

A Federated Learning Paradigm for Heart Sound Classification

Wanyong Qiu¹, Kun Qian^{*1}, *Senior Member, IEEE*, Zhihua Wang², Yi Chang³, Zhihao Bao¹, Bin Hu^{*1}, *Senior Member, IEEE*, Björn W. Schuller³, *Fellow, IEEE*, and Yoshiharu Yamamoto⁴, *Member, IEEE*

Abstract—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.

I. INTRODUCTION

In the U.S. alone, more than 30 % (equal to \$9 billion) of preventable hospitalisations are due to cardiovascular diseases (CVDs) [1]. AI auscultation-assisted diagnosis (non-invasive) needs to be developed to provide fast and effective pre-screen diagnosis and abnormal warning for patients. Existing healthcare data sits in data silos and machine learning (ML) cannot fully function. To address this issue, federated learning (FL) is expected to realise its potential and transition from research to clinical practice.

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. *Corresponding authors:* Kun Qian and Bin Hu.

¹Wanyong Qiu, Kun Qian, Bin Hu and Zhihao Bao are with the Laboratory of Brain Health Engineering (BHE), School of Medical Technology, Beijing Institute of Technology, No.5 Zhongguancun South Street, Haidian District, Beijing 100081, China. {qiuwy, qian, bh, bzh}@bit.edu.cn

²Zhihua Wang is with the School of Mechatronic Engineering, China University of Mining and Technology, Xuzhou, Jiangsu, China. w.z_hua@cumt.edu.cn

³Björn W. Schuller and Yi Chang are with GLAM – the Group on Language, Audio, & Music, Imperial College London, 180 Queen's Gate, London SW7 2AZ, UK. And also Björn W. Schuller is with the Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Eichleitnerstr.30, 86159 Augsburg, Germany. schuller@ieee.org, y.chang20@imperial.ac.uk

⁴Y. Yamamoto is with the Educational Physiology Laboratory, Graduate School of Education, The University of Tokyo, Japan. yamamoto@p.u-tokyo.ac.jp

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. 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 [2]–[4]. 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.

Healthcare data is usually typical non-independent identically distributed (Non-IID) data and distributed in various institutions collected by heterogeneous devices, which is in line with the FL scenario. Furthermore, in a large number of clinical studies [5]–[10], the focus is not on prediction, but on correlation analysis and hypothesis testing to understand the associations between different factors. The advantages of FL have recently been demonstrated in several practical applications, including medical image diagnosis [7], [9], prediction of disease risk factors [1], [10], and personalised medicine [8], which are particularly important in medical data applications. Therefore, the application of FL in healthcare is an urgent research area. FL-based assisted diagnostic classification tasks allow data maintainers to collaborate on modelling without sharing any private data. Our study builds on the work of *Federated Averaging (FedAvg)* [2] using real heart sound data to build FedAvg-MLP (multi-layer perceptron) and FedAvg-CNN (convolutional neural network) models.

The main contributions of this work can be summarised as follows:

1) The feasibility of a multi-institutional heart sound database for FL is implemented and evaluated; 2) aspects of the algorithmic models under centralised training and FL are compared; 3) the effectiveness of FL with Non-IID heart sound data is verified.

II. FL FOR HEART SOUND

In this section, we introduce the architecture of FL applied to heart sound analysis, and then describe the database and data preparation. Finally, the FL model and evaluation metrics are presented.

A. Architecture

Medical institutions are currently facing two important clinical realities: On the one hand, studies have shown that patients with CVDs, especially those with heart failure, are chronically accompanied by mental health such as anxiety and depression. On the other hand, issues such as data decentralisation and privacy protection urgently need to be resolved. These problems motivate us to explore the potential and value of FL in 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 [11]. As shown in Fig. 1, 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.

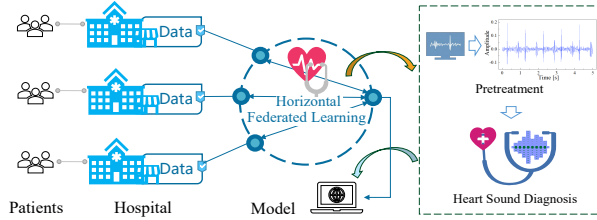


Fig. 1. A horizontal federated learning paradigm for heart sound analysis.

B. Materials and Methods

1) **Database:** The heart sound data from the PhysioNet/CinC Challenge 2016 [12]. The data consists of six independent databases that are non-independent identically distributed (Non-IID).

Data Heterogeneity: The data consisted of six independent heart sound databases (D_A - D_F), collected by different research teams in seven countries [12]. And these records were collected using heterogeneous equipment both clinically and non-clinically (e. g, home visits). The length of the records ranges from a few seconds to a few minutes.

System Heterogeneity: System heterogeneity in FL environment is mainly manifested in the difference of communication and computing capability of the devices at each institution.

As shown in Table I, 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.

TABLE I
DESCRIPTION OF PHYSIONET HEART SOUND DATABASE.

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	

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.

2) **Image Representations:** Basic heart sounds (S1, systolic, S2, diastolic) were analysed according to PCG signals. Referring to the work of [13], 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.

According to previous work [14], 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.

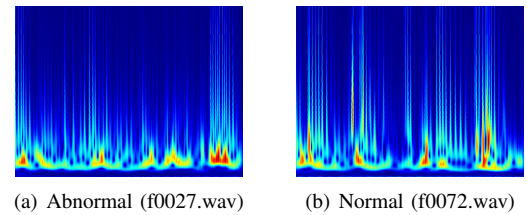


Fig. 2. The scalogram images are extracted from the abnormal / normal heart sounds using cwt.

3) **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.

Based on the algorithm of FedAvg [2], we adopt the work of [15] to reproduce the FedAvg-MLP, FedAvg- CNN_1 , and FedAvg- CNN_2 algorithm models for heart sound data. The MLP contains one hidden layer and one dropout optimisation term. CNN_1 performs a dropout term after two convolutional layers, and finally adds two fully connected layers. CNN_2 adds a pooling layer between two convolutional layers, and finally adds three fully connected layers. The specific process of FedAvg in this work is shown in Fig. 3.

Fig. 3. Architecture diagram of the FedAvg algorithm for heart sounds.

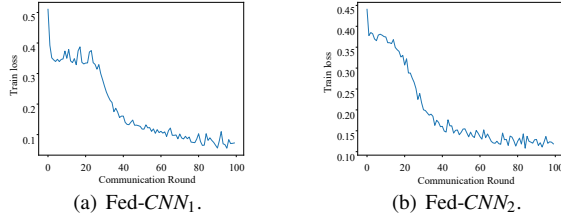


Fig. 4. Training error variation for Fed-CNN₁ and Fed-CNN₂.

TABLE II

TEST ACCURACY OF FEDAVG FOR IID DATA AND NON-IID DATA UNDER CENTRALISED TRAINING AND FEDERATED LEARNING (%).

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

Globally shared data

In experiment 3, we followed the scheme of [16] and selected the parameters $\beta = 10\%$ and $\alpha = 50\%$. The experimental results show that FedAvg-CNN₂ can reach 65.4% accuracy in testing, while the accuracy is only about 62.0% without a data sharing strategy. In contrast, FedAvg-CNN₂ has a higher sensitivity (59.2%) and specificity (65.9%), and the model performance is better as shown in Table III. 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.

III. DISCUSSION

FL has obvious privacy advantages over centralised training. Global computing involves the participation of multiple institutions, the ownership of data is retained in each institution and local computing is performed, and only updated model parameters are allowed to be shared with the server. The principle of data minimisation is embodied in the model update part, and the updates provided by each institution only need to be temporarily saved by the receiving server until the aggregation can be performed. In addition, it is crucial to optimise the hyperparameters of FedAvg, – especially the number of local epochs plays an important role in the convergence of the model.

IV. 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,

TABLE III

SENSITIVITY (Se), SPECIFICITY (Sp), AND FINAL SCORE ($Score$) ARE USED AS QUANTITATIVE METRICS FOR THE MODEL.

Model	Se	Sp	Score
Fed-MLP	47.1	56.2	51.6
Fed-CNN ₁	55.9	62.2	59.1
Fed-CNN ₂	59.2	65.9	62.5

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 [17] which is part of our current work.

REFERENCES

- [1] T. S. Brisimi *et al.*, “Federated learning of predictive models from federated electronic health records,” *International journal of medical informatics*, vol. 112, pp. 59–67, 2018.
- [2] B. McMahan *et al.*, “Communication-efficient learning of deep networks from decentralized data,” in *Artificial intelligence and statistics*. PMLR, 2017, pp. 1273–1282.
- [3] S. Lu, Y. Zhang, and Y. Wang, “Decentralized federated learning for electronic health records,” in *2020 54th Annual Conference on Information Sciences and Systems (CISS)*. IEEE, 2020, pp. 1–5.
- [4] Q. Yang *et al.*, “Federated machine learning: Concept and applications,” *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 10, no. 2, pp. 1–19, 2019.
- [5] J. B. Hernandez *et al.*, “Privacy-first health research with federated learning,” *medRxiv*, 2020.
- [6] M. J. Sheller *et al.*, “Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data,” *Scientific reports*, vol. 10, no. 1, pp. 1–12, 2020.
- [7] W. Li *et al.*, “Privacy-preserving federated brain tumour segmentation,” in *International workshop on machine learning in medical imaging*. Springer, 2019, pp. 133–141.
- [8] M. Grama *et al.*, “Robust aggregation for adaptive privacy preserving federated learning in healthcare,” *arXiv preprint arXiv:2009.08294*, 2020.
- [9] S. Silva *et al.*, “Federated learning in distributed medical databases: Meta-analysis of large-scale subcortical brain data,” in *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*. IEEE, 2019, pp. 270–274.
- [10] A. Vaid *et al.*, “Federated learning of electronic health records improves mortality prediction in patients hospitalized with covid-19,” *medRxiv*, 2020.
- [11] A. Raza *et al.*, “Designing ecg monitoring healthcare system with federated transfer learning and explainable ai,” *Knowledge-Based Systems*, vol. 236, p. 107763, 2022.
- [12] G. D. Clifford *et al.*, “Classification of normal/abnormal heart sound recordings: The physionet/computing in cardiology challenge 2016,” in *2016 Computing in cardiology conference (CinC)*. IEEE, 2016, pp. 609–612.
- [13] T. Nilanon *et al.*, “Normal/abnormal heart sound recordings classification using convolutional neural network,” in *2016 computing in cardiology conference (CinC)*. IEEE, 2016, pp. 585–588.
- [14] P. Langley and A. Murray, “Abnormal heart sounds detected from short duration unsegmented phonocardiograms by wavelet entropy,” in *2016 Computing in Cardiology Conference (CinC)*. IEEE, 2016, pp. 545–548.
- [15] S. Ji, *A PyTorch Implementation of Federated Learning*. Zenodo, Mar. 2018. [Online]. Available: <https://doi.org/10.5281/zenodo.4321561>
- [16] Y. Zhao *et al.*, “Federated learning with non-iid data,” *arXiv preprint arXiv:1806.00582*, 2018.
- [17] F. Dong *et al.*, “Machine listening for heart status monitoring: Introducing and benchmarking HSS—The Heart Sounds Shenzhen Corpus,” *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 7, pp. 2082–2092, 2020.