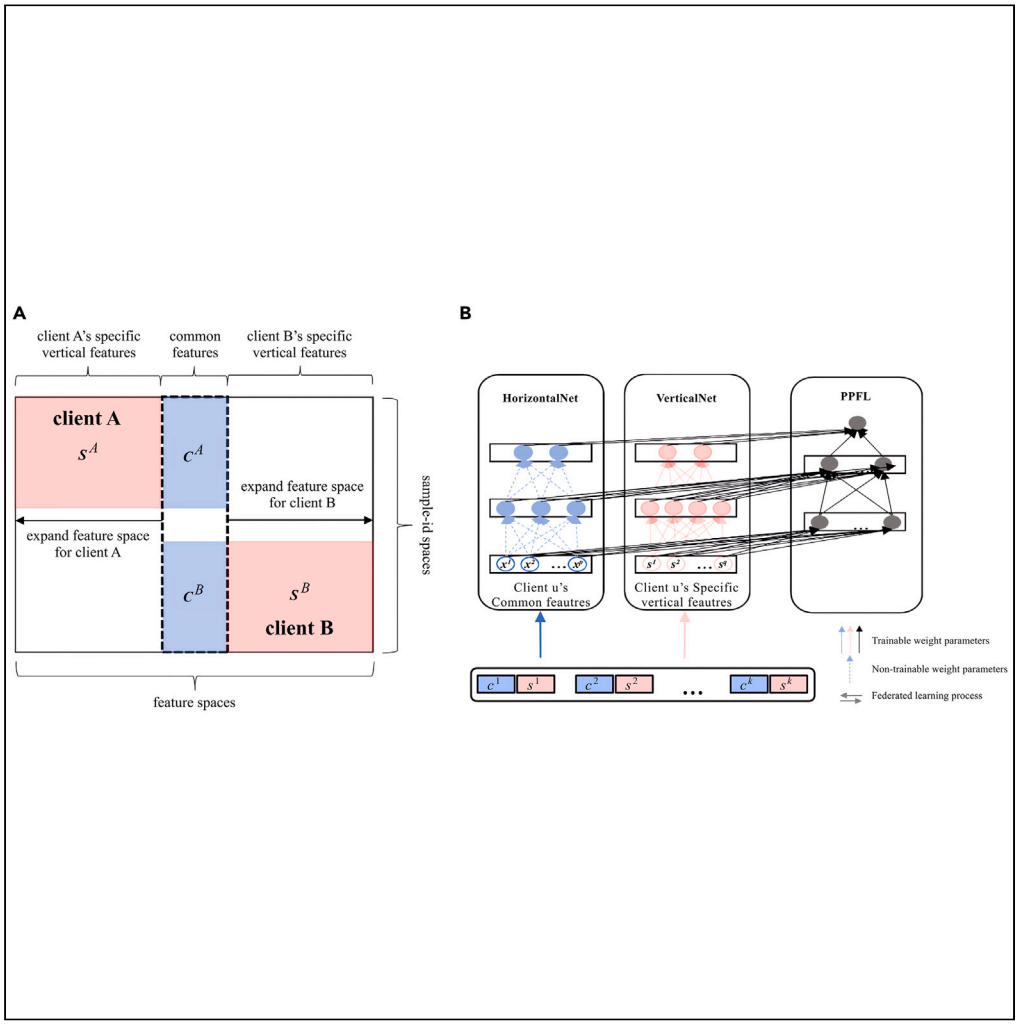
iScience

CellPress
OPEN ACCESS

Article

PPFL: A personalized progressive federated learning method for leveraging different healthcare institution-specific features



Tae Hyun Kim, Jae Yong Yu, Won Seok Jang, ..., JaeSeong Hong, KyungSoo Chung, Yu Rang Park

yurangpark@yuhs.ac

Highlights

We developed a privacy-preserving progressive personalized federated learning model

The proposed model learns client-specific features by transmitting layer-wise knowledge

PPFL outperformed the conventional models in predicting mortality on real-world datasets

Kim et al., iScience 27, 110943
October 18, 2024
https://doi.org/10.1016/j.isci.2024.110943

iScience

CellPress
OPEN ACCESS

Article

PPFL: A personalized progressive federated learning method for leveraging different healthcare institution-specific features

Tae Hyun Kim,[1,4] Jae Yong Yu,[2,4] Won Seok Jang,[1] Sun Cheol Heo,[1] MinDong Sung,[1,3] JaeSeong Hong,[1] KyungSoo Chung,[3] and Yu Rang Park[1,4,5,*]

SUMMARY

Federated learning (FL) in healthcare allows the collaborative training of models on distributed data sources, while ensuring privacy and leveraging collective knowledge. However, as each institution collects data separately, conventional FL cannot leverage the different features depending on the institution. We proposed a personalized progressive FL (PPFL) approach that leverages client-specific features and evaluated with real-world datasets. We compared the performance of in-hospital mortality prediction between our model and conventional models based on accuracy and area under the receiver operating characteristic (AUROC). PPFL achieved an accuracy of 0.941 and AUROC of 0.948, which were higher than the scores of the local models and FedAvg algorithm. We also observed that PPFL achieved a similar performance for cancer data. We identified client-specific features that can contribute to mortality. PPFL is a personalized federated algorithm for heterogeneously distributed clients that expands the feature space for client-specific vertical feature information.

INTRODUCTION

Federated learning (FL) is a collaborative machine learning approach used for solving data problems, such as data leakage, while leveraging the collective knowledge and preserving privacy in distributed environments across multiple devices and institutions in a communication-efficient manner.[1–4] Consequently, researchers are increasingly focusing on comprehensive research on FL for healthcare. Depending on the participating client settings in FL, FL is classified into two types: cross-silo and cross-device categories.[5] When considering the cross-silo aspect of FL, several studies based on various medical data collected from healthcare organizations have been presented. Dayan et al. proposed an electronic medical record (EMR) chest X-ray AI model, called EXAM, to conduct FL using data from multiple sites spread across four continents and under the oversight of different regulatory bodies, thus achieving a generalized model that can be applied to different regulated markets in an expedited way.[6] A study published in 2020[7] utilized FL for predicting mortality in patients hospitalized with coronavirus 2019 (COVID-19) based on data obtained from five clients, which is the study to evaluate the efficacy of applying FL to predict mortality in patients with COVID-19. Another study conducted in 2021 used cloud-based FL with two cloud tenants to identify different FL models that have achieved statistically significant performances.[8] In addition to electronic health record (EHR)-based FL, FL research focusing on medical image data, including functional magnetic resonance imaging, chest computed tomography, and breast density classification data, has also been suggested.[9] Most cross-silo FL studies in the healthcare domain are horizontal FL (HFL) studies,[2,10–15] which assume that clients have the same dataset, in which case client-specific features are ignored.

Recently, the development of cross-device FL has become prominent, and it has been adopted for healthcare applications. To handle heterogeneous data distribution in the internet of things field, Gao et al. suggested a cross-technology communication-based FL for considering different types of standardizations for wireless network technology.[16] To enhance robustness and precise performance, an optimizer was also studied using global model optimization.[17] Lian et al. proposed a decentralized, efficient, and privacy-enhanced federated edge learning method for cyber-physical systems in healthcare.[18] They designed a hierarchical ring topology to alleviate the centralization of the conventional algorithm and proposed an aggregation algorithm for distributed medical institutions to generate a global model.

In this study, we focused on solving the FL problem that has been identified in cross-silo studies. In real-world healthcare applications in cross-silo settings, different clients have different feature spaces, except for some common features, which implies that the FL model must

[1]Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea
[2]Research Institute for Data Science and AI (Artificial Intelligence), Hallym University, Chuncheon-si, Gangwon-do, Republic of Korea
[3]Division of Pulmonology, Department of Internal Medicine, Yonsei University Health System, Seoul, Republic of Korea
[4]These authors contributed equally
[5]Lead contact
*Correspondence: yurangpark@yuhs.ac
https://doi.org/10.1016/j.isci.2024.110943

Table 1. Description of data distribution based on the type of intensive care unit for different demographic variables obtained from Severance Hospital and electronic intensive care unit datasets

	Internal validation			External validation	
Variable	HICU (*n* = 5,266)	MICU (*n* = 14,550)	SICU (*n* = 10,644)	NSICU (*n* = 12,706)	*p* value *
Demographic					
Age, year, mean ± SD	65.0 ± 13.2	63.1 ± 17.2	62.3 ± 17.2	60.9 ± 17.7	<0.001
Gender (female), n (%)	1,604 (30.5)	7,017 (48.2)	4,602 (43.2)	6,087 (47.9)	<0.001
Height, cm, mean ± SD	164.3 ± 11.6	169.0 ± 12.8	170.1 ± 11.5	169.6 ± 12.4	<0.001
Weight, kg, mean ± SD	66.5 ± 13.3	83.8 ± 27.8	84.0 ± 26.5	83.0 ± 25.0	<0.001
Clinical					
HR, beats/min, mean ± SD	72.8 ± 18.0	103.8 ± 31.1	102.3 ± 29.7	95.6 ± 30.5	<0.001
RR, breaths/min, mean ± SD	17.4 ± 4.1	28.0 ± 15.0	24.7 ± 14.9	22.3 ± 14.2	<0.001
BP, mm Hg, mean ± SD	83.8 ± 13.3	86.1 ± 42.0	88.9 ± 42.1	97.2 ± 42.1	<0.001
TEMP, °C, mean ± SD	37.0 ± 0.5	36.4 ± 1.0	36.4 ± 0.9	36.5 ± 0.8	<0.001
RBC count, million/mm, mean ± SD	4.1 ± 0.6	3.7 ± 0.7	3.6 ± 0.7	3.9 ± 0.7	<0.001
HCT, vol %, mean ± SD	38.0 ± 5.4	32.9 ± 6.4	32.3 ± 6.0	35.2 ± 5.7	<0.001
Creatinine, umol/L, mean ± SD	4.0 ± 16.3	1.8 ± 1.9	1.5 ± 1.5	1.1 ± 1.2	<0.001
BUN, mg/dL, mean ± SD	18.4 ± 10.8	32.0 ± 23.6	24.3 ± 17.7	19.4 ± 15.4	<0.001
Sodium, mEq/L, mean ± SD	137.7 ± 7.8	137.5 ± 5.7	137.4 ± 4.7	138.6 ± 4.9	0.012
Potassium, mEq/L, mean ± SD	4.3 ± 1.4	4.1 ± 0.5	4.0 ± 0.4	4.0 ± 0.4	<0.001
In-hospital death, n (%)	368 (7.0)	1,887 (13.0)	949 (8.9)	1,056 (8.3%)	<0.001

HICU, heart intensive care unit; MICU, medical intensive care unit; SICU, surgical intensive care unit; NSICU, neurosurgical intensive care unit; HR, heart rate; RR, respiratory rate; BP, blood pressure; TEMP, temperature; RBC, red blood cell; HCT, hematocrit; BUN, blood urea nitrogen; SD, standard deviation.
* *p* value was calculated with one-way analysis of variance for continuous features and $\chi 2$ test for categorical features.

consider both HFL and vertical FL which assumed clients have different datasets but share the same feature space. Certain studies have addressed these limitations.[19,20] Lu et al. proposed FedAP to utilize client-specific layers by using a batch normalization layer to solve the distribution gaps among clients.[21] A model-agnostic meta-learning approach was also suggested for personalized learning based on gradient descent with respect to their own data.[22] Ma et al. proposed PerHeFed, a convolutional neural network-based representation aggregation for personalized layer.[23] A personalized FL framework for image data was proposed for aggregating local-sub models. Liu et al. suggested federated autonomous deep learning to predict patient mortality based on data from 58 hospitals.[24] They split the architecture into two parts: the first half of the neural network was trained globally using all data sources in a federated manner and the second half of the neural network was trained locally on data from each data source. Arivazhagan et al. developed FL with personalization layers (FedPer) by separating the model into base and personalized layers to address the common and personal knowledge.[20] Similar and shared weights were assigned for the base layer, and a distinct layer was added to adapt the vertical features. However, most of these algorithms exhibited limitations in the utilization of the same feature space and distinct layers for considering client-specific records, resulting in poor performance. Moreover, many of them required validation and generalization before being applied to diverse real-world datasets.

The aim of this study was to develop an approach called personalized progressive FL (PPFL), which combines FL with variants of progressive neural networks that can simultaneously consider common and client-specific features.[25]

RESULTS

Basic characteristics of experiment datasets

The intensive care unit (ICU)-related data for PPFL consisted of 5,266, 14,550, 10,644, and 12,706 data points from the medical ICU (MICU), surgical ICU (SICU), heart ICU (HICU), and neurosurgical ICU (NSICU), respectively. The mean (standard deviation [SD]) age of patients ranged from 60.9 (17.7) to 65.0 (13.2) years. Female patients accounted for 30.5% to 48.2% of these datasets. Among these, 7.0%, 13.0%, 8.9%, and 8.3% of data were of in-hospital mortality patients in the HICU, MICU, SICU, and NSICU datasets, respectively. Details of these characteristics are listed in Table 1.

The non-small cell lung cancer dataset for PPFL comprised 2,315 and 982 data points from the Memorial Sloan Kettering (MSK) Cancer Center and The Cancer Genome Atlas (TCGA) datasets, respectively. The mean (SD) ages of the patient were 56.3 (26.6) and 64.5 (14.4) years. Female patients comprised 60.3% and 40.2% of the data in the MSK and TCGA datasets, respectively. Among these data, 49.6% and 28.4% of the data were of deceased patients, respectively. Details of patient demographics are listed in Table S1.

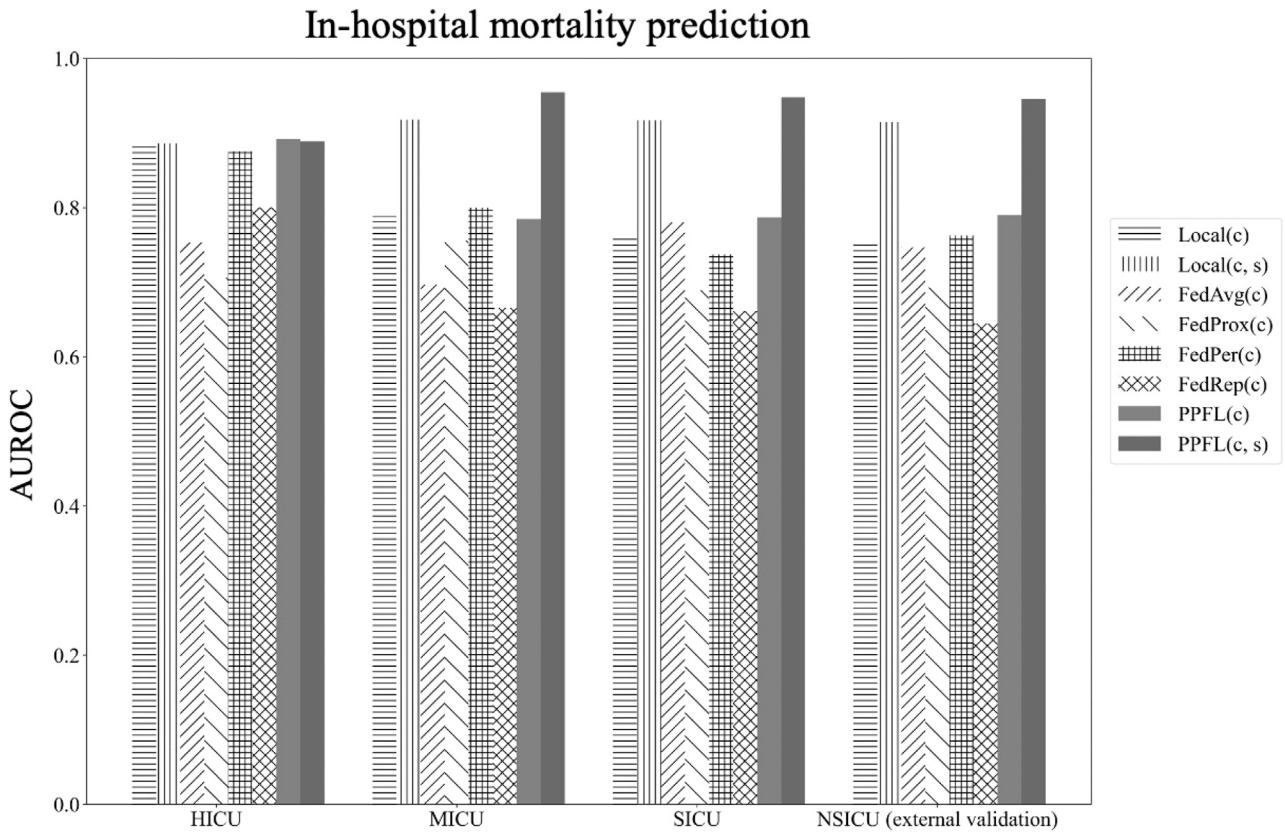


Figure 1. Performance comparison between PPFL and FedAvg, FedProx, FedPer, FedRep, Local (c), and Local (c, s) in terms of AUROC

PPFL (c, s) shows the highest AUROC score for predicting in-hospital mortality task. PPFL, personalized progressive federated learning; HICU, heart intensive care unit; MICU, medical intensive care unit; SICU, surgical intensive care unit; NSICU, neurosurgical intensive care unit; AUROC, area under the receiver operating characteristic; FedAvg, federated averaging; FedPer, federated learning with personalization layers; Local, local algorithms; (c), using only common features; (c, s), using both common and specific features.

Performance comparison between PPFL and conventional algorithms

When predicting in-hospital mortality, PPFL (c, s), which is based on both common and specific features, showed the highest score, with an accuracy of 0.939 and area under the receiver operating characteristic (AUROC) score of 0.934 on average in the internal assessment (Figure 1; Table 2).

In most ICU data, PPFL (c, s) achieved the highest AUROC, except for the HICU. In the HICU data, PPFL (c), which is based on only common features, exhibited the highest performance, with an AUROC of 0.892 and accuracy of 0.932, which were 0.1% and 0.3% higher than those of PPFL (c, s), respectively. Compared to the local algorithm using common and specific features (Local (c, s)) and federated averaging algorithm using common features (FedAvg (c)), the AUROC values of PPFL (c, s) improved by 2.5% and 19.7% on average, respectively. FedAvg (c) exhibited an average AUROC of 0.744, which was the lowest score among all models. Moreover, as the number of clients increased, the AUROC increased for all clients. In the external evaluation based on the NSICU data from the electronic ICU (eICU) dataset, PPFL (c, s) showed an AUROC of 0.948, which was 3.1% higher than that of the local algorithm based on both common and specific features (Local (c, s)).

When predicting mortality on non-small cell lung cancer data, PPFL (c, s) achieved the highest performance with AUROCs of 0.71 and 0.75 on MSK and TCGA datasets, when compared to the average values of 0.62 and 0.57 of the other algorithms, respectively, as shown in Figure S1. PPFL (c, s) achieved 7% and 30% higher AUROCs than those of PPFL (c) for MSK and TCGA datasets, respectively. Local (c) and PPFL (c) showed the lowest performances on the MSK and TCGA datasets, respectively.

Statistical analysis for performance comparison among algorithms

The AUROCs were compared based on the DeLong test, and the results are listed in Table 3. The p values of PPFL (c) and Local (c, s) on the HICU data (0.25 and 0.2) were not statistically different from those of PPFL (c, s). The p value of Local (c, s) on TCGA data (0.06) was also not different. Among the 28 results, 23 (82.1%) showed critical statistical significance ($p < 0.001$).

Table 2. Performance evaluation of PPFL in comparison with FedAvg, FedProx, FedPer, FedRep, and Local algorithms in internal and external validations using distributed real-world data

In-hospital mortality		
Model	Accuracy	AUROC
Internal validation		
Sev-HICU		
FedAvg (c)	0.929	0.753
FedProx (c)	0.930	0.706
FedPer (c)	0.930	0.876
FedRep (c)	0.931	0.800
PPFL (c)	0.932	0.892
PPFL (c, s)	0.931	0.889
Local (c)	0.930	0.886
Local (c, s)	0.927	0.886
eICU-MICU		
FedAvg (c)	0.812	0.697
FedProx (c)	0.870	0.756
FedPer (c)	0.874	0.800
FedRep (c)	0.848	0.666
PPFL (c)	0.867	0.785
PPFL (c, s)	**0.928**	**0.955**
Local (c)	0.881	0.789
Local (c, s)	0.906	0.918
eICU-SICU		
FedAvg (c)	0.890	0.747
FedProx (c)	0.911	0.693
FedPer (c)	0.910	0.763
FedRep (c)	0.890	0.645
PPFL (c)	0.916	0.790
PPFL (c, s)	**0.957**	**0.946**
Local (c)	0.911	0.752
Local (c, s)	0.927	0.915
External validation		
NSICU		
FedAvg (c)	0.890	0.780
FedProx (c)	0.917	0.689
FedPer (c)	0.875	0.737
FedRep (c)	0.902	0.661
PPFL (c)	0.918	0.787
PPFL (c, s)	**0.941**	**0.948**
Local (c)	0.915	0.765
Local (c, s)	0.908	0.917

PPFL, personalized progressive federated learning; c, common features; s, client-specific feature; HICU, heart intensive care unit; MICU, medical intensive care unit; SICU, surgical intensive care unit; NSICU, neurosurgical intensive care unit; eICU, electronic intensive care unit; Sev, Severance Hospital; FedAvg, federated averaging; FedPer, federated learning with personalization layers.

Table 3. Comparison between the area under receiver operating characteristic scores of PPFL (c, s) and other algorithms on two datasets

DeLong	PPFL (c)	FedAvg	FedProx	FedPer	FedRep	Local (c)	Local (c, s)
ICU							
HICU	***0.25***	<0.001	<0.001	0.008	<0.001	0.01	***0.2***
MICU	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
SICU	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.002
NSICU	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
NSCLC							
MSK	0.068	0.009	0.012	0.007	0.004	0.004	0.006
TCGA	0.001	0.007	0.013	0.009	0.008	0.006	***0.06***

ICU, intensive care unit; NSCLC, non-small cell lung cancer; PPFL, personalized progressive federated learning; HICU, heart intensive care unit; MICU, medical intensive care unit; SICU, surgical intensive care unit; NSICU, neurosurgical intensive care unit; NSCLC, non-small cell lung cancer; MSK, Memorial Sloan Kettering; TCGA, The Cancer Genome Atlas; c, common feature; cs, common and specific features *p* value was calculated using DeLong test for the AUROC comparison. *p* value > 0.05 indicates no statistical difference with the AUROC score of PPFL (c, s) and is presented in bold and italicized font.

Contribution of common and vertical features based on Shapley value for each client

The contributions of the common and vertical features from all clients in predicting in-hospital mortality are shown in Figure 2. The Shapley value patterns of the HICU varied from those of other ICUs (MICU, SICU, and NSICU). Among the common features, the heart rate and blood pressure exhibited the highest Shapley additive explanations (SHAPs) values in the HICU and other ICUs, respectively. Higher values of heart

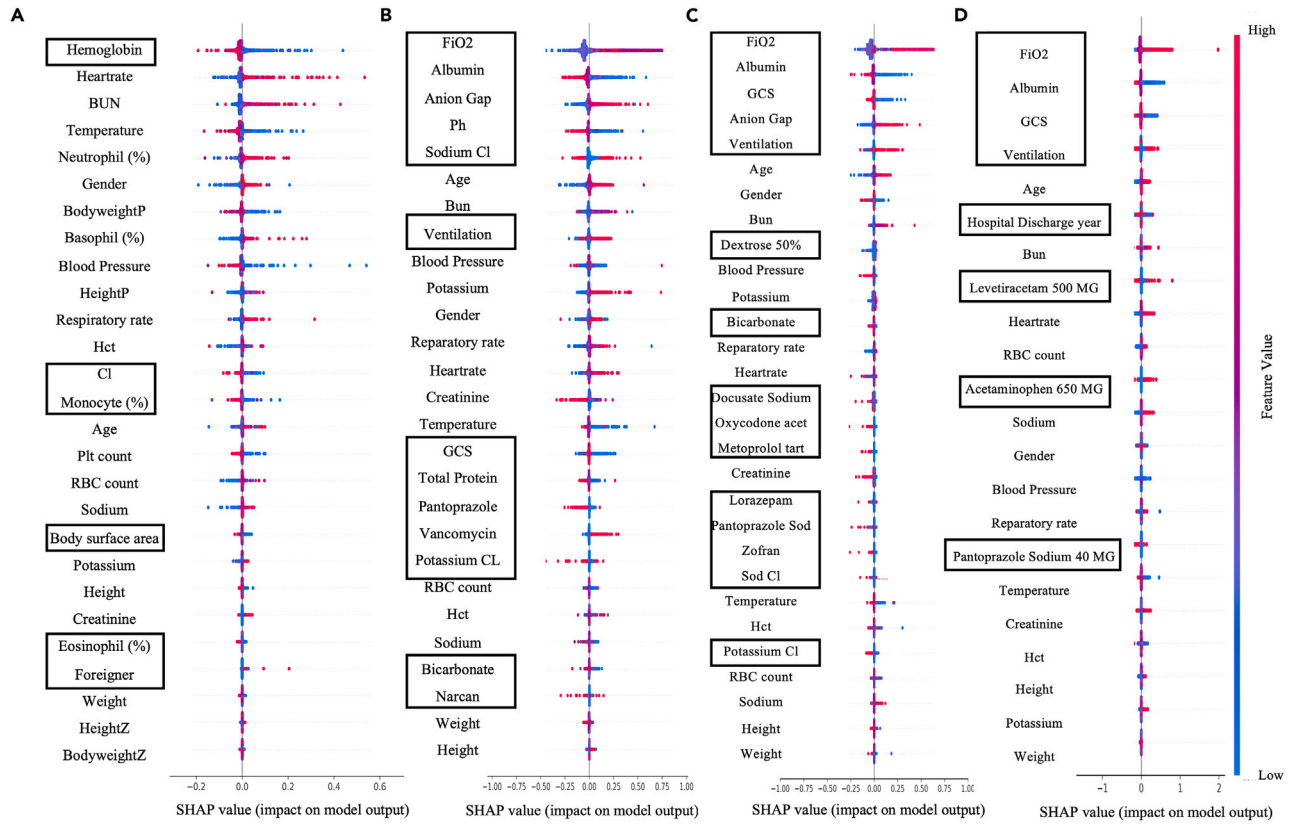


Figure 2. SHAP values of common and vertical features in predicting in-hospital mortality

(A) Heart intensive care unit.

(B) Medical intensive care unit.

(C) Surgical intensive care unit.

(D) Neurosurgical intensive care unit. Client-specific vertical features are highlighted with a black box. BUN, blood urea nitrogen; BP, blood pressure; RBC, red blood cell; HCT, hematocrit.

rate for all ICUs and blood urea nitrogen for the MICU, SICU, and SICU indicate a higher likelihood of mortality; conversely, smaller values of temperature for the HICU, SICU, and NSICU and blood pressure for the MICU, SICU, and NSICU indicated a greater likelihood of mortality. Among the vertical and common features, hemoglobin for HICU and fraction of inspired oxygen (FiO2) for the other ICUs exhibited the highest SHAP values. Larger value of FiO2 and smaller value of albumin in the MICU, SICU, and NSICU indicated a higher risk of mortality.

DISCUSSION

In this paper, we propose a PPFL algorithm for heterogeneously distributed clients that expands the feature space for client-specific vertical features. The proposed model not only learns global knowledge from common feature information but also expands the feature space related to client-specific vertical features by creating column networks. In previous studies, vertical feature information and its parameters were never transmitted or exchanged. We achieved the highest score, with AUROCs of 0.939 and 0.948 for the internal and external evaluations, respectively. We also determined client-specific feature importance using the proposed algorithm. This is the FL study in the medical field that considers not only common features but also client-specific vertical features by applying progressive learning.

Our proposed model can solve the primary limitation existing in clinical practice that features from different hospitals are typically collected in different ways. Recently, several studies have been conducted to train different properties of data based on personalized federated networks. However, they only utilized the weight allocation method with the same feature space using public open datasets. A study proposed in 2023 divided the contribution of training components into two: sharing and personalized components; however, the difference is not related to the feature set view, but it is related to patient-level separation.[26] Federated personalized networking using EMR was studied with respect to covariate adjusting with the propensity score matching method with the same feature space.[27] The proposed PPFL can consider all the collected features and samples of multiple clients that other models cannot utilize. Therefore, it exhibits superior performance when compared to the existing models. FedAvg exhibited a limited feature space because only common features from multiple clients are considered as inputs to the model in terms of its structure. The local model uses only a sample from each client; therefore, the number of samples is inevitably smaller than that of the PPFL input dataset. Moreover, the proposed model is generic and can be applied to other collaboratively aggregated methods based on deep learning models.

We identified differences in the importance of features in ICUs and feature type for predicting mortality based on the Shapley value, as shown in Figure 2. The HICU exhibited a different pattern in terms of feature importance. Blood pressure was the highest contributor in the MICU, SICU, and NSICU, whereas heart rate was the major factor in the HICU, which is consistent with the results of Fallah et al.[28] As PPFL can consider specific features of each client for modeling, we determined that larger FiO2 and smaller albumin levels in the MICU, SICU, and NSICU could contribute to mortality, which is consistent with previous research.[29,30] Moreover, all the most important features were client-specific vertical features. We identified this finding using the PPFL; however, a conventional model that considers only common features cannot achieve this.

As only a few studies have been conducted with real-world data scenarios on FL, and demands for experiments with real-world data are being emphasized,[28,31] this study used real-world clinical data from multiple ICU clients, which showed high performance. This is the rationale for the PPFL becoming a clinically applicable algorithm.

Communication costs in FL can affect real-world applications. The PPFL architecture for weight updates comprises two main parts. First, after updating the weight with a common vertical network, the vertical network was fine-tuned at each local client. Because we utilized the vertical network weight as an initial weight and only fine-tuning was conducted, the communication cost for constructing the models was significantly less. As shown in Figure S2, PPFL required around 25 epochs for the model to achieve a certain performance, whereas other models required 200 epochs which implies the applicability for real-world applications.

Limitations of the study

This study has several limitations. First, the study was not conducted in a real distributed environment. However, the database was allocated and treated as a dependent object to mimic real-world distribution. Second, an algorithm pipeline should be developed for clinical application and prospective validation. Each client must define specific parameters for the application. A prospective evaluation should be conducted after implementation, which is a future direction for our study. Third, the statistical challenge of FL when the local data are not independent and identically distributed (non-IID) was not considered. Data distribution can significantly affect the performance of the learning process. To consider non-IID settings, updating the PPFL, including distillation or allocation of weight, should be conducted in future studies. Finally, we did not consider the practical delay between obtaining data from different clients in the FL environment, which can affect communication costs. However, we measured the number of epochs required for each algorithm to calculate the communication cost.

In real-world healthcare problems, the feature spaces of data from different clients can differ. In addition to utilizing specific and common features from each client, we proposed a high-performance PPFL algorithm to personalize federated algorithms for heterogeneously distributed clients and expand the feature space for client-specific vertical feature information. Moreover, we investigated the performance improvement and robustness of our proposed model using real-world EMR data and validated the usefulness of the model.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and details of reagents should be directed to and will be fulfilled by the lead contact, Yu Rang Park (yurangpark@yuhs.ac).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The datasets generated and/or analyzed during the current study are available on the website (https://eicu-crd.mit.edu/gettingstarted/access/) for eICU data, and the Severance Hospital data can be obtained from the corresponding author upon reasonable request.
- The underlying code for this study is available at GitHub and can be accessed via the link https://github.com/DigitalHealthcareLab/Personalized-Progressive-Federated-Learning.

ACKNOWLEDGMENTS

This work was supported by the Bio-Industrial Technology Development Program (20014841), funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea).

AUTHOR CONTRIBUTIONS

Conceptualization, T.H.K. and Y.R.P.; methodology, T.H.K.; software, T.H.K.; validation, W.S.J. and S.C.H.; formal analysis, T.H.K.; investigation, T.H.K. and K.C.; resources, W.S.J. and J.H.; data curation, S.C.H., M.S., and K.C.; writing – original draft preparation, T.H.K. and J.Y.Y.; writing – review and editing, W.S.J. and Y.R.P.; visualization, T.H.K. and W.S.J.; supervision, Y.R.P.; project administration, Y.R.P.; funding acquisition, Y.R.P. All the authors have read and agreed to the published version of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Experimental setting and hyperparameters
- METHOD DETAILS
  - Problem setting and overall network architecture of PPFL
  - Horizontal network for common features
  - Vertical network for specific features
  - PPFL for common and specific features
  - Datasets for PPFL experiment in healthcare
  - Comparing the performance with conventional algorithms
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Contribution of common and specific features of each client
- ADDITIONAL RESOURCE

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.110943.

Received: December 4, 2023
Revised: March 29, 2024
Accepted: September 10, 2024
Published: September 13, 2024

REFERENCES

1. Kairouz, P., McMahan, H.B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A.N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., and D'Oliveira, R.G. (2021). Advances and Open Problems in Federated Learning. Found. Trends Mach. Learn. *14*, 1–210.
2. McMahan, B., Moore, E., Ramage, D., Hampson, S., and y Arcas, B.A. (2017). Communication-efficient Learning of Deep Networks from Decentralized Data (PMLR), pp. 1273–1282.
3. Rieke, N., Hancox, J., Li, W., Milletarì, F., Roth, H.R., Albarqouni, S., Bakas, S., Galtier, M.N., Landman, B.A., Maier-Hein, K., et al. (2020). The future of digital health with federated learning. NPJ Digit. Med. *3*, 119. https://doi.org/10.1038/s41746-020-00323-1.
4. Yang, Q., Liu, Y., Chen, T., and Tong, Y. (2019). Federated Machine Learning: Concept and Applications. ACM Trans. Intell. Syst. Techn. (TIST) *10*, 1.
5. Kholod, I., Yanaki, E., Fomichev, D., Shalugin, E., Novikova, E., Filippov, E., and Nordlund, M. (2020). Open-source federated learning frameworks for IoT: A comparative review and analysis. Sensors *21*, 167.
6. Dayan, I., Roth, H.R., Zhong, A., Harouni, A., Gentili, A., Abidin, A.Z., Liu, A., Costa, A.B., Wood, B.J., Tsai, C.-S., et al. (2021). Federated learning for predicting clinical outcomes in patients with COVID-19. Nat. Med. *27*, 1735–1743. https://doi.org/10.1038/s41591-021-01506-3.
7. Vaid, A., Jaladanki, S.K., Xu, J., Teng, S., Kumar, A., Lee, S., Somani, S., Paranjpe, I., De Freitas, J.K., Wanyan, T., et al. (2021). Federated Learning of Electronic Health Records to Improve Mortality Prediction in Hospitalized Patients With COVID-19: Machine Learning Approach. JMIR Med. Inform. *9*, e24207. https://doi.org/10.2196/24207.
8. Rajendran, S., Obeid, J.S., Binol, H., D Agostino, R., Jr., Foley, K., Zhang, W., Austin, P., Brakefield, J., Gurcan, M.N., and Topaloglu, U. (2021). Cloud-Based Federated Learning Implementation Across Medical Centers. JCO Clin. Cancer Inform. *5*, 1–11. https://doi.org/10.1200/cci.20.00060.

9. Nazir, S., and Kaleem, M. (2023). Federated Learning for Medical Image Analysis with Deep Neural Networks. Diagnostics *13*, 1532. https://doi.org/10.3390/diagnostics13091532.
10. Zhu, H., and Jin, Y. (2020). Multi-objective evolutionary federated learning. IEEE Trans. Neural Netw. Learn. Syst. *31*, 1310–1322.
11. Guha, N., Talwalkar, A., and Smith, V. (2019). One-shot Federated Learning. Preprint at arXiv. https://doi.org/10.48550/arXiv.1902.11175.
12. Lin, Y., Han, S., Mao, H., Wang, Y., and Dally, W.J. (2017). Deep Gradient Compression: Reducing the Communication Bandwidth for Distributed Training. Preprint at arXiv. https://doi.org/10.48550/arXiv.1712.01887.
13. Haddadpour, F., Kamani, M.M., Mokhtari, A., and Mahdavi, M. (2021). Federated Learning with Compression: Unified Analysis and Sharp Guarantees (PMLR), pp. 2350–2358.
14. Braverman, M., Garg, A., Ma, T., Nguyen, H.L., and Woodruff, D.P. (2016). Communication Lower Bounds for Statistical Estimation Problems via a Distributed Data Processing Inequality. Preprint at arXiv. https://doi.org/10.48550/arXiv.1506.07216.
15. Konečný, J., McMahan, H.B., Yu, F.X., Richtárik, P., Suresh, A.T., and Bacon, D. (2016). Federated Learning: Strategies for Improving Communication Efficiency. Preprint at arXiv. https://doi.org/10.48550/arXiv.1610.05492.
16. Gao, D., Wang, H., Guo, X., Wang, L., Gui, G., Wang, W., Yin, Z., Wang, S., Liu, Y., and He, T. (2023). Federated learning based on CTC for heterogeneous internet of things. IEEE Intern. Things J. *10*, 22673–22685.
17. Lian, Z., Zeng, Q., Wang, W., Xu, D., Meng, W., and Su, C. (2023). Traffic Sign Recognition using Optimized Federated Learning in Internet of Vehicles. IEEE Intern. Things J. *11*, 6722–6729.
18. Lian, Z., Yang, Q., Wang, W., Zeng, Q., Alazab, M., Zhao, H., and Su, C. (2022). DEEP-FEL: Decentralized, efficient and privacy-enhanced federated edge learning for healthcare cyber physical systems. IEEE Trans. Netw. Sci. Eng. *9*, 3558–3569.
19. Prayitno, Shyu, C.R., Putra, K.T., Chen, H.C., Tsai, Y.Y., Hossain, K.S.M.T., Jiang, W., and Shae, Z.Y. (2021). A systematic review of federated learning in the healthcare area: From the perspective of data properties and applications. Appl. Sci. *11*, 11191.
20. Arivazhagan, M.G., Aggarwal, V., Singh, A.K., and Choudhary, S. (2019). Federated Learning with Personalization Layers. Preprint at arXiv, 1. https://doi.org/10.48550/arXiv.1912.00818.
21. Lu, W., Wang, J., Chen, Y., Qin, X., Xu, R., Dimitriadis, D., and Qin, T. (2022). Personalized federated learning with adaptive batchnorm for healthcare. IEEE Trans. Big Data. 1–1.
22. Fallah, A., Mokhtari, A., and Ozdaglar, A. (2020). Personalized federated learning with theoretical guarantees: A model-agnostic meta-learning approach. Adv. Neural Inf. Process. Syst. *33*, 3557–3568.
23. Ma, L., Liao, Y., Zhou, B., and Xi, W. (2023). PerHeFed: A general framework of personalized federated learning for heterogeneous convolutional neural networks. World Wide Web *26*, 2027–2049.
24. Liu, D., Miller, T., Sayeed, R., and Mandl, K.D. (2018). FADL: Federated-Autonomous Deep Learning for Distributed Electronic Health Record. Preprint at arXiv. https://doi.org/10.48550/arXiv.1811.11400.
25. Rusu, A.A., Rabinowitz, N.C., Desjardins, G., Soyer, H., Kirkpatrick, J., Kavukcuoglu, K., Pascanu, R., and Hadsell, R. (2016). Progressive Neural Networks. Preprint at arXiv. https://doi.org/10.48550/arXiv.1606.04671.
26. Zhao, Y., Liu, Q., Liu, X., and He, K. (2023). Medical Federated Model with Mixture of Personalized and Sharing Components. Preprint at arXiv. https://doi.org/10.48550/arXiv.2306.14483.
27. Tarumi, S., Suzuki, M., Yoshida, H., Miyauchi, S., and Kurazume, R. (2023). Personalized Federated Learning for Institutional Prediction Model using Electronic Health Records: A Covariate Adjustment Approach. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. *2023*, 1–4. https://doi.org/10.1109/embc40787.2023.10339940.
28. Fallah, A., Mokhtari, A., and Ozdaglar, A. (2020). Personalized Federated Learning: A Meta-Learning Approach. Preprint at arXiv. https://doi.org/10.48550/arXiv.2002.07948.
29. Atrash, A.K., and de Vasconcellos, K. (2020). Low albumin levels are associated with mortality in the critically ill: A retrospective observational study in a multidisciplinary intensive care unit. South. Afr. J. Crit. Care *36*, 74–79. https://doi.org/10.7196/SAJCC.2020.v36i2.422.
30. Bi, H., Liu, X., Chen, C., Chen, L., Liu, X., Zhong, J., and Tang, Y. (2023). The PaO_2/FiO_2 is independently associated with 28-day mortality in patients with sepsis: a retrospective analysis from MIMIC-IV database. BMC Pulm. Med. *23*, 187. https://doi.org/10.1186/s12890-023-02491-8.
31. Tan, A.Z., Yu, H., Cui, L., and Yang, Q. (2023). Towards personalized federated learning. IEEE Trans. Neural Network. Learn. Syst. *34*, 9587.
32. Kingma, D.P., and Ba, J. (2014). Adam: A Method for Stochastic Optimization. Preprint at arXiv. https://doi.org/10.48550/arXiv.1412.6980.
33. Ruder, S. (2016). An Overview of Gradient Descent Optimization Algorithms. Preprint at arXiv. https://doi.org/10.48550/arXiv.1609.04747.
34. Pollard, T.J., Johnson, A.E.W., Raffa, J.D., Celi, L.A., Mark, R.G., and Badawi, O. (2018). The eICU Collaborative Research Database, a freely available multi-center database for critical care research. Sci. Data *5*, 180178.
35. Baraniuk, R. (2007). Compressive sensing [lecture notes]. IEEE Signal Process. Mag. *24*, 118–121.
36. Lundberg, S.M., and Lee, S.-I. (2017). A unified approach to interpreting model predictions. Adv. Neural Info. Process. Syst. *30*, 4765–4774.
37. Shrikumar, A., Greenside, P., Shcherbina, A., and Kundaje, A. (2016). Not Just a Black Box: Learning Important Features through Propagating Activation Differences. Preprint at arXiv. https://doi.org/10.48550/arXiv.1605.01713.
38. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., and Devin, M. (2016). Tensorflow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. Preprint at arXiv. https://doi.org/10.48550/arXiv.1603.04467.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
Python (version 3.7.15)	Python Software	https://www.python.org
Deposited data		
MIMIC open data	Physionet	https://physionet.org/content/mimiciv/3.0/
MSK	https://pubmed.ncbi.nlm.nih.gov/36357680/	https://www.cbioportal.org/study/clinicalData?id=nsclc_ctdx_msk_2022
TCGA	https://pubmed.ncbi.nlm.nih.gov/27158780/	https://www.cbioportal.org/datasets/
Code for PPFL	This paper	https://github.com/DigitalHealthcareLab/Personalized-Progressive-Federated-Learning

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Experimental setting and hyperparameters

We evaluated the performance of the binary classifications for in-hospital mortality as a binary class (dead or alive). We divided the training, validation, and test datasets in a ratio of 6:2:2 for each client participating in the PPFL training. The validation dataset was used to search for hyperparameters using a random-search algorithm.

Among the eICU datasets, the NSICU data were defined as unseen data for external validation. After selecting the common features from the HICU, SICU, and MICU data, we trained the algorithms to learn the weights via HorizontalNet. We optimized the weight parameters of the models by incorporating stochastic gradient descent using the Adam optimizer.[32] For the application of our proposed and comparative models for binary classification, we utilized the cross-entropy loss function. For hyperparameter tuning, 100 epochs were set for local training in FL and 30 rounds were used to aggregate the local models. This study was approved by the Institutional Review Board of Severance Hospital (IRB approval no. "4-2021-0820").

METHOD DETAILS

Problem setting and overall network architecture of PPFL

We propose the PPFL algorithm for conducting client-specific personalized inferences on heterogeneous data settings. The PPFL also addresses the limited information availability of FL design by leveraging not only common features but also client-specific vertical features across distributed clients. The proposed process comprises two major steps. First, we built an HFL on a central server using only the common features of distributed clients. Second, the pre-trained horizontal federated model was deployed for each client to learn personalized knowledge for client-specific inference tasks through the PPFL. The proposed PPFL considered both a horizontal FL network (HorizontalNet), which receives inputs as weights from a globally trained model based on common features of the client, and a vertical network (VerticalNet), which learns specific features of the client.

This study addressed the problem of solving the case in which both common and client-specific features exist (Figure S3). For better understanding, a mathematical description is presented in Table S2. Let us assume that an individual client k has a dataset $D_k := \{\boldsymbol{c}_i^k, \boldsymbol{s}_i^k, y_i^k\}_{i=1}^{m^{(k)}} = \{\mathbf{C}^k, \mathbf{S}^k, \boldsymbol{y}^k\}$ consisting of $m^{(k)}$ samples, where the client $k \in \mathcal{K} := \{1,\ldots,K\}$. The dataset of client k comprises a common feature matrix $\mathbf{C}^k \in \mathbb{R}^{m \times p}$ and client-specific vertical feature matrix $\mathbf{C}^k \in \mathbb{R}^{m \times q}$. The i-th sample of D_k can be represented using a common feature with a p-dimensional column vector $\boldsymbol{c}_i^k := \{c_i^{1(k)}, c_i^{2(k)}, \ldots, c_i^{p(k)}\} \in \mathbb{R}^{1 \times p}$ and a client-specific vertical feature with a q-dimensional column vector $\boldsymbol{s}_i^k := \{s_i^{1(k)}, s_i^{2(k)}, \ldots, s_i^{q(k)}\} \in \mathbb{R}^{1 \times q}$; moreover, the corresponding target variable is y_i^k. Note that the attributes and dimension $p^{(k)}$ of the common feature vector $\boldsymbol{c}_i^k$ are identical for all clients $k \in \mathcal{K}$; However, the attributes and dimension $q^{(k)}$ of the client-specific vertical feature vector $\boldsymbol{s}_i^k$ may not be identical for all clients.

The PPFL model is proposed to address the limitations of FL design with respect to data availability, and aims to personalize the client-specific inference of every client while maintaining the globally learned information. PPFL is a multi-model-based approach that generates different values for every client at the deployment step and aims to personalize the client-specific inference of every client while maintaining the globally learned information and solve the problem between globally generalized knowledge and client-specific knowledge. The concept of lateral connection was utilized in progressive neural networks,[25] which were proposed for leveraging transfer and avoiding catastrophic forgetting in multitask learning. Figure S3 shows the architecture of the PPFL model.

In PPFL, building a personalized model allows client-specific distributions to be learned from a globally learned FL model by transmitting layer-wise knowledge to different network columns. PPFL contains three network columns: HorizontalNet, VerticalNet, and PersonalizedNet.

The PPFL network, denoted by $f_k(\cdot)$, is trained and personalized for each client k. $f_k(\cdot)$ is a final function that maps the inputs c_i^k and s_i^k to estimated output y_i^k. We omit the notation k in the notations related to the weights and hidden layers in the three network columns that are subsequently.

Horizontal network for common features

HorizontalNet, which is the first column of PPFL, is a network that is transferred from the horizontal federated model. The internal weight parameters of the HorizontalNet column ω^c were initialized using parameter ω^c of the horizontal federated model.

Under the federated learning framework, the internal parameters of HorizontalNet ω^c was globally updated based on FedAvg through the communication between the central server and participating clients. In this communication, the weights of client k, was globally updated considering only common features.

Aggregate algorithm, FedAvg, integrates the weight parameters of the models within each client ω^k using weighted averaging to train internal weight parameters of HorizontalNet ω^c. The weight parameters of client k, ω^k is updated by solve the object function with loss function $l_i(\cdot)$ of the prediction on example of the prediction on example (c_i, y_i). m is the total sample size of K clients. Therefore, the objective function for solving the empirical risk minimization is:

$$\min_{\omega^c \in \mathbb{R}^d} \mathcal{L}^c(\omega^c) := \sum_{k=1}^{K} \frac{m^{(k)}}{m} \mathcal{L}_k^c(\omega^k), \text{ where } \mathcal{L}_k^c(\omega^k) := \sum_{c_i \in D_k} l_i(\omega^k). \quad \text{(Equation 1)}$$

This network column aims to pass generalized knowledge to personalized networks using the common feature data x^k of the client as input information. Note that the internal weight matrix ω^c in HorizontalNet, which is not connected, is "frozen" to train. However, the lateral weight parameter $\omega^{c'}$, which is the weight parameter connected between HorizontalNet and the final layer can be updated using an optimization algorithm.

The hidden layers h_l^c in the HorizontalNet column for the common feature vector x_k of the client are computed using the internal weight parameter $\omega_l^{c^{int}}$ of HorizontalNet is computed as follows:

$$h_{l+1}^c = \sigma\left(\omega_l^c h_l^c + b_l^c\right), \text{ where } h_0^c = c^k. \quad \text{(Equation 2)}$$

The output values of the HorizontalNet hidden layers h_l^c are transferred to PPFL via the lateral weight parameter $\omega^{c'}$ without overlaying the original internal weight parameter ω^c of HorizontalNet. Thus, the internal parameter of HorizontalNet ω^c is not a trainable weight parameter for retaining the globally learned knowledge for the common feature space, and its lateral weight parameter $\omega^{c'}$ allows the transfer of proper knowledge from h_l^c to the final layer h_l^p.

Vertical network for specific features

The second network column is VerticalNet. This network progressively expands the feature space with respect to the specific vertical features of the client. The input of VerticalNet is the client-specific vertical feature data, $s^k \in D_k$. The weight parameter ω^v is the internal weight parameter of VerticalNet, which is not connected to the last layer. The lateral weight parameter $\omega^{v'}$ consists of the parameters of VerticalNet and the last layer. Both ω^v and $\omega^{v'}$ can be learned through the training step because these parameters are constructed to expand the feature space and to connect to the layer used for an inference task. Thus, the internal and lateral weight parameters ω^v and $\omega^{v'}$, respectively, learn client-specific vertical feature information and transmit this knowledge to the last layer. The hidden layers h_l^v in the VerticalNet column with respect to the client-specific vertical feature s^k and internal weight parameter $\omega^{v^{int}}$ are obtained as follows:

$$h_{l+1}^v = \sigma\left(\omega_l^v h_l^v + b_l^v\right), \text{ where } h_0^v = s^k. \quad \text{(Equation 3)}$$

PPFL for common and specific features

We applied a lateral connection in a progressive neural network[25] to expand the layer-wise feature space from a globally pre-trained FL model. The weight parameters of the PersonalizedNet column learn the specific personalized knowledge of the client by acquiring the values of HorizontalNet, VerticalNet, and its previous layer as inputs.

The computation between network columns is achieved via a lateral connection, whose parameters $\omega^{c'}$ and $\omega^{v'}$ are lateral weight parameters. Thus, the lateral parameters $\omega^{c'}$ and $\omega^{v'}$ determine the amount of activation of the globally learned common feature information and vertical feature information within the client, respectively. Its internal parameter ω^p indicates the internal weight parameter used to mix the information from both HorizontalNet and VerticalNet to learn more complex information to achieve the inference tasks of individual clients. The hidden layers h_l^p in the PPFL column are computed as follows:

$$h_{l+1}^p = \sigma\left(\omega_l^{c'} h_l^c + \omega_l^{v'} h_l^v + \omega_l^p h_l^p + b_l^p\right), \text{ where } h_0^p = 0. \quad \text{(Equation 4)}$$

The proposed method can be applied even in the absence of client-specific vertical features. In this case, the hidden layer of the PPFL is expressed as

$$h_{l+1}^p = \sigma\left(\omega_l^{c'} h_l^c + \omega_l^p h_l^p + b_l^p\right), \text{ where } h_0^p = 0. \quad \text{(Equation 5)}$$

Algorithm 1. Learning procedure of horizontal federated model

Input: Dataset $D_k^{common} := \{(c_i^k, y_i^k)\}_{i=1}^{m^{(k)}}$, *where client* $k \in \mathcal{K} := \{1, \ldots, K\}$; B is the local mini-batch size, E is the number of local epochs, and η is the learning rate.

Output: Horizontal federated model C and its weight parameter $\boldsymbol{\omega}^c$

Central server execute:
Construct the horizontal federated model C and initialize its weight parameter $\boldsymbol{\omega}^c$
for each round t = 0, 1, 2 …, N **do**
  Randomly set the S_t from the clients with the number of $m \leftarrow \max(S \cdot K, 1)$, where $0 < S \leq 1$
  for each client $k \in S_t$ **in parallel do**
    $\boldsymbol{\omega}_{t+1}^k \leftarrow ClientCommonUpdate(k, \omega_t^k; x_i^k)$
  end for
  $\boldsymbol{\omega}_{t+1}^c \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} \omega_{t+1}^k$
end for
return ω^c to all clients
def $ClientCommonUpdate(k, \omega^k; x_i, y_i)$: *//Run on client k*
$\mathcal{B} \leftarrow$ (split D_k^{common} into batches of size B)
  for each local epoch e from 1 to E **do**
    for batch $b \in \mathcal{B}$ **do**
    $\boldsymbol{\omega}^{k+1} \leftarrow SGD(\omega^k, \ell, \eta^p; b)$
    end for
end for
return $\boldsymbol{\omega}_{t+1}^k$ to central server

The final predictors for a specific client with or without client-specific vertical features are learned as follows:

$$\hat{y} = \sigma\left(\boldsymbol{\omega}_{l=L}^{c'} h_{=L}^{c} + \boldsymbol{\omega}_{l=L}^{v'} h_{l=L}^{v} + \boldsymbol{\omega}_{l=L}^{p} h_{l=L}^{p} + b_{l=L}^{p}\right), \quad \text{(Equation 6)}$$

$$\hat{y} = \sigma\left(\boldsymbol{\omega}_{l=L}^{c'} h_{=L}^{c} + \boldsymbol{\omega}_{l=L}^{p} h_{l=L}^{p} + b_{l=L}^{p}\right). \quad \text{(Equation 7)}$$

These outputs are the results of the mapping function of PPFL $f_k(\cdot)$ on a specific client k. Note that the notation L denotes the last layer of the PPFL network. The objective function for the personalization based on a client k is defined as

$$\min_{\boldsymbol{\omega}^{c'}, \boldsymbol{\omega}^{s}, \boldsymbol{\omega}^{s'}, \boldsymbol{\omega}^{p}} \mathcal{L}_k(\boldsymbol{\omega}^c, \boldsymbol{\omega}^{c'}, \boldsymbol{\omega}^s, \boldsymbol{\omega}^{s'}, \boldsymbol{\omega}^p) = \sum_{\boldsymbol{c}_i, \boldsymbol{s}_i, y_i \in D_k} l_i^k(f_k(\boldsymbol{c}_i, \boldsymbol{s}_i), y_i). \quad \text{(Equation 8)}$$

This objective function can be solved using stochastic gradient descent algorithms. In Equation 8, $\boldsymbol{\omega}^c$ is excluded from the optimization because it is "frozen."

The processes for building the horizontal federated and PPFL models are presented in Algorithms 1 and 2, respectively. In Algorithm 1, the inputs include common feature vector from the participating clients and target variables. As the output of Algorithm 1, the horizontal federated model can be learned using common feature information from the participating clients in the FL. The outputs of Algorithm 1 and the dataset, including common features, vertical features, and target variables from the participating clients, are the inputs of Algorithm 2. Subsequently, the PPFL model, which comprises HorizontalNet and VerticalNet, is generated for each client. The HorizontalNet section of the PPFL model $\boldsymbol{\omega}^c$ is initialized using the weight parameter $\boldsymbol{\omega}^c$ from the horizontal federated model. The inputs are common feature and client-specific vertical feature vectors from the individual client and target variables. The weight parameters $\boldsymbol{\omega}^c$ of the HorizontalNet column are frozen to retain globally learned knowledge related to common features, where $\boldsymbol{\omega}^v$ is the internal weight parameter of VerticalNet for client-specific vertical features. The lateral weight parameters $\boldsymbol{\omega}^{c'}$ and $\boldsymbol{\omega}^{v'}$ transmit knowledge of layer-wise network columns, which are HorizontalNet and VerticalNet, respectively. The PPFL weight parameter $\boldsymbol{\omega}^p$ allows the learning of more complex information from the PersonalizedNet layer $\boldsymbol{h}_l^p$, which receives the values of $\boldsymbol{h}_l^c$ and $\boldsymbol{h}_l^v$. As an output for Algorithm 2, these parameters can be learned using optimization methods such as gradient descent optimization algorithms.[33] If no client-specific vertical features exist, the proposed model can be personalized using this process, excluding for the VerticalNet column.

Datasets for PPFL experiment in healthcare

The PPFL model was evaluated using a distributed ICU dataset obtained from three types of ICUs located in 208 institutions under the eICU[34] and from the HICU in Severance Hospital located in Seoul, South Korea. A total of 5,266 patients admitted to the HICU in Severance Hospital and 14,550 and 10,664 patients admitted to the MICU and SICU, respectively, in the eICU (208 hospitals) were selected for development and internal assessment. For external validation, 12,706 patients were selected from the NSICU of the eICU dataset. As shown in Table S3, we

Algorithm 2. Learning procedure of personalized progressive federated learning model

Input: Dataset $D_k := \{(\boldsymbol{c}_i^k, \boldsymbol{s}_i^k, y_i^k)\}_{i=1}^{m^{(k)}}$, *where client* $k \in \mathcal{K} := \{1, \ldots, K\}$; B^p is the local mini-batch size for personalization, E^p is the number of epochs for personalization, and η^p is the learning rate for personalization

Output: PPFL model $\boldsymbol{f}_k(\cdot)$

Client execute: // *Run on specific client* k
Receive the ω^c from central server
$\mathcal{B} \leftarrow$ (split D_k into batches of size B^p)
if client k has client-specific vertical feature s^k **then**
  $f_k \leftarrow$ PersonalizedVertical($c^k, \mathbf{s}^k$)
else
  $f_k \leftarrow$ PersonalizedCommon(c^k)
end if
def PersonalizedVertical(c^k, s^k)
Construct the PPFL model
$\boldsymbol{f}_k \leftarrow f(\boldsymbol{\omega}^c, \boldsymbol{\omega}^{c'}, \boldsymbol{\omega}^v, \boldsymbol{\omega}^{v'}, \boldsymbol{\omega}^p; \boldsymbol{c}^k, \boldsymbol{s}^k)$
initialize $\boldsymbol{\omega}^c$ to output of Algorithm 1 and freeze the training of $\boldsymbol{\omega}^c$
initialize $\boldsymbol{\omega}^{c'}, \boldsymbol{\omega}^v, \boldsymbol{\omega}^{v'}, \boldsymbol{\omega}^p$
for each personalization epoch e from 1 to E^p **do**
  for batch $b^p \in \mathcal{B}$ **do**
    $(\boldsymbol{\omega}_{e+1}^{c'}, \boldsymbol{\omega}_{e+1}^v, \boldsymbol{\omega}_{e+1}^{v'}, \boldsymbol{\omega}_{e+1}^p) \leftarrow SGD((\boldsymbol{\omega}_e^{c'}, \boldsymbol{\omega}_e^v, \boldsymbol{\omega}_e^{v'}, \boldsymbol{\omega}_e^p), \ell, \eta^p; b^p)$
  end for
end for
return the PPFL model f_k
def PersonalizedCommon(c^k)
Construct the PPFL model
$\boldsymbol{f}_k \leftarrow f(\boldsymbol{\omega}^c, \boldsymbol{\omega}^{c'}, \boldsymbol{\omega}^p; \boldsymbol{c}^k, \boldsymbol{s}^k)$
initialize $\boldsymbol{\omega}^c$ to output of Algorithm 1 and freeze the training of $\boldsymbol{\omega}^c$
initialize $\boldsymbol{\omega}^{c'}, \boldsymbol{\omega}^p$
for each personalization epoch e from 1 to E^p **do**
  for batch $b^p \in \mathcal{B}$ **do**
    $(\boldsymbol{\omega}_{e+1}^{c'}, \boldsymbol{\omega}_{e+1}^p) \leftarrow SGD((\boldsymbol{\omega}_e^{c'}, \boldsymbol{\omega}_e^p), \ell, \eta^p; b^p)$
  end for
end for
return the PPFL model f_k

identified 12 common features for each ICU; moreover, 23, 13, 7, and 8 different client-specific features of the HICU, MICU, SICU, and NSICU, respectively, were selected for each client using a feature selection method that utilizes linear models with an L1 penalty (L1-norm) added to the loss function. By fitting the linear model, the coefficients of certain features became zero, which exhibited a feature selection effect.[35]

For external validation, we evaluated the proposed model using non-small cell lung cancer datasets from two different studies (TCGA and MSK) in cBioPortal, a data portal for cancer genomics. A total of 2,621 and 11,44 records from the years 2008, 2015, 2013, and 2020 were included in the TCGA and MSK datasets, respectively. As shown in Table S4, we identified 8 common features and 13 and 8 client-specific features from TCGA and MSK, respectively.

Comparing the performance with conventional algorithms

To evaluate the performance of the PPFL algorithm, we compared the combinations of input data from eICU[34] and Severance Hospital. The MICU and SICU data from the eICU dataset and HICU data from the Severance Hospital dataset were utilized to develop the PPFL and validate it internally, whereas the NSICU data from the eICU dataset were used for external assessment. Details of the data are provided in the dataset section. We compared PPFL with the models described below. Here, (*c*) indicates that the model has learned only the common feature space, whereas (*c, s*) indicates that the model has learned both common and client-specific vertical features.

- **FedAvg** (*c*)**:** HorizontalNet learned using the FedAvg algorithm with common features.
- **FedProx** (*c*)**:** HorizontalNet learned using FedProx with common features by introducing an L_2 regularization term to the FedAvg objective function.
- **FedRep** (*c*): HorizontalNet learned using the FedRep algorithm with common features. The algorithm focused on extracting representations specifically from the personalized layers.
- **FedPer** (*c*): HorizontalNet learned using the FedPer algorithm with common features. The algorithm employed shallow base layers for high-level representation extraction and deep personalization layers for classification.

- **PPFL (*c*):** The PPFL model learned from data of individual clients by leveraging only common features.
- **PPFL (*c, s*):** The PPFL model learned from data of individual clients by leveraging both common and client-specific vertical features.
- **Local (*c*):** Multi-layer perceptron (MLP) models learned only from the common feature data of a specific client.
- **Local (*c, s*):** MLP models learned from both common and vertical feature data of a specific client.

Each client was selected independently. For each client, we compared the performances of FedAvg (c), FedProx (c), FedRep (c), FedPER (c), PPFL (c), PPFL (c, s), Local (c), and Local (c, s) for internal and external validations.

QUANTIFICATION AND STATISTICAL ANALYSIS

For the descriptive characteristics of the study, mean (SD) for continuous variables and frequency (percentages) for categorical variables were presented. Comparison tests were performed with analysis of variance for continuous variables and chi-square tests for categorical variables at 5% significance levels. To demonstrate the performance improvement for individual clients and robustness of the unseen distribution for the proposed model, we determined the accuracy and AUROC score for each ICU client. Other metrics, including sensitivity, specificity, positive predictive value, and negative predictive value at the Youden index, have also been reported. To evaluate the communication cost, we measured the AUROC over one epoch for each algorithm.

The AUROC curve (AUC) was computed to evaluate the model performance. DeLong's test was applied to demonstrate the statistical difference between the AUC values of our model and those of the comparative algorithms. Python and R programming software were used to compute the performance evaluation metrics and perform statistical analyses. Several publicly available packages were used in this study, including scikit-learn, SciPy, Matplotlib, NumPy, and Pandas for Python, and pROC for R.

Contribution of common and specific features of each client

To investigate the conceptual shift after the application of PPFL, we computed the feature importance using the SHAP value computed using Deep SHAP.[36] SHAP is a method for computing the contribution value of a data instance to explain the outcome. To calculate the contribution value for a specific feature, we considered all possible permutations of the features and computed the average contribution of the feature coalitions. Deep SHAP approximates the SHAP value using Deep Learning Important Features.[37]

ADDITIONAL RESOURCE

All experimental settings were implemented using TensorFlow 2.5.0.[38] The models were trained on a machine equipped with two NVIDIA QUADRO RTX 8000 CUDA 11.0 processors with 128 GB memory and one Intel Xeon Platinum 8253 2.2GHz CPU.