Personalized Progressive Federated Learning with Leveraging Client-Specific Vertical Features

Tae Hyun Kim, Won Seok Jang, Sun Cheol Heo,
MinDong Sung, and Yu Rang Park

Department of Biomedical Systems Informatics,
Yonsei University College of Medicine, Seoul, South Korea

Abstract

Federated learning (FL) has been used for model building across distributed clients. However, FL cannot leverage vertically partitioned features to increase the model complexity. In this study, we proposed a personalized progressive federated learning (PPFL) model, which is a multi-model PFL approach that allows the leveraging of vertically partitioned client-specific features. The performance of PPFL was evaluated using the Physionet Challenges 2012 dataset. We compared the performance of in-hospital mortality and length of stay prediction between our model and the FedAvg, FedProx, and local models. The PPFL showed an accuracy of 0.849 and AUROC of 0.790 in average in hospital mor-tality prediction, which are the highest scores compared to client-specific algorithm. For length-of-stay prediction, PPFL also showed an AUROC of 0.808 in average which was the highest among all comparators.

Keywords

Personalized Federated Learning, Vertical Federated Learning, Non-IID data

1. Introduction

Federated learning (FL) is a collaborative machine-learning approach used for solving data problems, such as data leakage, while preserving privacy in distributed environments [1–3]. Despite the numerous advantages of FL, such as privacy preservation, fulfillment of data requirements, and communication efficacy, it is still limited regarding the availability of information from conventional FL designs. FL designs (e.g., horizontal federated learning (HFL) and vertical federated learning (VFL)) can be categorized based on the data distribution among various parties (i.e., whether data are distributed based on the feature space or sample-ID space) [2]. HFL [3–9] can analyze large volumes of data using “identical feature spaces” from multiple clients. VFL [10-11] can be built from distributed feature spaces using only “identical sample IDs” across different clients.

However, in an HFL scenario, some clients might have specific feature information that is generated only within specific clients or is not allowed in a federated manner because of critical privacy concerns. For instance, there may be differences in the features collected among hospitals participating in federated learning, and these client-specific features may be excluded from the HFL scenario. Under a real-world VFL scenario, it is difficult for distributed clients to obtain sufficient identical samples to build a machine-learning model. These issues may degrade the performance of the model.

David C. Wyld et al. (Eds): CSML, NET, BDHI, SIPO, SOEA- 2023
pp. 65-77, 2023. CS & IT- CSCP 2023 DOI:10.5121/csit.2023.130106

In contrast, the main challenge for FL is the distributed setting of data heterogeneity and non-independent and identically distributed (non-IID) data from clients [12]. Previous studies [13, 14] have demonstrated that a FL model with a FedAvg [3] algorithm might perform poorly using statistical data heterogeneity, which slows down FL convergence.

The limitations of FL designs and data heterogeneity have motivated the development of a new approach to overcome both problems. In real-world situations, client-specific vertical features can be ignored in an HFL design, whereas identical sample IDs are insufficient in a VFL design, and data heterogeneity degrades performance. Therefore, we focused on leveraging client-specific vertical features while implementing a model that is well adapted to the heterogeneity of data across clients in a cross-silo environment.

In this study, we propose a novel approach called personalized progressive federated learning (PPFL) combining FL with variants of progressive neural networks [15]. In PPFL, building a personalized model allows the learning of client-specific distributions from a globally learned FL model by transmitting layer-wise knowledge to different network columns. The proposed model learns global knowledge from common feature information and expands the feature space related to client-specific vertical features by creating new column networks.

We applied the lateral connection in a progressive neural network [15] to expand the layer-wise feature space from a globally pre-trained FL model. Additionally, a pro-gressive neural network was proposed to address the forgetting problem [15,16]. Therefore, our model prevents the forgetting of previously learned global knowledge during the personalization phase. In this study, we experiment and validate the algorithm with real-world medical data.

2. RELATED WORKS

2.1. Federated Learning on Non-IID Data

FL is a machine-learning approach in which multiple clients collaboratively build a learning task while considering privacy issues and communication efficacy [3]. FL can be classified into HFL and VFL, depending on how the data are distributed among various clients [2]. HFL deals with a scenario in which each client has an identical feature space but different sample-ID spaces. FedAvg [3] is a collaborative machine-learning framework proposed for this HFL scenario. HFL approaches cannot utilize vertically partitioned features, which are specifically generated by individual clients and are not shared with the HFL frameworks, increasing the model complexity.

VFL deals with a scenario in which each client has a different feature space and identical sample ID space. Although secured machine-learning methods [10,32–35] for distributed features have been proposed, such methods cannot be used as deep learning approaches. In addition, despite the proposal of VFL approaches for deep learning [11,36,37], these methods have a limitation, in which every client must learn sufficient “identical sample-IDs” using a deep learning model.

2.2. Federated Learning on Non-IID Data

Data heterogeneity and non-IID data complicate the construction of a global FL model that can be applied to individual clients. FedAvg demonstrates a reduced model performance, including accuracy, under statistical data heterogeneity [14]. Additionally, the heterogeneity of the data slows down and destabilizes the convergence of FedAvg [13].

Previous studies [14,30,38,39] have focused on utilizing the data augmentation method in an FL manner to address the weight divergence on non-IID data during the FL process. This method has been proposed to smoothen the statistical heterogeneity across distributed clients. However, when data augmentation approaches FL, it suffers from privacy leakage because data sharing has not been eliminated. Client selection approaches, such as FAVOR [29] used to build the FL model from the more homogeneous data distributions, also exist.

Previous studies [31,40–45] proposed a personalized globally trained FL model for heterogeneous clients. Meta-learning-based approaches, such as personalized federated average (Per-FedAvg) [31], have been proposed to personalize an FL model by finding an optimal initialization for local personalization and learning of task-specific local representations based on a single global model design through meta-learning [40]. Multi-model personalization based on hierarchical clustering [41] was used to train an FL model for each cluster of clients. This framework involves training clusters of clients during each round of FL training. PFL approaches based on multi-task learning, model interpolation, and transfer learning build a model for each individual client through the FL process. The MOCHA algorithm was proposed as a personalization method for combining distributed multi-task learning and FL [42]. The model interpolation method [43] was proposed to handle the trade-off between a globally learned model and locally learned models with an adjustable penalty parameter. Transfer-learning-based approaches [44,45] aim to transfer the globally trained knowledge to the local models of individual clients through fine-tuning.

3. METHOD

We proposed a PPFL algorithm for conducting client-specific personalized inferences on data heterogeneity and non-IID data settings. PPFL also addresses the limited in-formation availability of FL design by leveraging not only common features but also client-specific vertical features across distributed clients. The proposed process involves two major steps. First, we built a HFL on a central server using only the common features from the distributed clients. Second, the pre-trained horizontal federated model was deployed for each client, learning personalized knowledge for client-specific inference task through a PPFL.

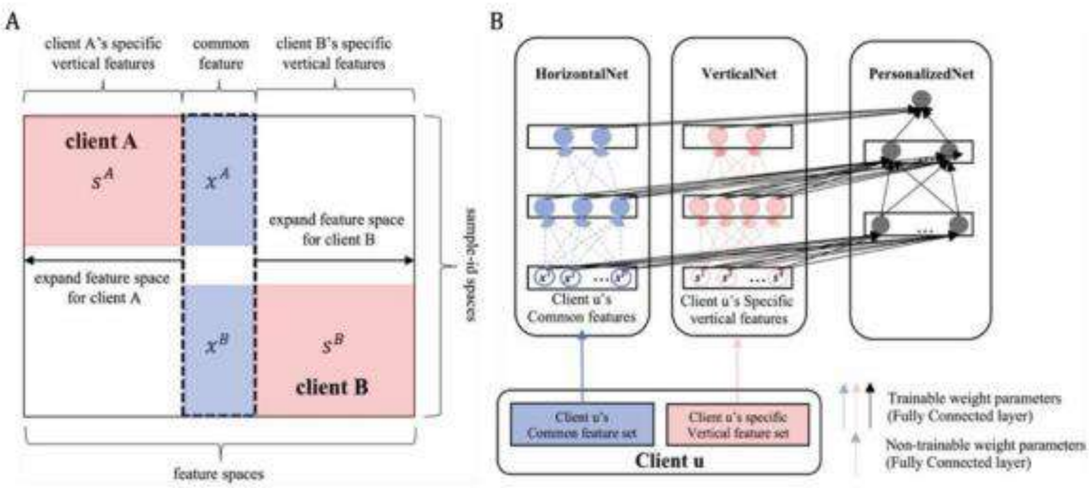

Figure 1: Problem setting and network architecture of the personalized progressive federated model.

3.1. Problem Formulation

This study aims to solve the case where the features of each client are common and client-specific cases exist (Figure 1 A). Before distinguishing common or client-specific vertical features of each client, all feature information should be shared among clients. Suppose that an individual client k has a dataset $D_k \coloneqq \{x_i^k, s_i^k, y_i^k\}_{i=1}^{m^{(k)}}$ consisting of $m^{(k)}$ samples, where the client $\in \mathcal{K} \coloneqq \{1, \dots, K\}$. The i-th sample of D_k can be represented using a common feature vector with p-dimension $\boldsymbol{x}_i^k \coloneqq \{x_i^{1(k)}, x_i^{2(k)}, \dots, x_i^{p(k)}\}$; the client's specific vertical feature vector with q-dimension $\boldsymbol{s}_i^k \coloneqq \{s_i^{1(k)}, s_i^{2(k)}, \dots, s_i^{q(k)}\}$, and the corresponding target variable y_i^k. Note that the attributes and dimension $p^{(k)}$of the common feature vector $\boldsymbol{x}_i^k$ are identical for all clients $k \in \mathcal{K}$. However, the attributes and dimension $q^{(k)}$ of the client's specific vertical feature vector $\boldsymbol{s}_i^k$ may not be the same for all clients.

3.2. Horizontal Federated Learning

A horizontal federated model learns global knowledge related to common features across multiple clients in a federated manner. The proposed model PPFL is generic and can be applied to other deep-learning-based approaches and aggregated methods. However, in this study, we applied our algorithm to the FedAvg as a base method for building a HFL because it is the most well-known and commonly used method.

m is the total sample size of K clients. then, $f_i(\boldsymbol{\omega})$ is the loss function of the prediction on example $(\boldsymbol{x}_i, y_i)$ where $\boldsymbol{x}_i$ is common feature vector. Therefore, the objective function is

$$\min_{\boldsymbol{\omega}^c \in \mathbb{R}^d} F(\boldsymbol{\omega}^c) \coloneqq \sum_{k=1}^{K} \frac{m_k}{m} F_k(\boldsymbol{\omega}^k), \qquad (1)$$
$$where \quad F_k(\boldsymbol{\omega}^k) \coloneqq \sum_{x_i^k \in D_k} f_i(\boldsymbol{\omega}^k)$$

3.3. Personalized Progressive Federated Learning

PPFL contains three network columns: HorizontalNet, VerticalNet, and Personalized Net. We utilized the concept of lateral connection in progressive neural networks [15], which is proposed for leveraging transfer and avoiding catastrophic forgetting in multi-task learning. Figure 1 B shows the architecture of the PPFL model.

3.3.1. Horizontal Network

HorizontalNet is a network column that is initialized from the horizontal federated model. The internal weight parameters of the HorizontalNet $\boldsymbol{\omega}^{c^{int}}$ were initialized using $\boldsymbol{\omega}^{\boldsymbol{c}}$described in the Section 2.2. This network aims to pass generalized knowledge to personalized networks with the common feature $\boldsymbol{x}^{\boldsymbol{k}}$ as input information. Note that the internal weight matrix $\boldsymbol{\omega}^{c^{int}}$ in HorizontalNet, which is not connected with PersonalizedNet, is "frozen" to train. However, the lateral weight parameter $\boldsymbol{\omega}^{c^{lat}}$, which is connected with PersonalizedNet, can be updated using an optimization algorithm. This approach avoids forgetting the generalized knowledge that has already been learned. The hidden layers $\boldsymbol{h}_l^c$in the HorizontalNet column are computed as

$$\boldsymbol{h}_{l+1}^c = \sigma\left(\boldsymbol{\omega}_l^{c^{int}} \boldsymbol{h}_l^c + \boldsymbol{b}_l^c\right), \; where \qquad \boldsymbol{h}_0^c = \boldsymbol{x}^k. \qquad (2)$$

3.3.2. Vertical Network

The second network column was the VerticalNet column. This network expanded the feature space with respect to the client-specific vertical features. The input of VerticalNet is the specific vertical feature data of the client $\boldsymbol{s}^k \in D_k$. The weight parameter $\boldsymbol{\omega}^{v^{int}}$ is the internal weight parameter of VerticalNet, which is not connected to PersonalizedNet. The lateral weight parameter $\boldsymbol{\omega}^{v^{lat}}$ is connected to PersonalizedNet. Both $\boldsymbol{\omega}^{v^{int}}$ and $\boldsymbol{\omega}^{v^{lat}}$ can be learned through the training step. Thus, the parameter $\boldsymbol{\omega}^{v^{int}}$ and $\boldsymbol{\omega}^{v^{lat}}$ learn client-specific vertical feature information and transmit their knowledge to PersonalizedNet. The hidden layers $\boldsymbol{h}_l^v$ with respect to the client-specific vertical feature $\boldsymbol{s}^k$ and internal weight parameter $\boldsymbol{\omega}^{v^{int}}$ are

$$\boldsymbol{h}_{l+1}^v = \sigma\left(\boldsymbol{\omega}_l^{v^{int}}\boldsymbol{h}_l^v + \boldsymbol{b}_l^v\right), \qquad where\ \ \boldsymbol{h}_0^v = \boldsymbol{s}^k \qquad (3)$$

3.3.3. Personalized Network

The Personalized Net learn the specific personalized knowledge of the client by acquiring the value of $\boldsymbol{h}_l^c$, $\boldsymbol{h}_l^v$, and its previous layer as inputs. The computation between network columns is made possible through a lateral connection, the parameters of which, $\boldsymbol{\omega}^{c^{lat}}$ and $\boldsymbol{\omega}^{v^{lat}}$, are lateral weight parameters. Therefore, $\boldsymbol{\omega}^{c^{lat}}$ and $\boldsymbol{\omega}^{v^{lat}}$ determine the amount of activation of the globally learned common feature information and vertical feature information within the client, respectively. Its internal parameters $\boldsymbol{\omega}^p$ are the internal weight parameters learn more complex information to achieve the inference tasks of individual clients. The hidden layers $\boldsymbol{h}_l^p$ are computed using Equation (4).

$$\boldsymbol{h}_{l+1}^p = \sigma\left(\boldsymbol{\omega}_{l+1}^{c^{lat}}\boldsymbol{h}_{l+1}^c + \boldsymbol{\omega}_{l+1}^{v^{lat}}\boldsymbol{h}_{l+1}^v + \boldsymbol{\omega}_{l+1}^p\boldsymbol{h}_l^p + \boldsymbol{b}_l^p\right), where\ \ \boldsymbol{h}_0^p = \boldsymbol{0} \qquad (4)$$

Note that the proposed method can be applied even in the absence of client-specific vertical features. In this case, the hidden layer of a personalized progressive network is expressed as

$$\boldsymbol{h}_{l+1}^p = \sigma\left(\boldsymbol{\omega}_{l+1}^{c^{lat}}\boldsymbol{h}_{l+1}^c + \boldsymbol{\omega}_{l+1}^p\boldsymbol{h}_l^p + \boldsymbol{b}_l^p\right), \qquad where\ \ \boldsymbol{h}_0^p = 0 \quad (5)$$

In this process, if there is no client-specific vertical features, it can be personalized except the VerticalNet.

3.4. Study Design

We compared PPFL with the models described below. (x) indicates that the model has learned only the common feature space, and (x,s) indicates the model has learned both common features and client-specific vertical features.

- **FedAvg(*x*): The** FedAvg algorithm with common features.
- **FedProx(*x*): The** FedProx algorithm with common features.
- **PPFL(*x*):** The PPFL learns by leveraging only common features.
- **PPFL(*x, s*):** The PPFL learns by leveraging both common features and client-specific vertical features.
- **Local(*x*):** Multi-layer perceptron (MLP) learned only from common feature data of a specific client.
- **Local(*x, s*):** MLP learned from both common and vertical feature data of a specific client.

We divided the training, validation, and test datasets in the ratio of 6:2:2 for each client. The validation dataset was used to search for hyper parameters using a random-search algorithm. We optimized the weight parameters of the models by the Adam optimizer [21]. We utilized the cross-entropy loss for the binary classification. We implemented them while providing accuracy and an area under the receiver operating characteristic (AUROC) to demonstrate the performance.

3.5. Dataset

The performance of the PPFL model was evaluated on a public EMR dataset called Physionet Challenge 2012 [19]. The Physionet Challenge 2012, which was extracted from the MIMIC-II database [22], consists of information regarding 8,000 ICU patients. These records contained 36 time-series features (i.e., laboratory tests, vital signs, and mechanical ventilation) and five demographic features, including ICU-type information. In this study, we aggregated ICU information for 48 h in an average manner because we did not focus on time-series data. Each ICU, with a total of 6,000 samples, was considered an individual client. Coronary care unit (CCU), cardiac surgery recovery unit (CSRU), medical ICU (MICU), and surgical ICU (SICU) retained 889, 1,219, 2,216, 1,676, and 2,000 ICU stay samples, respectively. The remaining 2,000 samples were used as external ICUs, configured without client separation. The external ICU was not used during the PPFL training. In this dataset, we assumed that the common feature set comprised demographic and mechanical ventilation information. In contrast, client-specific vertical features comprised vital signs and laboratory tests for all clients. The description of data distribution by the ICU for common features of the Physionet Challenge 2012 data set is presented in Supplementary Table 1 and 2.

3.6. Experiments

For each client, we compared the performance for both internal and external validations. Internal performance was measured using a test set from a local client. For external validation, we used external dataset that were set aside when partitioning ICU data. We evaluated the performance of binary classifications for the following two cases: in-hospital mortality as a binary class and length of stay.

We computed feature importance using the SHAP value computed by Deep SHAP to investigate the concept shift after the application of PPFL [24, 25].

All experimental settings were implemented using TensorFlow 2.5.0 [26]. The models were trained on a machine equipped with two NVIDIA QUADRO RTX 8000 CUDA 11.0, 128 GB memory and one Intel Xeon Platinum 8253 2.2 GHz CPU.

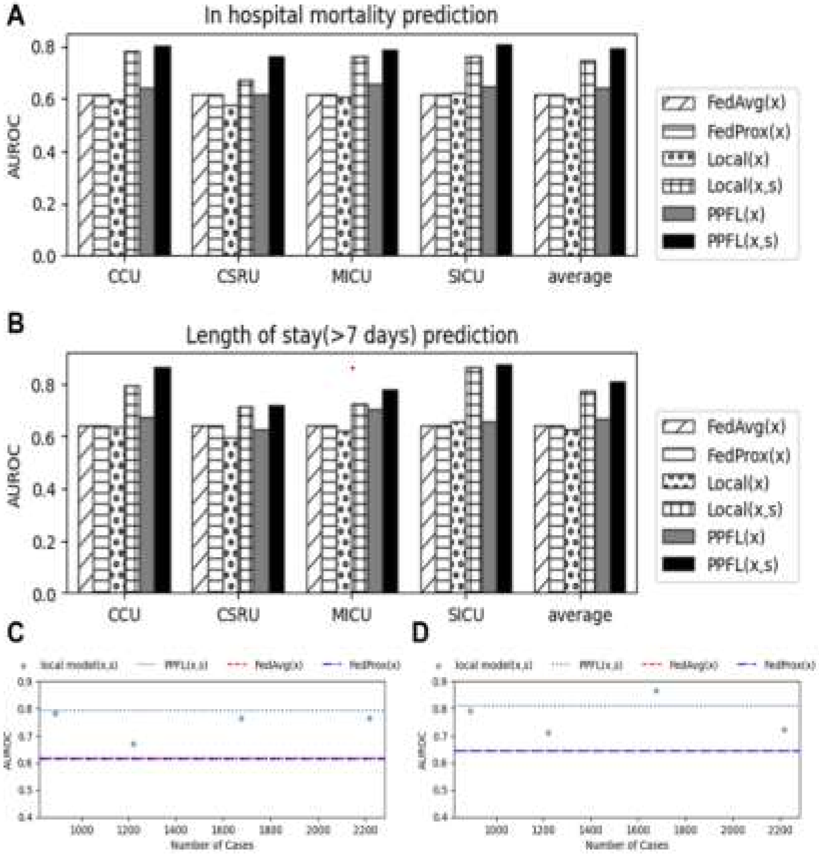

Figure 2. Performance evaluation of PPFL compared to FedAvg (x), FedProx (x), Local (x), and Local (x,s) in terms of AUROC on external validation. PPFL (x,s) shows the highest score in every task A. AUROC comparison for in-hospital mortality prediction task. B. AUROC score comparison for the length of stay prediction task. C. AUROC score

4. RESULTS

4.1. Performance of PPFL

PPFL(x,s) showed the highest performance for every ICU client on external validation. The PPFL(x) showed an average of 0.790 AUROC for the in-hospital mortality task and 0.808 AUROC for the length of the stay task (Figures 2A and 3B, respectively). Where FedAvg(x) and FedProx(x) showed performance (AUROC) by 0.616 and 0.615 in mortality prediction, respectively. In addition, PPFL(x,s) higher performance than FedAvg(x) and FedProx(x) both in hospital mortality and length of stay prediction. The average AUROC of FedAvg(x) was 0.643 and 0.643 for FedProx in length of stay prediction. Compared with Local(x,s), PPFL(x,s) show that all AUROC performances of PPFL(x,s) outperform in external validations. The average AUROC for local(x,s) in external validation was 0.743 in in hospital mortality prediction, and 0.773 in length of stay prediction. In average, PPFL(x,s) showed higher performance than local(x,s) models in external validation (Figure 2A, Figure 2B, Supplementary Table 3). Comparing the average AUROC of PPFL(x,s) to Local(x,s) in Figure 2C and Figure 2D, our model showed higher performance in in hospital mortality task. However, in length of stay prediction, the SICU showed 0.865, which was higher than the average AUROC performance than PPFL(x,s). Overall, PPFL(x,s) showed the highest AUROC compared to other local model(x,s) in average (Figure 2C, Figure 2D). Figures 4 shows the contributions of common and vertical features for all clients in predicting in-hospital mortality. Within common features, age and mechanical ventilation (MechVent) features had the highest shape value in all clients (age was 0.5 or more in all clients and MechVent was 0.3 or more in three clients). Among the verticalfeatures, the Glasgow Coma Scale (GCS) had the highest shape value for all clients (0.025 orhigher for all clients). Mechanical ventilation still had a high ranking for CCU and SICU. We also compared FedAvg(x) to PPFL(x,s) to evaluate whether leveraging client specific features shows high performance. PPFL(x,s) showed higher performance than FedAvg (x) (Supplementary Table 4). For the MICU, the SHAP value for MechVent was not lower than those of the other clients. However, in terms of vertical features, vital signs, such as GCS, blood urea nitrogen, fraction of inspired oxygen, heart rate, and absolute blood pressure, have higher SHAP values than those for mechanical ventilation.

4.2. Analysis of Concept Drift

Figure 3 shows the contributions of common and vertical features for all clients in predicting in-hospital mortality. Within common features, age and mechanical ventilation (MechVent) features had the highest shape value in all clients, in that order (age was 0.5 or more in all clients and MechVent was 0.3 or more in 3 clients). Among the vertical features, the Glasgow Coma Scale (GCS) had the highest shape value for all clients (0.025 or higher for all clients). Mechanical ventilation still had a high ranking for CCU and SICU.

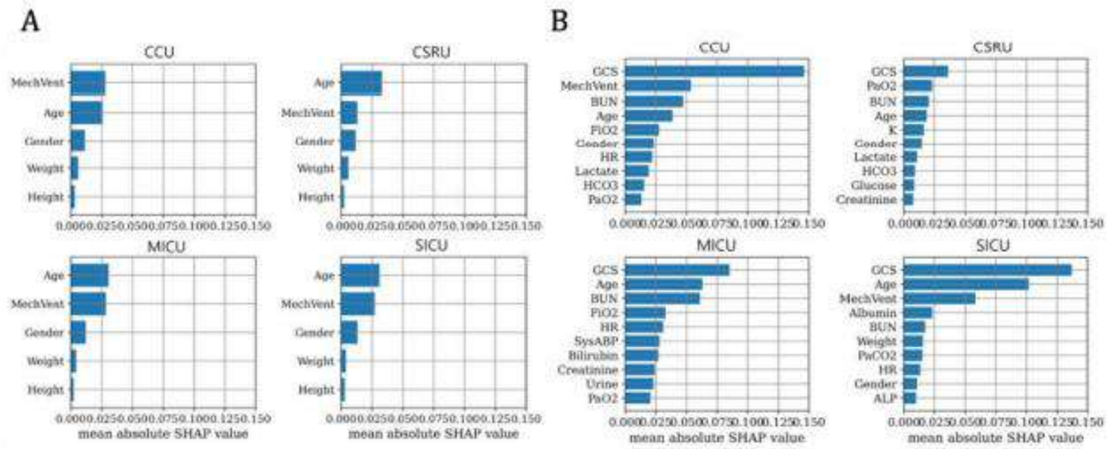

Figure 3.Mean absolute SHAP values of common and vertical features in predicting in-hospital mortality. 4A. SHAP values in common features. 4B. top 10 highest SHAP value features with vertical features.

5. DISCUSSION

The usage of federated learning in analyzing distributed medical data is a well-known research topic [17,27]. Therefore, research on federated learning that can potentially protect data privacy has been conducted in various medical fields [28]. However, most current studies consider learning common features among clients. In this study, we proposed a personalized progressive federated learning (PPFL) algorithm for heterogeneously distributed clients that expands the feature space for client-specific vertical features. This study is the first federated learning study that considers common features and client-specific vertical features by applying progressive learning. PPFL shows a robust performance compared to other algorithms based on the comparison of PPFL with existing federated learning models and local models in various settings.

Compared to FedAvg, which is suitable for a horizontally partitioned data environment [3–9], PPFL is a novel federated learning framework that leverages the idea of progressive learning to perform learning in both horizontally and vertically partitioned environments. PPFL can utilize more features and samples than other models (Figure 2, Supplementary Table 3), resulting in better performance compared to existing local and federated learning models. For example, FedAvg and FedProx have a limited feature space because only the common features from multiple clients are input into the model in terms of its structure. The local model uses only the sample of each client; thus, the number of samples is inevitably smaller than that of the PPFL input dataset. PPFL demonstrated a higher performance than the existing model by inputting all the collected features and samples of multi-clients.

The effectiveness of the proposed model is the greatest for clients who are significantly different from the overall data distribution since CSRU has the most different label distribution from an external client and the most severe class imbalance.

For all internal validations of the clients, except for the in-hospital mortality task for some clients, HorizontalNet(x), learned through FedAvg, exhibits a degraded performance compared to that with Local(x). Previous studies have confirmed that FL performance may decrease when the distribution among clients is heterogeneous [13,14]. Additionally, the data we tested was statistically significant heterogeneous across clients (Supplementary Table 1). We found that the hospital stay of SICU patients was significantly longer than that of other ICU patients (Supplementary Table 1). Moreover, we found that the performance of the local(x) model using only local data was higher than our proposed ppfl(x,s) (Figure 2D). This indicates that extreme data heterogeneity in FL can lead to lower performance than that of local models. However, we emphasize that our model still outperforms FedAvg and FedProx, and the performance difference with the local model (SICU) is negligible.

Although client-specific vertical features contain more information, our proposed model is effective in terms of robustness. This shows that PPFL is robust to the global knowledge forgetting problem in the personalization process of the FL models.

6. LIMITATIONS

Our study has several limitations. First, there is little difference in the computing time and resources when verifying the PPFL in the same network bandwidth. However, additional research on the computation time and resources between physically distant networks is required for multi-client from multi-country studies. Second, this PPFL algorithm was written assuming that information on the features of multiple clients is shared; however, information about common and vertical features of each client may not be provided in the real world. Research on an automatic feature selection process based on the characteristics of input data among the features of multiple clients is essential. Third, Yang et al. (2019) reported that there is a possibility of indirect privacy leakage to raw federated learning systems [2]. We plan to further our studies in strengthening PPFL from these issues. Fourth, although only MLP modules based on linear layer have been applied to the PPFL framework in this study, we will also apply them to other neural network structures such as sequential-based layers in future studies.

7. CONCLUSION

We proposed the 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 EHR data and validated the usefulness of the model. Our model showed higher performance than FedAvg and FedProx. We plan to further our studies in improving the PPFL compared to other models in FL.

ACKNOWLEDGEMENTS

This research was supported by a grant from the Korea Health Technology RD Project through the Korea Health Industry Development Institute (KHIDI) funded by the Ministry of Health Welfare (HI19C1015), and the Bio-Industrial Technology Development Program (20014841) funded by the Ministry of Trade, Industry Energy (MOTIE, Korea), Republic of Korea.

REFERENCE

[1] Kairouz, P.; McMahan, H.B.; Avent, B.; Bellet, A.; Bennis, M.; Bhagoji, A.N.; Bonawitz, K.; Charles, Z.; Cormode, G.; Cummings, R.; et al. Advances and Open Problems in Federated Learning. 2019.

[2] Yang, Q.; Liu, Y.; Chen, T.; Tong, Y. Federated Machine Learning. ACM Trans Intell Syst Technol2019, 10, 1–19, doi:10.1145/3298981.

[3] McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; Arcas, B.A. y Communication-Efficient Learning of Deep Networks from Decentralized Data. In Proceedings of the Proceedings of the 20th International Conference on Artificial Intelligence and Statistics; Singh, A., Zhu, J., Eds.; PMLR, October 2017; Vol. 54, pp. 1273–1282.

[4] Konečný, J.; McMahan, H.B.; Yu, F.X.; Richtárik, P.; Suresh, A.T.; Bacon, D. Federated Learning: Strategies for Improving Communication Efficiency. 2016.

[5] Braverman, M.; Garg, A.; Ma, T.; Nguyen, H.L.; Woodruff, D.P. Communication Lower Bounds for Statistical Estimation Problems via a Distributed Data Processing Inequality. In Proceedings of the Proceedings of the forty-eighth annual ACM symposium on Theory of Computing; ACM: New York, NY, USA, June 19 2016; pp. 1011–1020.

[6] Haddadpour, F.; Kamani, M.M.; Mokhtari, A.; Mahdavi, M. Federated Learning with Compression: Unified Analysis and Sharp Guarantees. 2020.

[7] Lin, Y.; Han, S.; Mao, H.; Wang, Y.; Dally, W.J. Deep Gradient Compression: Reducing the Communication Bandwidth for Distributed Training. 2017.

[8] Guha, N.; Talwalkar, A.; Smith, V. One-Shot Federated Learning. 2019.

[9] Zhu, H.; Jin, Y. Multi-Objective Evolutionary Federated Learning. IEEE Trans Neural Netw Learn Syst2020, 31, 1310–1322, doi:10.1109/TNNLS.2019.2919699.

[10] Hardy, S.; Henecka, W.; Ivey-Law, H.; Nock, R.; Patrini, G.; Smith, G.; Thorne, B. Private Federated Learning on Vertically Partitioned Data via Entity Resolution and Additively Homomorphic Encryption. 2017.

[11] Hu, Y.; Niu, D.; Yang, J.; Zhou, S. FEDML. In Proceedings of the Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; ACM: New York, NY, USA, July 25 2019; pp. 2232–2240.

[12] Darrell, T.; Kloft, M.; Pontil, M.; Rätsch, G.; Rodner, E. Machine Learning with Interdependent and Non-Identically Distributed Data (Dagstuhl Seminar 15152). In Proceedings of the Dagstuhl Reports; 2015; Vol. 5.

[13] Li, X.; Huang, K.; Yang, W.; Wang, S.; Zhang, Z. On the Convergence of FedAvg on Non-IID Data. 2019.

[14] Zhao, Y.; Li, M.; Lai, L.; Suda, N.; Civin, D.; Chandra, V. Federated Learning with Non-IID Data. 2018, doi:10.48550/arXiv.1806.00582.

[15] Rusu, A.A.; Rabinowitz, N.C.; Desjardins, G.; Soyer, H.; Kirkpatrick, J.; Kavukcuoglu, K.; Pascanu, R.; Hadsell, R. Progressive Neural Networks. 2016.

[16] Kirkpatrick, J.; Pascanu, R.; Rabinowitz, N.; Veness, J.; Desjardins, G.; Rusu, A.A.; Milan, K.; Quan,

J.; Ramalho, T.; Grabska-Barwinska, A.; et al. Overcoming Catastrophic Forgetting in Neural Networks. Proceedings of the National Academy of Sciences2017, 114, 3521–3526, doi:10.1073/pnas.1611835114.
[17] 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. The Future of Digital Health with Federated Learning. NPJ Digit Med2020, 3, 119, doi:10.1038/s41746-020-00323-1.
[18] Ruder, S. An Overview of Gradient Descent Optimization Algorithms. 2016.
[19] Silva, I.; Moody, G.; Scott, D.J.; Celi, L.A.; Mark, R.G. Predicting In-Hospital Mortality of ICU Patients: The PhysioNet/Computing in Cardiology Challenge 2012. In Proceedings of the 2012 Computing in Cardiology; 2012; pp. 245–248.
[20] Pollard, T.J.; Johnson, A.E.W.; Raffa, J.D.; Celi, L.A.; Mark, R.G.; Badawi, O. The EICU Collaborative Research Database, a Freely Available Multi-Center Database for Critical Care Research. Sci Data2018, 5, 180178, doi:10.1038/sdata.2018.178.
[21] Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. 2014.
[22] Saeed, M.; Lieu, C.; Raber, G.; Mark, R.G. MIMIC II: A Massive Temporal ICU Patient Database to Support Research in Intelligent Patient Monitoring. Comput Cardiol2002, 29, 641–644.
[23] Baraniuk, R. Compressive Sensing [Lecture Notes]. IEEE Signal Process Mag2007, 24, 118–121, doi:10.1109/MSP.2007.4286571.
[24] Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In Proceedings of the Advances in Neural Information Processing Systems; Guyon, I., Luxburg, U. von, Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc., 2017; Vol. 30.
[25] Shrikumar, A.; Greenside, P.; Shcherbina, A.; Kundaje, A. Not Just a Black Box: Learning Important Features Through Propagating Activation Differences. 2016.
[26] Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. 2016.
[27] Ciampi, M.; Sicuranza, M.; Silvestri, S. A Privacy-Preserving and Standard-Based Architecture for Secondary Use of Clinical Data. Information2022, 13, 87, doi:10.3390/info13020087.
[28] Kaissis, G.A.; Makowski, M.R.; Rückert, D.; Braren, R.F. Secure, Privacy-Preserving and Federated Machine Learning in Medical Imaging. Nat Mach Intell2020, 2, 305–311, doi:10.1038/s42256-020-0186-1.
[29] Wang, H.; Kaplan, Z.; Niu, D.; Li, B. Optimizing Federated Learning on Non-IID Data with Reinforcement Learning. In Proceedings of the IEEE INFOCOM 2020 - IEEE Conference on Computer Communications; IEEE, July 2020; pp. 1698–1707.
[30] Tan, A.Z.; Yu, H.; Cui, L.; Yang, Q. Towards Personalized Federated Learning. IEEE Trans Neural Netw Learn Syst2022, 1–17, doi:10.1109/TNNLS.2022.3160699.
[31] Fallah, A.; Mokhtari, A.; Ozdaglar, A. Personalized Federated Learning: A Meta-Learning Approach. 2020.
[32] Mohassel, P.; Zhang, Y. SecureML: A System for Scalable Privacy-Preserving Machine Learning. In Proceedings of the 2017 IEEE Symposium on Security and Privacy (SP); IEEE, May 2017; pp. 19–38.
[33] Cheng, K.; Fan, T.; Jin, Y.; Liu, Y.; Chen, T.; Papadopoulos, D.; Yang, Q. SecureBoost: A Lossless Federated Learning Framework. 2019.
[34] Yang, S.; Ren, B.; Zhou, X.; Liu, L. Parallel Distributed Logistic Regression for Vertical Federated Learning without Third-Party Coordinator. 2019.
[35] Feng, S.; Yu, H. Multi-Participant Multi-Class Vertical Federated Learning. 2020.
[36] Gu, B.; Dang, Z.; Li, X.; Huang, H. Federated Doubly Stochastic Kernel Learning for Vertically Partitioned Data. In Proceedings of the Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; ACM: New York, NY, USA, August 23 2020; pp. 2483–2493.
[37] Chen, T.; Jin, X.; Sun, Y.; Yin, W. VAFL: A Method of Vertical Asynchronous Federated Learning. 2020.
[38] Jeong, E.; Oh, S.; Kim, H.; Park, J.; Bennis, M.; Kim, S.-L. Communication-Efficient On-Device Machine Learning: Federated Distillation and Augmentation under Non-IID Private Data. 2018.
[39] Duan, M.; Liu, D.; Chen, X.; Liu, R.; Tan, Y.; Liang, L. Self-Balancing Federated Learning With Global Imbalanced Data in Mobile Systems. IEEE Transactions on Parallel and Distributed Systems2021, 32, 59–71, doi:10.1109/TPDS.2020.3009406.
[40] Finn, C.; Abbeel, P.; Levine, S. Model-Agnostic Meta-Learning for Fast Adaptation of Deep

Networks. In Proceedings of the Proceedings of the 34th International Conference on Machine Learning; Precup, D., Teh, Y.W., Eds.; PMLR, October 2017; Vol. 70, pp. 1126–1135.
[41] Briggs, C.; Fan, Z.; Andras, P. Federated Learning with Hierarchical Clustering of Local Updates to Improve Training on Non-IID Data. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN); IEEE, July 2020; pp. 1–9.
[42] Smith, V.; Chiang, C.-K.; Sanjabi, M.; Talwalkar, A. Federated Multi-Task Learning. 2017.
[43] Hanzely, F.; Richtárik, P. Federated Learning of a Mixture of Global and Local Models. 2020.
[44] Chen, Y.; Wang, J.; Yu, C.; Gao, W.; Qin, X. FedHealth: A Federated Transfer Learning Framework for Wearable Healthcare. 2019.
[45] Yang, M.; Wang, X.; Zhu, H.; Wang, H.; Qian, H. Federated Learning with Class Imbalance Reduction. In Proceedings of the 2021 29th European Signal Processing Conference (EUSIPCO); IEEE, August 23 2021; pp. 2174–2178.
[46] Pfitzner, B.; Steckhan, N.; Arnrich, B. Federated Learning in a Medical Context: A Systematic Literature Review. ACM Trans Internet Technol2021, 21, 1–31, doi:10.1145/3412357.
[47] Vepakomma, P.; Gupta, O.; Swedish, T.; Raskar, R. Split Learning for Health: Distributed Deep Learning without Sharing Raw Patient Data. 2018.

SUPPLEMENTARY

Table 1. Description of data distribution by icu for common variables of Physionet Challenge 2012 data set.

		CCU (n=889)	CSRU (n=1,219)	MICU (n=2,216)	SICU (n=1,676)	ExternalICU (n=2,000)	P-Value *
Age		69.4(14.6)	67.6(13.1)	63.5(18.1)	60.3(19.3)	64.1(12.2)	<0.001
Gender	Female	357(40.2)	453(35.8)	1075 (50.1)	706(41.6)	241(45.2)	<0.001
	Male	531(59.8)	812(64.2)	1070 (49.9)	992(58.4)	292(54.8)	
Height		170.6(17.8)	169.9(10.5)	168.3(19.7)	170.1(17.3)	169.3(23.2)	<0.001
Weight		80.7(21.8)	87.4(20.0)	82.3(27.2)	83.0(25.8)	81.9(23.3)	<0.001
In-hospitaldeath	Alive	773(87.0)	1205 (95.2)	1724 (80.3)	1457 (85.8)	453(85.0)	<0.001
	Death	115(13.0)	61(4.8)	423(19.7)	242(14.2)	80(15.0)	
Length of stay	<7days	396(44.6)	455(35.9)	801(37.3)	453(26.7)	189(35.5)	<0.001
	>7days	492(55.4)	811(64.1)	1346 (62.7)	1246 (73.3)	344(64.5)	

*One-way analysis ofvariance (ANOVA) for continuous features; χ^2-test for categorical features.

Table 2.Performance evaluation of PPFL compared to FedAvg, Local (using common features), Local (using common and specific features) in internal and external validation.

In hospital mortality										
Client	Client-specific vertical features									
1 CCU	DiasABP	PaO2	pH	SysABP	Lactate	HR	SaO2	Bilirubin	ALP	Platelets
2 CSRU	Na	Albumin	PaO2	FiO2	SaO2	Urine	pH	Lactate	Creatinine	SysABP
3 MICU	PaCO2	Temp	Na	K	PaO2	Creatinine	HCT	SysABP	Bilirubin	pH
4 SICU	pH	HCT	MAP	SysABP	Albumin	Mg	Platelets	DiasABP	K	FiO2

Table 3.Performance evaluation of PPFL compared to FedAvg, Local (using common features), Local (using common and specific features) in internal and external validation.

		In hospital mortality				Length of stay (>7)			
		Local		External		Local		External	
Client	Model	Accuracy	AUROC	Accuracy	AUROC	Accuracy	AUROC	Accuracy	AUROC
1.CCU	FedAvg (x)	0.857	0.671	0.818	0.616	0.650	0.690	0.710	0.643
	PPFL(x)	0.862	0.773	0.860	0.640	0.862	0.715	0.860	0.671
	PPFL(x,s)	**0.879**	**0.827**	**0.845**	**0.803**	**0.871**	**0.853**	**0.862**	**0.861**
	Local(x)	0.860	0.657	0.823	0.598	0.852	0.803	0.839	0.636
	Local(x,s)	0.871	0.810	0.835	0.781	0.864	0.822	0.847	0.792
2.CSRU	FedAvg (x)	0.951	0.614	0.818	0.616	0.535	0.661	0.710	0.643
	PPFL(x)	0.937	0.643	0.814	0.617	0.923	0.690	0.816	0.625
	PPFL(x,s)	**0.954**	**0.873**	**0.836**	**0.762**	**0.954**	**0.833**	0.856	**0.719**
	Local(x)	0.952	0.635	0.818	0.576	0.927	0.691	0.851	0.596
	Local(x,s)	0.926	0.824	0.818	0.671	0.931	0.714	**0.860**	0.710
3.MICU	FedAvg (x)	0.809	0.616	0.818	0.616	0.640	0.593	0.710	0.643
	PPFL(x)	0.812	0.643	0.820	0.655	0.815	0.643	0.860	0.703
	PPFL(x,s)	0.815	**0.715**	**0.847**	**0.789**	**0.864**	**0.695**	**0.868**	**0.779**
	Local(x)	0.809	0.631	0.818	0.604	0.805	0.619	0.860	0.619
	Local(x,s)	**0.818**	0.709	0.841	0.765	0.805	0.690	0.852	0.722
4.SICU	FedAvg (x)	0.833	0.659	0.818	0.616	0.643	0.617	0.710	0.643
	PPFL(x)	0.855	0.672	0.860	0.648	0.851	0.689	0.860	0.659
	PPFL(x,s)	**0.860**	**0.835**	**0.867**	**0.807**	**0.856**	**0.853**	0.864	**0.873**
	Local(x)	0.803	0.665	0.818	0.622	0.741	0.692	0.858	0.657
	Local(x,s)	0.846	0.792	0.862	0.764	0.851	0.796	**0.871**	0.865

Table 4. Internal and external validation of using client-specific features in each client.

		In hospital mortality			
		Internal		External	
Client	Model	Accuracy	AUROC	Accuracy	AUROC
CCU	FedAvg (x)	0.857	0.671	0.818	0.616
	PPFL (x,s)	**0.871**	**0.838**	**0.862**	**0.723**
CSRU	FedAvg (x)	0.951	0.614	0.818	0.616
	PPFL (x,s)	**0.954**	**0.847**	**0.861**	**0.760**
MICU	FedAvg (x)	**0.809**	0.616	0.818	0.616
	PPFL (x,s)	0.805	**0.774**	**0.860**	**0.745**
SICU	FedAvg (x)	0.833	0.659	0.818	0.616
	PPFL (x,s)	**0.860**	**0.781**	**0.865**	**0.772**