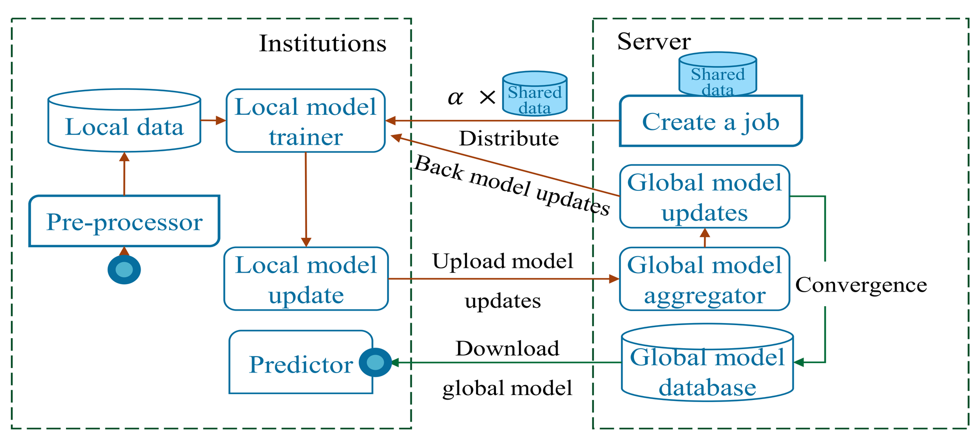


Fig. 3. Architecture diagram of the FedAvg algorithm for heart sounds. Based on the algorithm of FedAvg, we reproduce the FedAvg-MLP, FedAvg-CNN_1, and FedAvg-CNN_2 algorithm models for heart sound data.

Server-side Initialisation: The server-side distribution initialisation model uses global shared data and distributes randomised portions of the shared data.

Local Model Update: Local institutions combine private data and partial global shared data for model update.

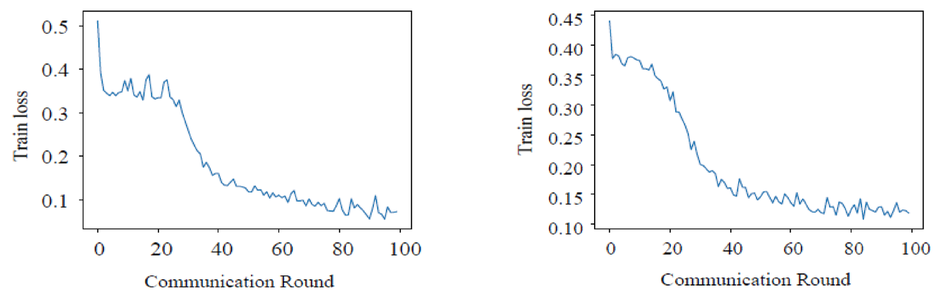
Global Model Aggregation: The server receives model updates from each institution and then updates the global model using a weighted average.

Iteration: The above two processes are iterated for the maximum round.

Experiments and Results

Experimental Design. The ultimate goal of FL is to obtain a robust global model on the server side, so validation and testing are performed on the server side. **Centralised Environment:** Considers the skewed label distribution and quantity for each of the Non-IID heart sound databases, and helps avoid a small sample size for the validation and test sets. We divide the training set as a merged set (D_C, D_D, D_E, D_F), the validation set as D_A, and the test set as D_B. Then, the MLP and two CNN models are trained separately. **Federated Environment:** To examine the difference in model performance between IID and Non-IID data, we set up a control group experiment (1-3).

Results. Federated vs data-centralised training: Due to the limitation of database sample volume, simple neural networks are used for FL. Although the accuracy of each model is poor under centralised training, the FL environment is effective for models to avoid overfitting and keep performance. Despite the large variation in the Non-IID heart sound database across institutions, FedAvg still achieved acceptable classification accuracy. This indicates that the unbalanced number of iterations across institutions can be efficiently handled by the FedAvg algorithm. In addition, although FedAvg is robust to unbalanced and Non-IID data, the accuracy of Experiment 2 is considerably lower than that of Experiment 1. We found as reason in the work that such precision reduction can be explained by weight divergence, which can be quantified by the earth movers distance (EMD) between the distribution over classes and the quantity distribution on each institution.



(a) Fed-CNN_1. (b) Fed-CNN_2.

Fig. 4. Training error variation for Fed-CNN_1 and Fed-CNN_2.

Model	Avg-loss. of IID	Acc. of IID	Acc. of Non-IID	
			Non-shared	Shared
MLP	0.0117	57.9	-	-
CNN_1	0.0114	68.1	-	-
CNN_2	0.1180	76.2	-	-
Fed-*MLP*	0.2753	52.6	45.3	47.6
Fed-CNN_1	0.1629	66.9	57.8	62.1
Fed-CNN_2	0.1215	72.1	**62.0**	**65.4**

Table 2. Test accuracy of FedAvg for IID data and Non-IID data under centralised training and federated learning (%).

Globally shared data: In experiment 3, we selected the parameters β = 10 % and α = 50 %. The experimental results show that FedAvg-CNN_2 can reach 65.4 % accuracy in testing, while the accuracy is only about 62.0 % without a data sharing strategy. In contrast, FedAvg-CNN_2 has a higher sensitivity (59.2 %) and specificity (65.9 %), and the model performance is better. This process is executed only once during server-side initialisation, hence, there is no communication overhead. In addition, instead of random weights during the FedAvg initialisation, the global sharing strategy reduces the EMD of the institution and therefore improves the test accuracy, while D is a separate database with no privacy implications.

Conclusion and Future Work

We applied FL to real-world healthcare data and evaluated the first FL system for a multi-institutional heart sound database. We validated the potential of FL in terms of performance and data protection for a real heart sound database, which is essential for processing sensitive healthcare data. Note that privacy is not a default guarantee in any FL setting, and secure aggregation of communication parameters between servers and institutions is our work plan. Importantly, the limited number of participating institutions and the lack of data volumes are currently the biggest contributing factors. We are working on the establishment of a more powerful heart sound database based on the cooperation of several medical institutions, such as the HSS-The Heart Sounds Shenzhen Corpus which is part of our current work.

Acknowledgment

This work was partially supported by the Ministry of Science and Technology of the People's Republic of China (2021ZD0201900), the BIT Teli Young Fellow Program from the Beijing Institute of Technology, China, and the Grants-in-Aid for Scientific Research (No. 20H00569) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan.

Collaborators

Imperial College London
東京大学 THE UNIVERSITY OF TOKYO
Universität Augsburg University
中国矿业大学 CHINA UNIVERSITY OF MINING AND TECHNOLOGY