



Feasibility Study of Federated Learning on the Distributed Research Network of OMOP Common Data Model

Geun Hyeong Lee^{1,2}, Jonggul Park³, Jihyeong Kim³, Yeesuk Kim⁴, Byungjin Choi⁵, Rae Woong Park^{5,6}, Sang Youl Rhee^{7,8}, Soo-Yong Shin^{1,2,9}

¹Department of Digital Health, SAIHST, Sungkyunkwan University, Seoul, Korea

²Department of Intelligent Precision Healthcare Convergence, Sungkyunkwan University, Suwon, Korea

³Evidnet Inc., Seongnam, Korea

⁴Department of Orthopedic Surgery, Hanyang University College of Medicine, Seoul, Korea

⁵Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Korea

⁶Department of Biomedical Sciences, Graduate School of Medicine, Ajou University, Suwon, Korea

⁷Department of Endocrinology and Metabolism, College of Medicine, Kyung Hee University, Seoul, Korea

⁸Center for Digital Health, Medical Science Research Institute, Kyung Hee University Medical Center, Seoul, Korea

⁹Center for Research Resource Standardization, Samsung Medical Center, Seoul, Korea

Objectives: Since protecting patients' privacy is a major concern in clinical research, there has been a growing need for privacy-preserving data analysis platforms. For this purpose, a federated learning (FL) method based on the Observational Medical Outcomes Partnership (OMOP) common data model (CDM) was implemented, and its feasibility was demonstrated.

Methods: We implemented an FL platform on FeederNet, which is a distributed clinical data analysis platform based on the OMOP CDM in Korea. We trained it through an artificial neural network (ANN) using data from patients who received steroid prescriptions or injections, with the aim of predicting the occurrence of side effects depending on the prescribed dose. The ANN was trained using the FL platform with the OMOP CDMs of Kyung Hee University Medical Center (KHMC) and Ajou University Hospital (AUH). **Results:** The area under the receiver operating characteristic curves (AUROCs) for predicting bone fracture, osteonecrosis, and osteoporosis using only data from each hospital were 0.8426, 0.6920, and 0.7727 for KHMC and 0.7891, 0.7049, and 0.7544 for AUH, respectively. In contrast, when using FL, the corresponding AUROCs were 0.8260, 0.7001, and 0.7928 for KHMC and 0.7912, 0.8076, and 0.7441 for AUH, respectively. In particular, FL led to a 14% improvement in performance for osteonecrosis at AUH. **Conclusions:** FL can be performed with the OMOP CDM, and FL often shows better performance than using only a single institution's data. Therefore, research using OMOP CDM has been expanded from statistical analysis to machine learning so that researchers can conduct more diverse research.

Keywords: Machine Learning, Deep Learning, Distributed Systems, Privacy, Common Data Element

Submitted: September 2, 2022, **Revised:** January 21, 2023, **Accepted:** February 8, 2023

Corresponding Author

Soo-Yong Shin

Department of Digital Health, SAIHST, Sungkyunkwan University, 81, Irwon-ro, Gangnam-gu, Seoul 06351, Korea. Tel: +82-2-3410-1449, E-mail: sy.shin@skku.edu (<https://orcid.org/0000-0002-2410-6120>)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

I. Introduction

Federated learning (FL) is a new machine learning method that can be performed when data are distributed and difficult to centralize [1]. Initially, it was proposed by Google for use in mobile devices, but it is emerging as a suitable learning method in the medical area because it can achieve the effects of large-scale data learning without sharing the original data from multiple institutions [1-3]. FL performs learning using only the data held, without sharing the original data, and only shares the model weights to update the model. This method has two advantages: it can (1) reduce the risk of data leakage and (2) achieve privacy protection of data [4,5]. In particular, FL is very suitable for adoption in the medical domain because medical data are difficult to share with other institutions due to personal privacy protection reasons [6-9]. Since FL allows models to be updated by exchanging only the weights, it enables multi-institution research using medical data. As a consequence, FL can achieve higher performance than research conducted by each institution individually [6-9].

Before FL, it was common to share statistical analysis models through a distributed research network and perform meta-analyses of the values reported from each institution [10-12]. A representative distributed research network is the Observational Health Data Sciences and Informatics (OHD-SI) [13]. In South Korea, the FeederNet was built based on

the Observational Medical Outcomes Partnership (OMOP) common data model (CDM) used in the OHDSI. FeederNet is a distributed medical data analysis platform currently involving 33 institutions [14]. FL is performed in a data-distributed environment, so FeederNet is a very suitable environment for FL. The FL system requires a client module that performs machine learning at each institution, and a server module that aggregates model parameters and communicates with clients. Building this system on FeederNet enables high-quality research based on machine learning methods. To verify the feasibility of this FL platform, we conducted a pilot study using data from patients who received steroid prescriptions or injections, with the aim of predicting side effects such as bone fractures, osteonecrosis, and osteoporosis that could occur depending on the prescribed dose [15,16].

II. Methods

1. Implementation of the FL Platform

In order to efficiently perform FL on FeederNet, a distributed medical data analysis platform in Korea, the FL system was designed with a server-centered pulling method, unlike the client-centered FL method. Assuming that the CDM of each institution was the client, client control was only possible through FeederNet. Therefore, we designed a server-centered structure that could directly control FeederNet and

Table 1. List of federated learning APIs

API	Method	API content	API detail
/login	POST	Login	Check a user authentication to perform federated learning
/researchFileForML	POST	Research file upload	Transmit the python script file for machine learning to each institution
/project/{PROJECT_ID}/analysis/{ANALYSIS_ID}/execution?cdmId={CDM_ID}	POST	Research execution	Run Python script file Access to CDM to inquire data, perform pre-processing, and perform learning using the defined model.
/execution/{EXECUTION_ID}	GET	Research status check	Query the execution status of a Python script Has the following execution status: Waiting, Preparing, Running, Saving, Finished, Failed
/project/{PROJECT_ID}/analysis/{ANALYSIS_ID}/execution/{EXECUTION_ID}/resultFile	GET	Research result file list and ID lookup	Lookup the file list and ID created as a result of running the Python script The list of created files is as follows: log.txt, local_weight.json, local_socres.csv
/project/{PROJECT_ID}/analysis/{ANALYSIS_ID}/execution/{EXECUTION_ID}/resultFile/{FILE_ID}/export	GET	Research result file download	Download the file created as a result of running the Python script

API: application programming interface, CDM: common data model.

aggregate functions from the outside. For this, application programming interfaces (APIs) were defined and implemented. The details of each API are shown in Table 1. The APIs include login, research file upload, research execution, status check, result inquiry, and downloading results.

The developed platform can be used after logging into FeederNet. The server manages scripts that perform pre-defined pre-processing and learning processes, and the server transmits this script to FeederNet. FeederNet runs the transmitted script in the machine learning environment of each institutional client and updates the progress. The client performs training and testing through script execution and saves the result in an independent space, which FeederNet manages. The server can then download the client's weights and the results of the script after it runs, stored in an independent space from FeederNet. The server updates the global weights by performing federated averaging, which calculates average values by aggregating weight values from the results downloaded from FeederNet. Next, the server repeats this process by transmitting the script, including the global weights, back to FeederNet to proceed with the next round of learning. This process is shown in a sequential diagram in Figure 1.

This approach has the advantage that the server manages only the number of rounds and weight information, and each client is able to perform training, testing, and saving the results on the platform stably.

2. Steroid Side Effects Study

To study steroid side effects, the CDMs of Kyung Hee University Medical Center (KHMC) and Ajou University Hospital (AUH) linked to FeederNet were used. The CDM versions KHMC_5.3.1_0.2 and AUH_5.3.1_0.6 were used. The subjects of this study were patients over 20 years old who had been prescribed oral or injected steroids from January 1, 2001 to December 31, 2019. We used SNOMED-CT codes for each disease and RxNorm codes for specific steroid drugs to retrieve the data. We excluded patients who had no records of hospital visits within 90 days after the prescription date, patients who had no records of hospital visits within 365 days before the prescription date, and patients for whom data errors make recognition impossible. The collected items are shown in Table 2.

Learning was conducted using daily average dose, changes in vital signs, total dose, and duration of dose, which were calculated by extracting data from the CDM. The duration of the dose was calculated based on the start and the end dates of the steroid prescription for each patient. The total dose

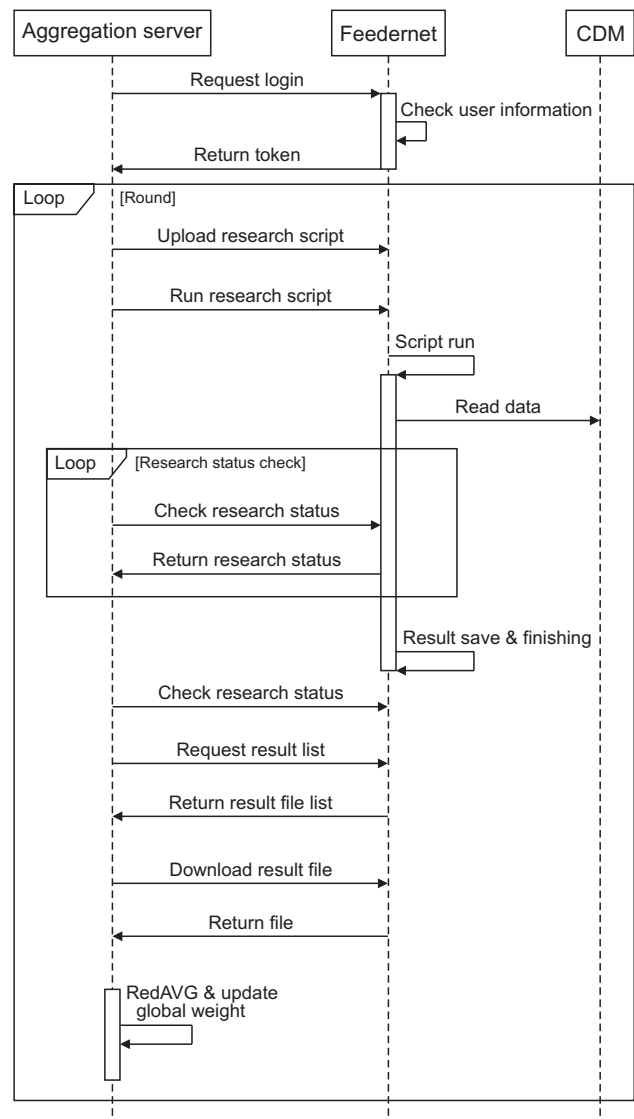


Figure 1. Federated learning sequence diagram.

was calculated as the cumulative dose during this period, and the average daily dose was calculated over the same period. We also divided the duration of steroid use into short, intermediate, and long-acting intervals using 90-day intervals. Moreover, changes in vital signs were checked when the steroid was injected. The predicted outcomes were bone fracture, osteonecrosis, and osteoporosis. For each disease, positive patients were labeled “true” and negative patients were labeled “false.”

3. Machine Learning Model

Using the pre-processed data, each client trained the artificial neural network. Seventy percent of the data from each institution was used for training, and 30% was used for testing. For performance evaluation, the area under the receiver

Table 2. Collected items

Category	Description
Collection item	Gender, age, diagnosis Drugs taken, duration of drugs, dosage, usage Visit date, period and number of visits Treatment details, period, method
Classification of steroid use groups	According to steroid potency, it is classified into short, intermediate, and long-acting, and each drug is applied in clinical practice. A cohort will be established according to each drug and group.
Classification of steroid usage	Difference in potency of each drug, the potency is calculated based on the prednisolone, which is a commonly used drug, and the cumulative dosage and daily average dose of the drug are calculated. Investigate the number of cohorts according to the calculated results and construct cohorts by dividing groups into 1:9 or 2:8 combinations.

Table 3. Number of patients queried in each CDM

CDM	Diagnosis	Outcome ^{a)}	
		False	True
KHMC (n = 11,058)	Bone fracture	10,599	459 (4.15)
	Osteonecrosis	10,993	65 (0.59)
	Osteoporosis	9,936	1,122 (10.15)
AUH (n = 28,596)	Bone fracture	28,312	284 (0.99)
	Osteonecrosis	28,562	34 (0.12)
	Osteoporosis	27,819	777 (2.72)

Values are presented as number of patients (%).

CDM: common data model, KHMC: Kyung Hee University Medical Center, AUH: Ajou University Hospital.

^{a)}For each disease, positive patients were labeled “true” and negative patients were labeled “false.”

operating characteristic curve (AUROC) was calculated in the verification phase.

The model used for training consisted of a total of three layers, including an input layer, a fully connected layer, and an output layer, which applied the sigmoid activation function. Training was performed in 100 iterations.

For FL, federated averaging was performed using “coef_” which denotes the coefficient in the decision function, and “intercept_” which refers to the intercept value in the decision function, among the hyperparameters generated as a result of learning. Nineteen rounds of learning were conducted.

III. Results

The total number of patients for whom data were collected was 11,058 at KHMC and 28,596 at AUH. The numbers of

Table 4. Learning results, expressed as AUROCs, using each institution's data

CDM	Diagnosis	AUROC
KHMC	Bone fracture	0.8426
	Osteonecrosis	0.692
	Osteoporosis	0.7727
AUH	Bone fracture	0.7891
	Osteonecrosis	0.7049
	Osteoporosis	0.7544

AUROC: area under the receiver operating characteristic curve, CDM: common data model, KHMC: Kyung Hee University Medical Center, AUH: Ajou University Hospital.

Table 5. Learning results expressed as AUROCs using federated learning

CDM	Diagnosis	AUROC
KHMC	Bone fracture	0.8260 (-1.9%)
	Osteonecrosis	0.7001 (+1.16%)
	Osteoporosis	0.7978 (+0.27%)
AUH	Bone fracture	0.7912 (+2.7%)
	Osteonecrosis	0.8076 (+14%)
	Osteoporosis	0.7441 (-1.3%)

AUROC: area under the receiver operating characteristic curve, CDM: common data model, KHMC: Kyung Hee University Medical Center, AUH: Ajou University Hospital.

cases of bone fracture, osteonecrosis, and osteoporosis were 459, 65, and 1,122 at KHMC and 284, 34, and 777 at AUH, respectively. Detailed information on the numbers of patients is shown in Table 3.

As a result of learning using only data from each institu-

tion, for bone fracture, osteoporosis, and osteoporosis, the AUROCs were 0.8426, 0.6920, and 0.7727, respectively, for KHMC, and 0.7891, 0.7049, and 0.7544 for AUH, respectively. Details are shown in Table 4.

The results of 19 rounds of FL for bone fracture, osteonecrosis, and osteoporosis are shown in Table 5. For KHMC, the AUROCs were 0.8260, 0.7001, and 0.7978, showing changes of -1.9% , $+1.16\%$, and $+0.27\%$, respectively, while the AUROCs for AUH were 0.7912, 0.8076, and 0.7441, showing changes of $+2.7\%$, $+14\%$, and -1.3% respectively.

IV. Discussion

The results of this study showed that FL improved the overall performance of disease prediction compared to using only data from each institution. For KHMC, performance was improved by 1.16% and 0.27% for osteonecrosis and osteoporosis, respectively. For AUH, performance was improved by 2.7% and 14% for bone fracture and osteonecrosis, respectively. In particular, the performance for osteonecrosis significantly improved by 1.16% at KHMC and 14% at AUH. Considering the small number of positive data points for osteonecrosis, it can be seen that the performance was significantly improved by incorporating data from both institutions in the learning process. In some cases, the performance was low; an explanation for this may be that the model we used had only one fully connected layer, so the convergence was insufficient for a relatively large number positive data points. To solve this, conducting more rounds of iteration or training a complex model using more layers would seem to be required.

Since FL through FeederNet is performed in an independent virtual environment, the risk of personal information leakage is quite low. In addition, the server collects only numeric weights; therefore, it is impossible to guess the original data. These aspects of FL make it possible to conduct multi-institution research using medical data dealing with sensitive information and have the major advantage of protecting personal information.

We demonstrated that the FL platform designed through experiments worked well in a distributed research environment. In particular, in the past, statistical analysis was frequently performed using OMOP-CDM, but it has been demonstrated that FL enables artificial intelligence learning using multi-institution OMOP-CDM. Based on this study, we think that this method will provide an opportunity for more active multi-institutional research using medical data through FL in the future.

Conflict of Interest

This study was conducted with support by Evidnet. Jonggul Park and Jihyeong Kim supported the use of FeederNet. And Rae Woong Park is an editorial member of Healthcare Informatics Research; however, he did not involve in the peer reviewer selection, evaluation, and decision process of this article.

Acknowledgments

This study was supported by the research funding from Evidnet Inc.

ORCID

Geun Hyeong Lee (<https://orcid.org/0000-0002-4365-8510>)
 Jonggul Park (<https://orcid.org/0000-0002-5796-4826>)
 Jihyeong Kim (<https://orcid.org/0000-0002-3662-3184>)
 Yeesuk Kim (<https://orcid.org/0000-0003-4956-0693>)
 Byungjin Choi (<https://orcid.org/0000-0002-1445-5888>)
 Rae Woong Park (<https://orcid.org/0000-0003-4989-3287>)
 Sang Youl Rhee (<https://orcid.org/0000-0003-0119-5818>)
 Soo-Yong Shin (<https://orcid.org/0000-0002-2410-6120>)

References

1. McMahan B, Moore E, Ramage D, Hampson S, y Arcas BA. Communication-efficient learning of deep networks from decentralized data. *Proc Mach Learn Res* 2017;54: 1273-82.
2. Bonawitz K, Eichner H, Grieskamp W, Huba D, Ingerman A, Ivanov V, et al. Towards federated learning at scale: System design. *Proceedings of Machine Learning and Systems (MLSys)*; 2019 Mar 31–Apr 2; Stanford, CA. p. 374-88.
3. Konecny J, McMahan HB, Yu FX, Riechtrik P, Suresh AT, Bacon D. Federated learning: strategies for improving communication efficiency [Internet]. Ithaca (NY): arXiv.org; 2016 [cited at 2023 Mar 20]. Available from: <https://doi.org/10.48550/arXiv.1610.05492>.
4. Xu J, Glicksberg BS, Su C, Walker P, Bian J, Wang F. Federated learning for healthcare informatics. *J Health Inform Res* 2021;5(1):1-19. <https://doi.org/10.1007/s41666-020-00082-4>
5. Rieke N, Hancox J, Li W, Milletari F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. *NPJ Digit Med* 2020;3:119. <https://doi.org/10.1038/s41746-020-0103-8>

- org/10.1038/s41746-020-00323-1
6. Lee GH, Shin SY. Federated learning on clinical benchmark data: performance assessment. *J Med Internet Res* 2020;22(10):e20891. <https://doi.org/10.2196/20891>
 7. Vaid A, Jaladanki SK, Xu J, Teng S, Kumar A, Lee S, et al. Federated learning of electronic health records to improve mortality prediction in hospitalized patients with COVID-19: machine learning approach. *JMIR Med Inform* 2021;9(1):e24207. <https://doi.org/10.2196/24207>
 8. Li W, Milletari F, Xu D, Rieke N, Hancox J, Zhu W, et al. Privacy-preserving federated brain tumour segmentation. In: Suk HI, Liu M, Yan P, Lian C, editors. *Machine learning in medical imaging*. Cham, Switzerland: Springer; 2019. p. 133-41. https://doi.org/10.1007/978-3-030-32692-0_16
 9. Sheller MJ, Edwards B, Reina GA, Martin J, Pati S, Kotrotsou A, et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Sci Rep* 2020;10(1):12598. <https://doi.org/10.1038/s41598-020-69250-1>
 10. Brown JS, Holmes JH, Shah K, Hall K, Lazarus R, Platt R. Distributed health data networks: a practical and preferred approach to multi-institutional evaluations of comparative effectiveness, safety, and quality of care. *Med Care* 2010;48(6 Suppl):S45-51. <https://doi.org/10.1097/MLR.0b013e3181d9919f>
 11. Hansen RA, Zeng P, Ryan P, Gao J, Sonawane K, Teeter B, et al. Exploration of heterogeneity in distributed research network drug safety analyses. *Res Synth Methods* 2014;5(4):352-70. <https://doi.org/10.1002/jrsm.1121>
 12. Toh S, Gagne JJ, Rassen JA, Fireman BH, Kulldorff M, Brown JS. Confounding adjustment in comparative effectiveness research conducted within distributed research networks. *Med Care* 2013;51(8 Suppl 3):S4-10. <https://doi.org/10.1097/MLR.0b013e31829b1bb1>
 13. Observational Health Data Sciences and Informatics (OHDSI) [Internet]. [place unknown]: OHDSI; 2022 [cited at 2023 Mar 20]. Available from: <https://www.ohdsi.org/>.
 14. FeederNet: a distributed clinical data analysis platform in Korea [Internet]. Seongnam, Korea: Evidnet; c2022 [cited at 2023 Mar 20]. Available from: <https://feeder.net.com/member/main>.
 15. Kim Y, Tian Y, Yang J, Huser V, Jin P, Lambert CG, et al. Comparative safety and effectiveness of alendronate versus raloxifene in women with osteoporosis. *Sci Rep* 2020;10(1):11115. <https://doi.org/10.1038/s41598-020-68037-8>
 16. Suchard MA, Schuemie MJ, Krumholz HM, You SC, Chen R, Pratt N, et al. Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis. *Lancet* 2019;394(10211):1816-26. [https://doi.org/10.1016/S0140-6736\(19\)32317-7](https://doi.org/10.1016/S0140-6736(19)32317-7)