GSC Advanced Research and Reviews

eISSN: 2582-4597 CODEN (USA): GARRC2

Cross Ref DOI: 10.30574/gscarr

Journal homepage: https://gsconlinepress.com/journals/gscarr/

(RESEARCH ARTICLE)

Check for updates

Efficiency measurement of FL algorithms for image classification

Mushfiqur Rahman Abir, Asif Zaman * and Sawon Mursalin

Department of Computer Science & Engineering, American International University-Bangladesh, Khilkhet, Dhaka 1229, Bangladesh.

GSC Advanced Research and Reviews, 2024, 18(03), 356–366

Publication history: Received on 02 February 2024; revised on 16 March 2024; accepted on 19 March 2024

Article DOI: https://doi.org/10.30574/gscarr.2024.18.3.0110

Abstract

Federated Learning (FL) has emerged as a promising approach to collaborative machine learning without the need to share raw data. It enables decentralized model updates while preserving the privacy of each device and reducing the communication overhead. This experiment evaluates the effectiveness of the personalized FL algorithms, namely FedAvg, APPLE, FedBABU and FedProto, in a decentralized setting, with a particular focus on the Fashion MNIST dataset, which is characterized by a non-ideal data distribution. The objective is to identify which algorithm performs optimally in image classification tasks. The experimental results show that both FedProto and APPLE have nearly equivalent and better performance compared to FedBABU and FedAvg. Interestingly, increasing the number of uploads in FedBABU leads to similar results to APPLE and FedProto. However, under limited upload conditions, FedBABU performs similarly to FedAvg. These results provide valuable insights into the differential performance of personalized FL algorithms in non-id data scenarios and provide guidance for their application in distributed environments, especially in sensitive domains such as medical, military and confidential image analysis tasks where privacy and communication efficiency are paramount concerns.

Keywords*:* Federated learning (FL); Data privacy; Model personalization; Image classification

1. Introduction

Traditional machine learning (ML) methods come with limitations related to data privacy, data storage, and computational expenses [1]. This raise concerns about privacy and security, as the centralization of data makes them susceptible to breaches. Moreover, this centralization requires significant storage capacity and infrastructure, resulting in increased costs and data management complexity. The computational costs associated with central training and data processing can be burdensome and time-consuming, limiting scalability and overall efficiency. In addition, the transmission of raw data for training purposes can strain bandwidth and be challenging in scenarios with limited network connectivity. Traditional ML methods also have difficulties to comply with privacy regulations and data protection laws. In response to these issues, FL has emerged as a privacy-preserving alternative that decentralizes data, reduces data transfer requirements, prioritizes privacy and minimizes computational overhead [2].

FL varies from ML in some key aspects where in ML, data is mostly centralized and stored in a single location, such as a cloud or a server platform. On the other hand, in FL the data is distributed on devices or edge servers. FL maintains data privacy by keeping it local and not distributing it with a central server, instead exchanging model updates in between the central server and devices [3]. Moreover, FL fosters collaborative learning, enabling multiple entities to contribute their data and models. FL offers several advantages, including the preservation of privacy by keeping sensitive data on the devices, efficient utilization of diverse data distributed across multiple devices, reduced communication costs as only model updates are exchanged, enhanced scalability to accommodate numerous devices, and adaptability to dynamic environments where devices can join or leave the network without requiring centralized retraining [2].

* Corresponding author: Asif Zaman.

Nonetheless, conventional FL primarily concentrates on amalgamating model updates from various devices to construct a global model. This approach may encounter challenges in terms of personalization and may grapple with the inherent diversity in devices and data sources [4]. Personalized federated learning (PFL) overcome the limitations of traditional FL by introducing user-specific model updates, considering device heterogeneity and enhancing user trust [5]. Customized global models that take into account each user's preferences are possible with PFL, which enhances accuracy and relevance. Additionally, it provides more precise control over privacy and data sharing, boost-ing user involvement and confidence. PFL is a potential strategy for a variety of applications since it combines the advantages of FL with personalized learning. [6].

In the ever-evolving landscape of image classification methodologies, a significant imperative lies in conducting a comprehensive study to assess the efficacy of personalized federated learning. This endeavor assumes particular relevance in light of the burgeoning significance placed on data privacy and the need to tailor outcomes to individual preferences. Consequently, this paper delves into the evaluation of diverse federated learning algorithms deployed for image classification tasks. Notably, the primary contribution of this research encompasses a meticulous performance comparison of personalized federated learning algorithms: FedAvg [2], APPLE [7], FedBABU [8], and FedProto

[9] using the Fashion-MNIST dataset [10]. The comparison will give an idea of how PFL algorithms perform on heterogeneous image classification tasks in a non-iid setup and which PFL algorithm provides better accuracy.

The rest of the paper is segmented as follows: Section 2 discussed the related work done in various domains of FL: Section 3 described the algorithms used in this experiment; Section 4 analyzed the dataset, experimental set-up and the results of the experiment; and finally, Section 5 discussed the overall contribution and conclude the paper.

2. Related Work

2.1. Non identical and independently distributed (non-iid) federated learning

The systemic privacy concerns are reduced by FL which enables clients to cooperatively train a model without sharing their data. FedAvg [2] the most used FL method, apply this by aggregating updated copies of global model by using the averaging strategy. Due to the worries on FedAvg's behaviour regarding non-iid data started to raise [11] research has been done on FL's robustness on the non-iid data. In 2020, FedProx was proposed which penalises local update when client is in distant from prox-center [11]. Then SCAFFOLD was introduced [12], which uses control variates to adjust the local gradient caused by client-drift. To mitigate the effects of data heterogeneity, FedDynp [13] was introduced which dynamically changes the regularizer in empirical risk.

2.2. Heterogeneous federated learning

The non-iid problem, or statistical heterogeneity among clients, is the biggest obstacle facing FL. To improve each client's local model, FedProx [11] suggested a local regularization term. Recent research [14] trains personalized models to take use of both the individualized component and globally available information. The third option is to give numerous global models by grouping the local models into different groups or clusters [15]. In order to address the issues posed by heterogeneity, self-supervised learning algorithms have recently been integrated into the local training phase [16]. For personalized FL, [17] uses a meta-learning technique. Heterogeneous model design is another difficult FL situation. Without having access to the local training data and architectures, a collective learning platform is developed to manage heterogeneous architectures [18].

2.3. Personalized federated learning

A new branch of FL, the personalized FL, has been created to methodically reduce the influence of data het- erogeneity [6]. Personalized FL, especially when the data are taken from multiple distributions, permits alternative models for each client rather than being constrained by the global consensus model [7]. The goal of personalized federated learning is to acquire locally tailored, individualized models for each client [19]. A logical approach in this line of work is to adjust the global model for each client [20]. It was then found that fine-tuning the global model may have a negative impact on its ability to generalize to new data [21]. Local models can be created without federation, although this approach has data issues. Thus, several additional techniques, like as clustering, multi-task learning, transfer learning, regularized loss function, and meta-learning, have been used to FL in order to preserve the advantages of the federation and individualized models [19]. Studies have also concentrated on the relationship between FL and meta-learning [22]. In addition to methods that need further adjusting the learned models, MOCHA was presented [23] , which makes use of multi-task learning to discover the connections between various clients. Furthermore, various technologies and algorithms for personalized FL have been presented and studies are going on.

3. Algorithms

The algorithms FedAvg, APPLE, FedBABU, and FedProto are all part of our evaluation. The pseudocode for these algorithms is provided in this section.

3.1. Federated Averaging (FedAvg)

FedAvg is used for ML model training in a decentralized manner across different divices or clients [2]. The concept is to divide the training process over a group of clients rather than gathering all client data on a single server. Each client uses its own data to update its local model, which it then sends back to the server. The server then combines all the models to create a new global model, which is distributed to the clients to begin a new training cycle.

Algorithms 1 shows the working procedure of FedAvg.

Algorithm 1 FedAvg

Fire up the symphony with an initial note w_0

for each enchanting round $t = 0, 1, \ldots$ **do**

Summon a magical gathering S_t of mystical clients

Let m be the spellbound count, at least one to ensure the cosmic dance

for each mesmerized client $k \in S_t$ **in parallel do**

Allow $w^k{}_{t+1}$ to be crafted through the sacred ritual **ClientUpdate**(k, w_t)

end for

Unleash the unity $w_{t+1} = \sum k \in S_1 \frac{nk}{n} W^k{}_{1+1}$ of their mystical energies

end for

3.2. Adaptive Personalized Cross-Silo Federated Learning (APPLE)

APPLE is a personalized cross-silo FL framework that adaptively learns how much each client can benefit from other clients models [7]. Like most FL approaches, the training in APPLE progresses in rounds. Each client downloads and uploads the model's parameters throughout each iteration. A core model serves as the foundation for each client's customised model that is uploaded to APPLE. On the central server, the fundamental models that clients upload are likewise kept up to date.

Algorithms 2 shows the working procedure of APPLE.

3.3. Federated Averaging with Body Aggregation and Body Update (FedBABU)

FedBABU is a type of FL algorithm that only updates the model's body while training (the head is initialized randomly and never updated), and the head is fixed for personalization in the evaluation process [8].

Algorithms 3 shows the working procedure of FedBABU.

3.4. Federated Prototype Learning (FedProto)

FedProto or federated prototype learning is a framework where the clients and server communicate abstract class prototypes instead of gradients [9].

Algorithms 4 shows the working procedure of FedProto.

To regularize the training of local models, FedProto collects the local prototypes from many customers and aggre- gates them before sending the global prototypes back to each client. The training on each client seeks to reduce the classification error on the local data while keeping the resultant local prototypes sufficiently near to the corresponding global ones.

Algorithm 2 Adaptive Personalized Cross-Silo Federated Learning

- 1: Summon N curious minds, each equipped with the elixir of learning rates η_1 and η_2, a treasure map with R
- destinations, the magical coefficients $\lambda(r)$, μ, and the mystical proximal center p_0.
- 2: Unleash the spirit of randomness upon N disciples, bestowing upon them the primal essence of knowledge, encapsulated in the enigmatic core model $w_i^{(c)}$ on the sacred server.
- 3: Ignite the flames of parallelism within each disciple's soul as they embark on their individual quests, armed with local DR vectors p_i as companions.
- 4: **for** $r \leftarrow 1, 2, \ldots, R$ **do**
- 5: **for** $i \leftarrow 1, 2, \ldots, N$ in a synchronized dance **do**
- 6: Summons echo from the server, acquiring core models as dictated by destiny.
- 7: Commence the ritual of iterative enlightenment for the local core model$_i w^{(c)}$ and its loyal companion p_i:
- 8: Conjure the personalized model $w_i^{(p)}$ through the harmonious blend of $p_{i,j}$ and $w_j^{(c)}$.
- 9: Invoke the oracle to reveal the empirical risk $F_i(w_i^{(p)})$, a potion concocted from the sacred elixir of data and
- the ethereal dance of loss functions.
- 10: Perform the sacred dance of updates, guiding $w_i^{(c)}$ and the DR vector p_i through the mystical gradients of enlightenment.
- 11: Upon the completion of the mystical dance, offer the sanctified local core model $w_i^{(c)}$ to the server for communion.
- 12**end for**
- 13: **end for**
- 14: **Emerge Victorious:** Receive the personalized models $w_1^{(p)}, w_2^{(p)}, \ldots, w_N^{(p)}$, each a unique relic forged in the crucible
- of collective wisdom, bestowed upon the respective disciples.

Algorithm 3 Federated Averaging with Body Aggregation and Body Update

1: Initialize the magical global parameters $\theta^0{}_G = \{\theta^0{}_{G,ext}, \theta^0{}_{G,clx}\}$

2: **for** every enchanting round $k = 1, \ldots, K$ **do**

3: $m \leftarrow \max(\lfloor N f \rfloor, 1)$

4: C^k conjures a random subset of m clients

5: **for** each spellbound client $C^k{}_i \in C^k{}_i$ **simultaneously do**

6: $\theta_i{}^k(0) \leftarrow \theta^k{}_G{}^{-1} = [\theta^{k-1}{}_{G,ext}, \theta^0{}_{G,cls}]$

7: $\theta^k{}_{i,ext}(\tau I^k) \leftarrow$ **ClientBodyUpdate**$(\theta^k{}_i(0), \tau)$

8: **end for** m
k

9: $\theta^K{}_{G,ext} \leftarrow \sum^m{}_{i=1} \frac{nCki}{n} \theta^K{}_{i,ext}(\tau I^k{}_i), n = \sum^m{}_{i=1} {}^nC^k{}_i$

10: **end for**

11: **return** the mystical $\theta^K{}_G = \{\theta^K{}_{G,ext}, \theta^0{}_{G,cls}\}$

Algorithm 4 Federated Prototype Learning

1: *Harmonize the Essence:* Initialize the individual essence D_i, the unique harmonizing factor w_i, where i gracefully dances through the classes 1 to m, and the ethereal global prototype set $\bar{C}(j)$ for each enchanting class.

2: **for** each magical round $T = 1, 2, \ldots$ **do**

3: **for** each participant i **in parallel do**

4: $C_i \leftarrow$ **Local Enchantment** $(i, \bar{C} i)$

5: **end for**

6: *Unveil the Collective Symphony:* Update the global prototype, letting $\bar{C}^{(j)}$ bloom like a garden of shared dreams, nurtured by the magical collaboration of all entities in N_j.

7: *Empower Local Identities:* Infuse the local essence C_i with the awakened prototypes from $[\bar{C}^{(j)}]$.

8: **end for**

9: **return** the transcendent global prototype set $\bar{C}^{(j)}$

4. Results and discussion

4.1. Experimental Setup

We base our benchmarks on the Fashion MNIST [10] dataset of apparel items with non-iid data distribution. The Fashion-MNIST dataset comprises images of various clothing items, including T-shirts, dresses, pants, and shoes. Each apparel category within the dataset is composed of 60,000 training images and 10,000 test images. These images are grayscale and have dimensions of 28x28 pixels, with pixel values spanning the range from 0 to 255.

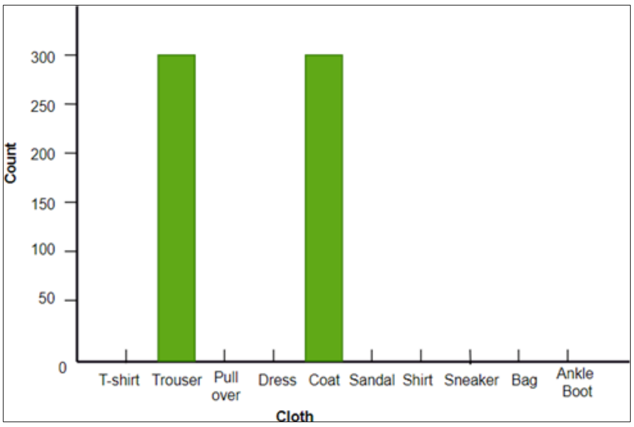

Figure 1 Data Non-i.i.d. distribution

Non-iid data samples are intentionally assigned to specific clients based on their unique characteristics. This customization was achieved through a process involving data sorting and segmentation. [2]. The step-by-step pro- cess involves organizing the data, dividing it into uniform-sized segments, and subsequently distributing a specified quantity of these segments to individual clients. In the case of Fashion MNIST, each of the 20 clients received two segments, each comprising 300 data points. Fig. 1 visualizes how our non-iid distributions are different from each other.

Table 1 Machines specifications

Machine	Central Processing Unit	Memory	Operating System
A	i9-9900K	31.75 GiB	Linux
B	i9-9900K	31.75 GiB	Linux
C	i9-10900K	31.76 GiB	Linux
D	i9-10900K	31.76 GiB	Linux

To complete the experiment we used four computers in a star topology, that were Ethernet-connected to one another. All computers used Arch Linux 2023.04.01 (x86 64); kernel version 6.2.2-arch1-1. The Table 1 provide lists the details of each machine.

We divided our 20 clients as evenly as we could among all of our distributed benchmarks, using one of our machines as the server. Using a convolutional neural network (CNN) artificial neural network architecture that is specifically created for image classification tasks of 10 different classes, we cross-validate FedAvg, APPLE, Fed- BABU, and FedProto on the Fashion MNIST. The overall design is appropriate for image classification tasks using grayscale input images that are at least 28x28 pixels in size. The design has two convolutional layers, each followed by a fully connected layer with a ReLU activation function and a max-pooling layer, and an output layer with a soft- max function. Fig. 2 illustrates the architecture of artificial neural network used in our experiment. For searching the suitable hyperparameter values, we used random searches [24]. The initial set of hyperparameters was directly derived from the default set of hyperparameters in [2] for the Fashion MNIST dataset.

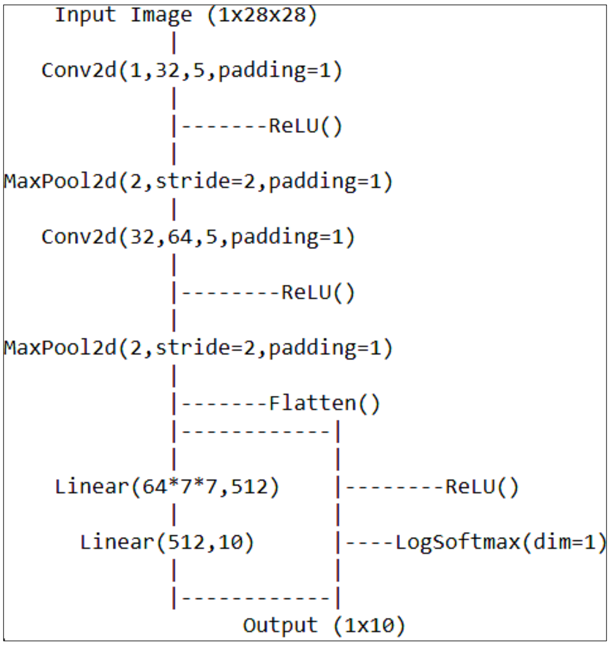

Figure 2 Convolutional neural network architecture

Table 2 Mean test accuracy of four Federated Learning algorithms

Algorithm	**Mean test accuracy**
FedAvg	0.7604
APPLE	0.9921
FedBABU	0.7462
FedProto	0.9940

4.2. Result Analysis

Table 2 displays the mean top test accuracy outcomes for four Federated Learning techniques used in image classification. The algorithms' effectiveness is ranked according to their mean accuracy scores: FedAvg achieved 76.04%, APPLE performed exceptionally well with 99.21%, FedBABU scored 74.62%, and FedProto displayed an impressive average accuracy of 99.40%. These findings highlight the different capabilities of these algorithms, with APPLE and FedProto showing remarkable image classification performance, and FedAvg and FedBABU also deliv- ering competitive results.

Table 3 Comparisons, signing whether an algorithm in a row is upper (+), lower (), or nearly equal (=) to an algorithm in a column.

	FedAvg	**APPLE**	**FedBABU**	**FedProto**
FedAvg		-	=	-
APPLE	+	-	+	=
FedBABU	=			-
FedProto	+	=	+	

Nonetheless, the federated algorithms faced limitations due to the restricted number of uploads, potentially creat- ing an unfair basis for comparison. After 200 global updates, FedAvg, APPLE, and FedProto demonstrated minimal improvement, leaving uncertainty about how much FedBABU could progress with greater communication. To address this, an experiment was conducted, granting FedBABU a larger communication budget. It was discovered that when given 1.5 times more uploads, FedBABU reached a performance level practically on par with APPLE and FedProto, surpassing FedAvg, as illustrated in Fig 7. The overall results of the tests are visualized in Fig. 3-6 in brief.

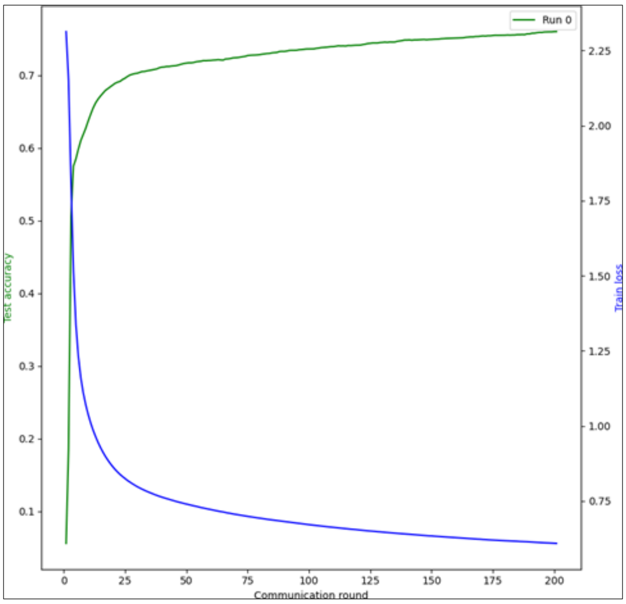

Figure 3 FedAvg test accuracy, train loss with respect to communication round

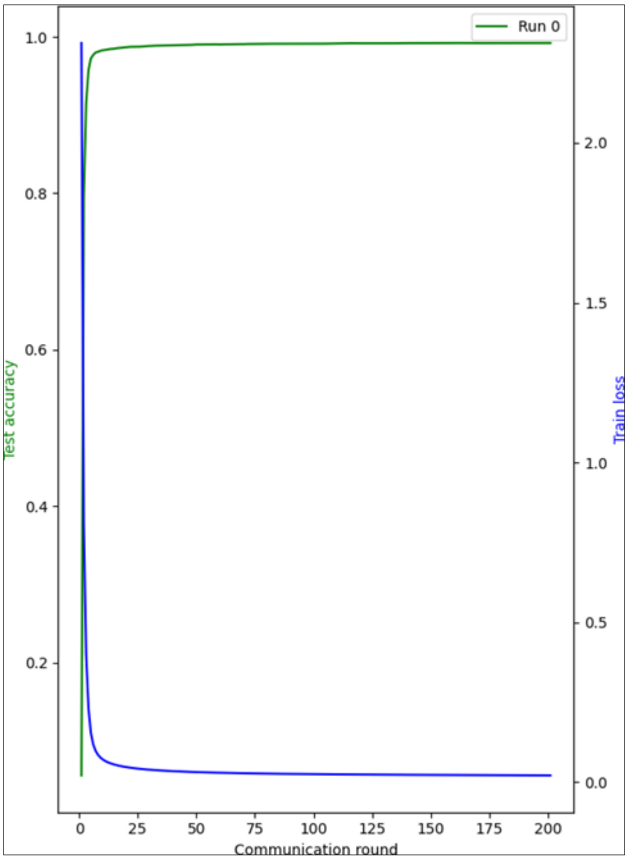

Figure 4 APPLE test accuracy and train loss with respect to communication round

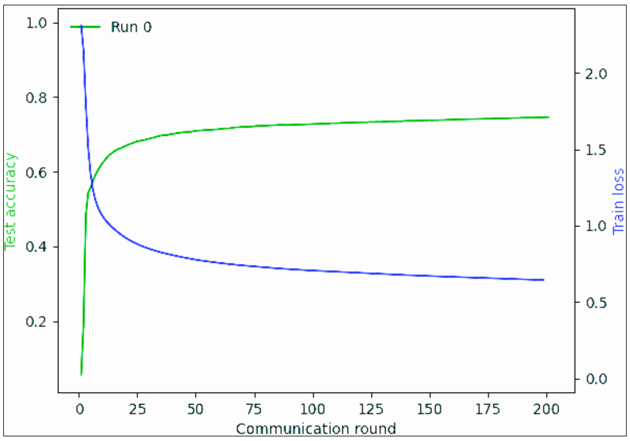

Figure 5 FedBABU test accuracy and train loss with respect to communication round

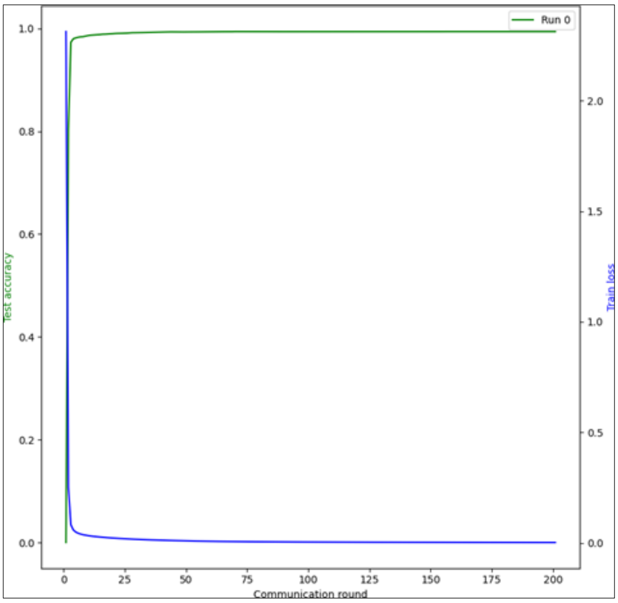

Figure 6 FedProto test accuracy and train loss with respect to communication round

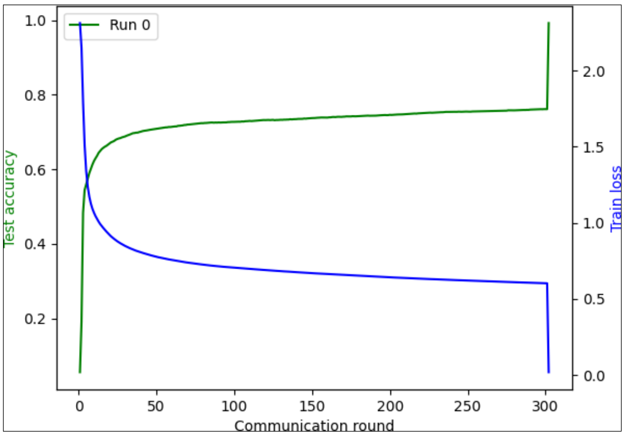

Figure 7 FedBABU test accuracy, train loss with respect to Communication round on extended upload

In this research, the study assessed and contrasted the performance of the federated learning algorithms FedAvg, APPLE, FedBABU, and FedProto using the Fashion MNIST dataset. In the experiments conducted with a multi-layer perceptron model, it was observed that APPLE and FedProto exhibited significantly superior performance on Fashion MNIST. Table 3 provides a summary of the comparison of these four personalized federated learning algorithms as part of the study.

5. Conclusion

This study assessed four Federated Learning algorithms for image classification and found that FedProto and APPLE performed nearly equally well, outperforming the FedBABU and FedAvg algorithms. Initially, the experi- ments used balanced data distributions to provide equal data for all clients, which is a useful starting point. However, this approach may not fully capture the complexity of the diverse federal environment. Consequently, extending our investigation to evaluate performance in scenarios where client data is unevenly distributed would be a logical pro- gression of our research. By investigating different levels of data imbalance, we can gain a deeper understanding of how heterogeneity affects the efficiency and resilience of federated learning algorithms. The results of this study have important implications for image classification tasks with heterogeneous data, such as in medical, military, or other applications where privacy is a critical concern.

Compliance with ethical standards

Acknowledgements

We would like to thank Dr. M. F. Mridha for his valuable insights.

Disclosure of conflict of interest

No conflict of interest to be disclosed.

Declarations

An older version of this article has been preprinted at https://www.researchsquare.com/article/rs-3056513/v1.

References

[1] S. Niknam, H. S. Dhillon, and J. H. Reed, "Federated learning for wireless communications: Motivation, opportunities, and challenges,"

[2] *IEEE Communications Magazine*, vol. 58, no. 6, pp. 46–51, 2020.

[3] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*, pp. 1273–1282, PMLR, 2017.

[4] Z. Li, V. Sharma, and S. P. Mohanty, "Preserving data privacy via federated learning: Challenges and solutions," *IEEE Consumer Electronics Magazine*, vol. 9, no. 3, pp. 8–16, 2020.

[5] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings, *et al.*, "Advances and open problems in federated learning," 2019.

[6] M. G. Arivazhagan, V. Aggarwal, A. K. Singh, and S. Choudhary, "Federated learning with personalization layers," *arXiv preprint arXiv:1912.00818*, 2019.

[7] V. Kulkarni, M. Kulkarni, and A. Pant, "Survey of personalization techniques for federated learning," in *2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)*, pp. 794–797, IEEE, 2020.

[8] J. Luo and S. Wu, "Adapt to adaptation: Learning personalization for cross-silo federated learning," *arXiv preprint arXiv:2110.08394*, 2021.

[9] J. Oh, S. Kim, and S.-Y. Yun, "Fedbabu: Towards enhanced representation for federated image classification," *arXiv preprint arXiv:2106.06042*, 2021.

[10] Y. Tan, G. Long, L. Liu, T. Zhou, Q. Lu, J. Jiang, and C. Zhang, "Fedproto: Federated prototype learning across heterogeneous clients," in

[11] *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, pp. 8432–8440, 2022.

[12] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms," *arXiv e-prints*,

[13] pp. arXiv–1708, 2017.

[14] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," *Proceedings of Machine learning and systems*, vol. 2, pp. 429–450, 2020.

[15] S. P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, and A. T. Suresh, "Scaffold: Stochastic controlled averaging for federated learning," in *International Conference on Machine Learning*, pp. 5132–5143, PMLR, 2020.

[16] A. E. Durmus, Z. Yue, M. Ramon, M. Matthew, W. Paul, and S. Venkatesh, "Federated learning based on dynamic regularization," in

[17] *International Conference on Learning Representations*, 2021.

[18] P. P. Liang, T. Liu, L. Ziyin, N. B. Allen, R. P. Auerbach, D. Brent, R. Salakhutdinov, and L.-P. Morency, "Think locally, act globally: Federated learning with local and global representations," *arXiv preprint arXiv:2001.01523*, 2020.

[19] Y. Mansour, M. Mohri, J. Ro, and A. T. Suresh, "Three approaches for personalization with applications to federated learning," *arXiv preprint arXiv:2002.10619*, 2020.

[20] Q. Li, B. He, and D. Song, "Model-contrastive federated learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10713–10722, 2021.

[21] A. Fallah, A. Mokhtari, and A. Ozdaglar, "Personalized federated learning with theoretical guarantees: A model-agnostic meta-learning approach," *Advances in Neural Information Processing Systems*, vol. 33, pp. 3557–3568, 2020.

[22] M. Hoang, N. Hoang, B. K. H. Low, and C. Kingsford, "Collective model fusion for multiple black-box experts," in *International Conference on Machine Learning*, pp. 2742–2750, PMLR, 2019.

[23] J. H. OH, S. Kim, and S. Yun, "Fedbabu: Toward enhanced representation for federated image classification," in *10th International Conference on Learning Representations, ICLR 2022*, International Conference on Learning Representations (ICLR), 2022.

[24] K. Wang, R. Mathews, C. Kiddon, H. Eichner, F. Beaufays, and D. Ramage, "Federated evaluation of on-device personalization," *arXiv preprint arXiv:1910.10252*, 2019.

[25] Y. Jiang, J. Konecˇny`, K. Rush, and S. Kannan, "Improving federated learning personalization via model agnostic meta learning," *arXiv preprint arXiv:1909.12488*, 2019.

[26] M. Khodak, M.-F. F. Balcan, and A. S. Talwalkar, "Adaptive gradient-based meta-learning methods," *Advances in Neural Information Pro- cessing Systems*, vol. 32, 2019.

[27] V. Smith, C.-K. Chiang, M. Sanjabi, and A. S. Talwalkar, "Federated multi-task learning," *Advances in neural information processing systems*, vol. 30, 2017.

[28] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization.," *Journal of machine learning research*, vol. 13, no. 2, 2012.