Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/mbe

MBE, 20(6): 10725–10740.
DOI: 10.3934/mbe.2023475
Received: 29 January 2023
Revised: 28 March 2023
Accepted: 28 March 2023
Published: 17 April 2023

Research article

Personalized federated learning for heterogeneous data: A distributed edge clustering approach

Muhammad Firdaus[1], Siwan Noh[2], Zhuohao Qian[2], Harashta Tatimma Larasati[3] and Kyung-Hyune Rhee[4,*]

[1] Department of Artificial Intelligence Convergence, Pukyong National University, Busan 48513, Republic of Korea
[2] Department of Information Security, Pukyong National University, Busan 48513, Republic of Korea
[3] School of Electrical Engineering and Informatics, Institut Teknologi Bandung, Bandung 40132, Indonesia
[4] College of Information Technology and Convergence, Division of Computer Engineering and AI, Pukyong National University, Busan 48513, Republic of Korea

* **Correspondence:** Email: khrhee@pknu.ac.kr; Tel: +82516296247; Fax: +82516264887.

Abstract: Federated learning (FL) is a distributed machine learning technique that allows multiple devices (e.g., smartphones and IoT devices) to collaborate in the training of a shared model with each device preserving the privacy of its local data. However, the highly heterogeneous distribution of data among clients in FL can result in poor convergence. In addressing this issue, the concept of personalized federated learning (PFL) has emerged. PFL aims to tackle the effects of non-independent and identically distributed data and statistical heterogeneity and to achieve personalized models with rapid model convergence. One approach is clustering-based PFL, which utilizes group-level client relationships to achieve personalization. However, this method still relies on a centralized approach, whereby the server coordinates all processes. To address these shortcomings, this study introduces a blockchain-enabled distributed edge cluster for PFL (BPFL) that combines the benefits of blockchain and edge computing. Blockchain technology can be used to enhance client privacy and security by recording transactions on immutable distributed ledger networks, thereby improving client selection and clustering. The edge computing system offers reliable storage and computation such that computational processing is locally performed in the edge infrastructure to be closer to clients. Thus, the real-time services and low-latency communication of PFL are improved. However, further work is required to develop a representative dataset for the examination of related types of attacks and defenses for a robust BPFL protocol.

Keywords: personalized FL; non-IID data; blockchain; edge computing; client clustering

1. Introduction

The emergence of the Internet of Things (IoT) and the widespread utilization of mobile devices with advanced computing capabilities have instigated the need for large-scale data acquisition. These data can be harnessed to train advanced artificial intelligence (AI) models that can provide various smart services, benefiting society in diverse aspects. Nevertheless, traditional data acquisition models that rely on centralized machine learning models entail security and privacy challenges and reduce participation in data contribution. Moreover, with the recent establishment of data privacy preservation regulations, such as the General Data Protection Regulation [1] and the Health Insurance Portability and Accountability act [2], the need for privacy-preserving AI has been continuously growing. Hence, federated learning (FL), which allows clients to train their data collaboratively without exposing their private data to each other, has become more prevalent as a promising approach for tackling privacy-preserving AI issues. The introduction of FL was first used by Google for Gboard's next-word prediction [3]. Subsequently, empowered by the success of Gboard, the utilization of FL was promoted in a myriad of applications, such as healthcare [4], industrial IoT [5], vehicular networks [6], finance [7] and so forth.

Despite the benefits of general FL, the federated averaging (FedAvg) [8] approach, has suffered from poor convergence from the highly heterogeneous and non-independent nature of data distributions across clients [9]. A global model is trained using data from multiple clients; however, the data from each client are not necessarily identically distributed. This approach can lead to poor model performance for certain clients, whereby the negative impact can lead to the reluctance or refusal of clients to participate in the FL training process. Thus, the concept of personalized federated learning (PFL), as a variation of FL, was developed to alleviate the impact of non-identically and independently distributed (non-IID) data and statistical heterogeneity issues for the efficient creation of tailored models. PFL addresses these problems by personalizing the model for each client, considering the specific characteristics of their data. Moreover, it allows for better performance and increases clients' FL participation.

One approach to PFL is clustering-based PFL [10], which enables personalization using a multi-model approach with group-level client interactions. By mapping client associations, this approach facilitates the learning of personalized models by each client in conjunction with affiliated clients within similar clusters. However, the existing cluster-based PFL method still depends on the centralized strategy where the server orchestrates all processes and requires high communication, incurring computational costs. To address these shortcomings, this study introduces a blockchain-enabled distributed edge cluster for PFL (BPFL) that exploits the benefits of two cutting-edge technologies, blockchain and edge computing, which have the potential to revolutionize the way data are stored, processed, and shared. Specifically, in clustering-based PFL, the integration of blockchain and edge computing enables the creation of clusters based on real-time data, eliminating the need for a central authority to manage and process the data, thereby providing a robust and decentralized solution for data management and processing. In this study, blockchain is used to enrich client privacy and security by recording all transactions in immutable distributed ledger networks to enhance efficient client selection and clustering. Blockchain can be utilized to establish a decentralized network of devices, where each device maintains a copy of the same data, making the process more resilient to data loss or tampering. Likewise, to offer appropriate storage and

computation of PFL, we employ an edge computing system, where computational processing is locally performed in the edge infrastructure of PFL to be nearer to clients as data providers. Thus, this approach allows for faster and more efficient data processing and provides proper computation capability by improving real-time services while reducing the latency and bandwidth requirements of PFL.

This paper is organized as follows. Section 2 provides background knowledge related to edge-AI, FL challenges, and blockchain. Section 3 explains the current works related to PFL. In Section 4, we present the proposed model which is based on a distributed edge cluster for PFL. In Section 5, numerical results of BPFL are discussed, and several related concerns are explored. Finally, Section 6 concludes the paper.

2. Background

2.1. The emergence of edge-AI

In 2014, the European Telecommunications Standards Institute introduced the concept of edge computing to optimize the user experience through low latency, high bandwidth, and real-time communication capabilities [11]. Edge computing leverages local infrastructure to enhance response speed and minimize transmission latency during the transaction process by strategically placing the servers in the edge network [12], thus emphasizing proximity to end users [13]. Whereas, edge-AI, also known as edge intelligence, offers the utilization of AI technologies at the perimeter of a network as opposed to a centralized cloud infrastructure. Specifically, data collection, processing, transmission, and utilization occur at the network edge. The approach enables model training across network nodes, allowing the preservation of privacy and confidentiality [14, 15]. Moreover, this approach can enhance the responsiveness and efficiency of the system by decreasing the volume of data that needs to be transmitted over the network [16–18]. Edge-AI has been increasingly gaining popularity in recent years in both industry and academia. Leading companies such as Google, Microsoft, Intel, and IBM have initiated pilot projects to showcase the benefits of edge computing in the last mile of AI [19].

2.2. Federated learning

The traditional client-server architecture in machine learning involves training on a server, with clients providing the data. However, this approach raises privacy concerns. Clients serve only as providers of data, whereas the server also undertakes the task of data training and aggregation. Various concerns are associated with this classical machine learning strategy, in particular, regarding user privacy. To address this issue, Google introduced FedAvg, a novel communication-efficient optimization algorithm for FL. FedAvg, as outlined in [8], is an efficient method for the distribution-based training of models, allowing distributed mobile devices to collaborate in model training without centralizing the training data and keeping local data stored on mobile devices, thereby improving privacy by blocking access to local data. It also reduces the number of communication rounds, making it more efficient than conventional distributed methods.

In the FedAvg algorithm, the server, acting as a model provider, initially sends the global model to the clients. Each client, as a participant user, downloads the global model from the central server,

generates a model update by training the current global model on local data, and subsequently uploads the trained model to the aggregator server. The central server, then acting as an aggregator, gathers and aggregates all model updates from the clients to produce a new global model for the next iteration. Thus, FedAvg significantly enhances client privacy by blocking attacks from straightforward access to the local training data, as cited in [20]. FedAvg is much more communication-efficient than conventional distributed stochastic gradient descent because of fewer communication rounds. Furthermore, FedAvg frequently leads to improved performance, as demonstrated in various learning-related issues such as predictive models in health, low latency vehicle-to-vehicle communication, vocabulary estimation, and next-word prediction, as cited in [3, 21–23].

2.3. The effect of heterogeneous data in Federated learning

FL offers numerous benefits for various practical applications. However, the highly heterogeneous data distribution among clients also poses several challenges, leading to a lack of personalization and poor convergence. Specifically, the data distribution among clients is highly non-IID, which makes it challenging to train a single model with effective performance for all clients. The non-IID data also significantly affect the accuracy of FedAvg. Since the distribution of each local dataset differs significantly from the global distribution, the local objectives of each client are incompatible with the global optimum, which leads to a drift in local updates, causing each model to be updated towards its own local optimum, which may be far from the global optimum [24]. Especially if there are many significant local updates (i.e., an enormous number of local epochs), the averaged model might also be far from the global optimum [25, 26]. Consequently, the convergence of the global model provides a substantially less accurate solution than that associated with the IID setting.

2.4. Blockchain: Distributed ledger technology

In 2008, Bitcoin, a digital currency system based on blockchain technology, was proposed by Nakamoto for financial transactions. Blockchain is a technology that enables participating nodes to share and validate transactions on a network, which are then imprinted with timestamps and stored in an unchangeable database using a specific consensus mechanism. Blockchain has three primary features: decentralized storage, a distributed ledger, and the ability to support distributed services through the use of smart contracts [27]. Because of these benefits, many researchers from various fields are currently exploring the development of blockchain technology. The advantages of blockchain include anonymity and privacy for users, the immutability of stored data, a decentralized approach that eliminates single points of failure, transparency of transactions, as well as trustful and distributed transactions that do not require a central authority [5]. The feature of immutability ensures that data cannot be deleted or modified from the network, whereas the decentralized approach allows for open participation, provides immunity from particular attacks, and eliminates single points of failure, resulting in consistent, reliable, and widely accessible data, timestamped for recorded transactions. In addition, every user has access to transparent transactions, thereby enabling every node to share and validate transactions in a distributed manner [28].

3. Related works

PFL learning emerged as a response to the problems caused by the non-IID data distribution and statistical heterogeneity of the current FL. The use of FedAvg-based methods for non-IID data result in a decrease in accuracy due to client drift. To overcome these issues, two strategies have been proposed in [29], i.e., the global model personalization strategy, which involves training a single global model, and personalized model learning strategies, which involve training PFL models individually. Most personalization methods for the global FL model typically involve two distinct processes [30]: creating a global model through collaboration or using private client information to personalize the global model [31]. Essentially, PFL uses FedAvg as the standard approach for general FL training settings, with the added step of personalizing the global model using local client data after training. The personalized model strategy for learning aims to achieve PFL by applying different learning algorithms and modifying the FL aggregation process through similarity and architecture-based approaches. Furthermore, PFL seeks to train personalized models for a group of clients by utilizing the non-IID nature of all clients' private data while maintaining their privacy. Thus, to improve the practicality of PFL, the work in [32] suggested that the following three goals be simultaneously addressed: achieving rapid model convergence in a reduced number of training rounds, improving personalized models that benefit a large number of clients, and creating more accurate global models that aid clients with limited private data for personalization.

Recently, research on PFL has been gaining popularity as it aims to address one of the main challenges of current FL. One of the earliest works on PFL was the manager, owner, consultant, helper, approver (MOCHA) framework which was proposed in 2017, as a multi-task learning approach [9]. MOCHA simultaneously learns client task settings and a similarity matrix, to address the issue of federated multi-task learning. The distributed multi-task issues addressed by MOCHA include fault tolerance, stragglers, and communication limitations. In addition, the authors of [10] proposed a context-based client clustering approach to facilitate multiple global models and handle changes in client populations over time. In this approach, FL models are trained for each homogeneous group of clients, whereby clients are divided into clusters based on the cosine similarity of the gradient updates from the clients. There are multiple methods for implementing the PFL framework, such as data augmentation [33], meta-learning [32], transfer learning [34], and fine-tuning [25]. It is worth noting that this study focuses on context-based client clustering with a multi-task approach.

4. Distributed edge cluster for PFL

In this section, we introduce the BPFL approach, which combines the benefits of edge computing and blockchain technology to enable PFL through distributed edge clusters. Unlike traditional methods that rely on a single global model, our approach utilizes a multi-model approach with group-level client associations to achieve personalization. Blockchain technology is employed to securely store these models in a distributed manner with smart contracts to improve efficient client selection and clustering through the evaluation of client relationships. The proposed model utilizes a consortium blockchain to ensure that the participating edge clusters are preselected based on their trustworthiness, thereby enhancing the security and reliability of the overall system. Consortium

blockchain guarantees the authenticity of the transactions and mitigates the risk of eavesdropping, tampering, and compromising of the BPFL network edge servers by malicious clients. To verify the validity of the transaction, we employed the proof of training quality concept, where the accuracy of the local model is the primary verification parameter [35]. Moreover, edge computing servers were utilized to reduce communication and computation costs by providing local storage, communication, and computation capabilities, allowing the computational processing to be conducted closer to clients as data providers. As illustrated in Figure 1, the proposed model comprises the steps of system initialization, collaborative edge cluster establishment, personalized local training models, and personalized global model aggregation. The specific procedures are explained in further detail as follows.

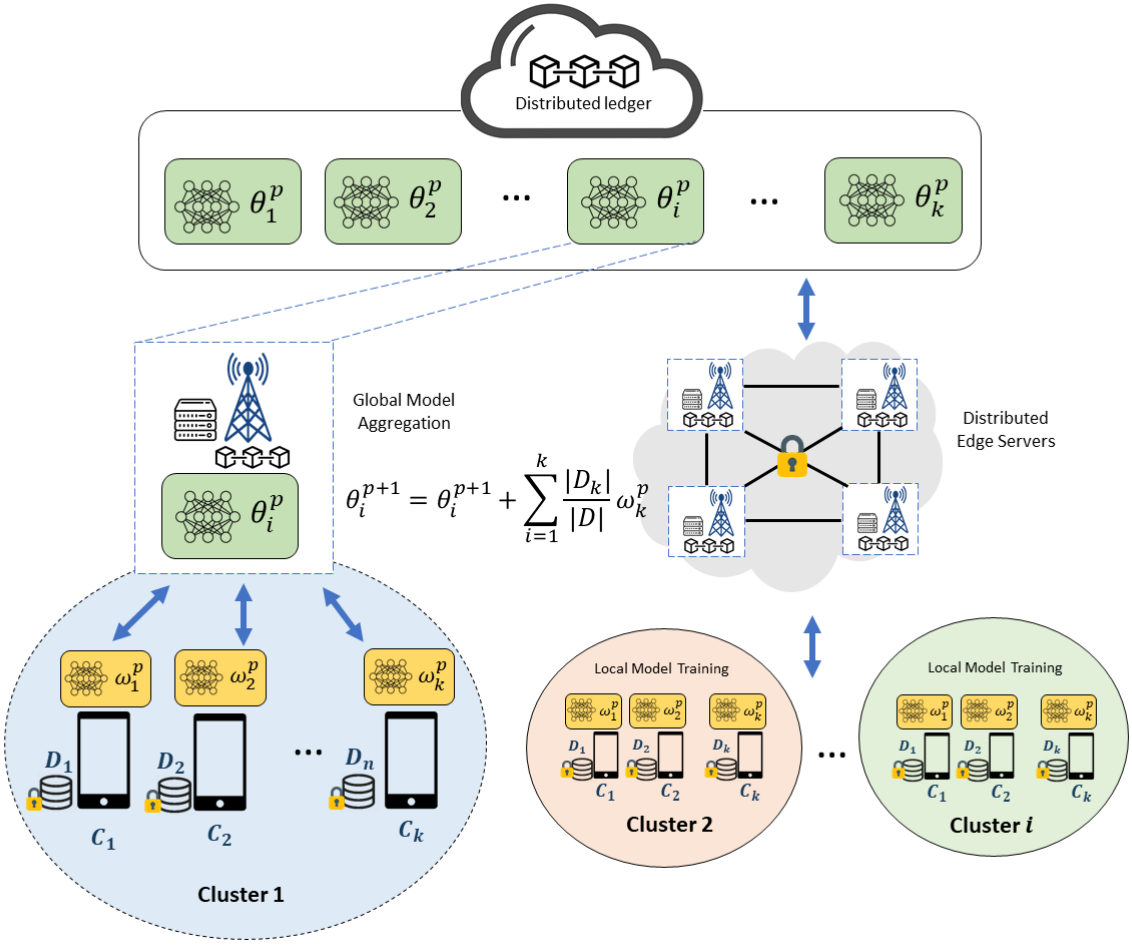


Figure 1. Distributed edge cluster for personalized federated learning.

4.1. System initialization

The initial model parameters of the global model (θ^{in}) are stored in a distributed ledger blockchain (which can be integrated with off-chain storage, e.g., the InterPlanetary File System (IPFS)) as the initial learning model process. It is maintained by edge servers that are placed throughout clusters.

Then, the number of K clients (C_k) need to register to the BPFL system by sending asset statement (refers to [36]) validation with their pseudo-public key address Pub_C^{ps} through Eq (4.1).

$$\begin{aligned} Tx\_C_{k(Asset)} = Pub_{C_k}^{ps} \rightarrow \Big\{\Big(Pub_{C_k\_Asset} = Pub^{Hash(C_k\_Asset)}, \\ \beta_j\_C_k = \left(Hash_{(j)}.Pub^{Hash(C_k\_Asset)}\right)^{Hash(C_k\_Asset)}\Big), \text{"}Asset\_sum\text{"}\Big\}, \\ where, Pub_{C_k}^{ps} \in \{{Pub_1}^{Sec\_C}, {Pub_2}^{Sec\_C}, ..., {Pub_k}^{Sec\_C}\} \end{aligned} \tag{4.1}$$

Equation (4.1) describes the process of recording a client's asset information in the system, which includes the client's public key related to the asset pseudonym $Pub_{C_k\_Asset}$, as evidence that the client owns the asset $\beta_j\_C_k$, and general asset information "$Asset\_sum$" (e.g., its format, topic, and data size). The public key and ownership proof are essential for verifying that the client owns the asset and keeping the asset information anonymous. These elements are created using a secure hash function $Hash(C_k\_Asset)$ to map the client's assets into unique public and private keys ${Pub_k}^{Sec\_C}$ to maintain the client's anonymity. In short, only validated clients can download θ^{in} and access the BPFL system.

4.2. Collaborative edge cluster establishment and personalized local training models

In this step, BPFL can distinguish incongruent clients after converging to a stationary point [10] as well as generate a multi-model method (θ_k) that is maintained by distributed edge servers. Hence, according to the client's asset statement, the collaborative edge clusters (ECl_i) are established based on the client's data distribution (D_k) similarity, where their datasets are non-IID. In this sense, BPFL splits the clients into two clusters (ECl_1, ECl_2) in a manner that optimizes the minimization of the maximum similarity between clients belonging to different clusters (see Eq (4.2)). Additionally, the group-level client similarity (δ) can be counted using specific partition techniques, such as Euclidean distance, cosine similarity, and Gaussian mixture. Our model relies on the cosine similarity technique (see Eq (4.3)).

$$ECl_1, ECl_2 \leftarrow \arg \min_{ECL_1 \cup ECL_2} (\max \delta_{ECl_1, ECl_2}) \tag{4.2}$$

$$\delta_{C_1,C_2} := \delta(w_1^p, w_2^p) := \frac{(w_1^p, w_2^p)}{\left\|w_1^p\right\| \left\|w_2^p\right\|} \tag{4.3}$$

In the $p-$th iteration, C_k in each ECl_i downloads global parameter θ_i^p from the blockchain and performs local training using their datasets D_k to obtain a personalized model (w_k^p) owned by client C_k by using the following formula [37]:

$$w_k^p = \arg \min_{w \in \mathbb{R}^d} F_i(w) + \frac{\mu}{2\alpha_p} \|w - \theta_k^p\| \tag{4.4}$$

where F_i is a loss function associated with the $i-$th data point on C_k; $w \in \mathbb{R}^d$ encodes the parameters of local model w, μ represents the regularization parameter, and α_p is the step size of gradient descent in iteration p.

Algorithm 1 A summary of the BPFL algorithm. D^i is the local datasets; K number of clients C_k; number of iteration, p; E is the number of local epochs; and η is learning rate.

procedure EdgeServerUpdate:
  ECL_i initialize θ^{in} **the initial models are stored in the blockchain*
  for each round p = 1,2, ..., P **do**
    $C_K \leftarrow max(Tx\_C_{k(Asset)})$
    $ECl_1, ECl_2 \leftarrow \arg\min_{ECL_1 \cup ECL_2}(\max \delta_{ECl_1, ECl_2})$ **similarity-based clients clustering*
    for each client $i \in C_K$ **in parallel do**
      $w_k^p \leftarrow$ UserUpdate(i, w^p) **$\forall$ updates the personalized model*
      $\theta_i^{(p+1)} \leftarrow \theta_i^p + \sum_{i=1}^{k} \frac{|D_k|}{|D|} w_k^p$ **aggregating the gathered models*
    end for
  end for
end procedure
procedure ClientUpdate:(i, w^p)
  //Executes on client i
  $w_k^p \leftarrow w^p$
  for each local epoch j from 1 to E **do**
    for each batch $b = \{x, y\}$ of D^i **do**
      $w_k^p \leftarrow \arg\min_{w \in \mathbb{R}^d} F_i(w) + \frac{\mu}{2\alpha_p} \|w - \theta_k^p\|$; **$\forall$ local training of personalized model*
    end for
  end for
  return w_k^p for aggregation
end procedure
procedure Incentive_Mechanism($Incv\_C_k$) **incentivized using Ethereum platform*
  ECL_i collects the list of participating clients $C_1, C_2, ..., C_k$
  for $C_1, C_2, ..., C_k$; ECL_i **do**
    $ECL_i \leftarrow$ ConfirmTransaction $H(w_1^p, w_2^p, ..., w_k^P)$ **ECL_i has the list of clients*
    $Incv\_C_k$ are given to $C_1, C_2, ..., C_k$ **the rewards are distributed to the clients*
  end for
end procedure

4.3. Personalized global model aggregation and incentive distribution

After the clients collaboratively upload their w_k^p to the distributed edge cluster, the cluster global model aggregation is executed as follows:

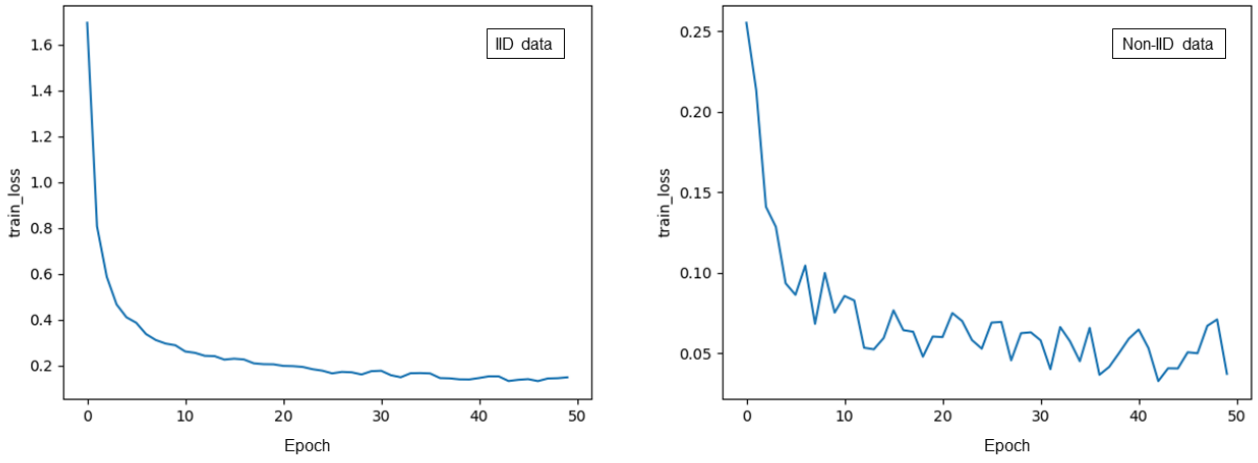
$$\theta_i^{(p+1)} = \theta_i^p + \sum_{i=1}^{k} \frac{|D_k|}{|D|} w_k^p \tag{4.5}$$

where $\theta_i^{(p+1)}$ is a new personalized global model obtained for the next iteration $(p + 1)$. In this step, all participants in the BPFL system can download $\theta_i^{(p+1)}$ through the distributed ledger blockchain. Consequently, iterations continue until the model achieves precise accuracy or reaches the maximum number of iterations. Finally, clients are incentivized through the implementation of a smart contract

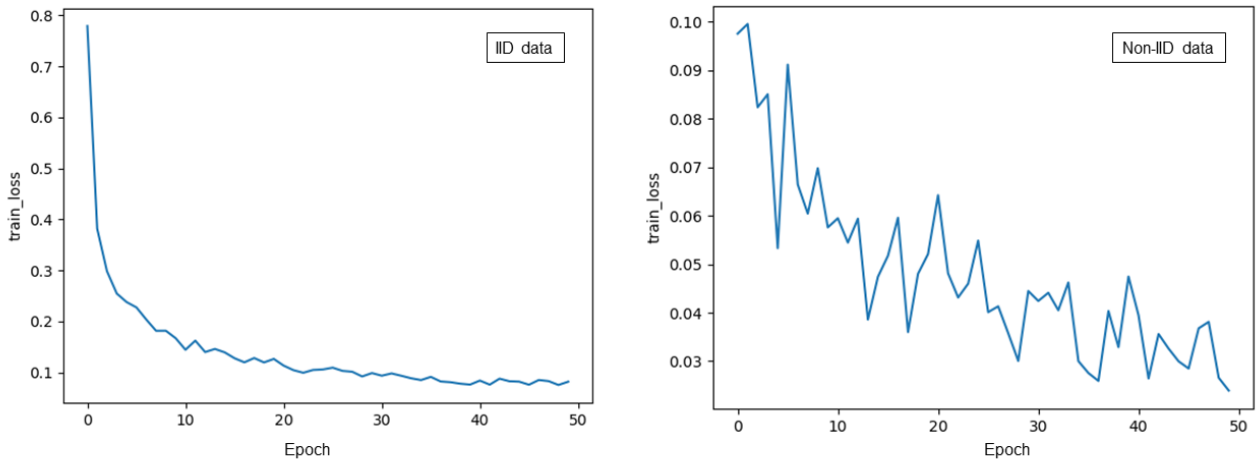
blockchain, whereby they are rewarded for fulfilling the transaction requirements of the BPFL platform. The incentive, represented by $Incv\_C_k$, is calculated based on the following formula [38]:

$$Incv\_C_k = \sum_i contrib_{C_k} \cdot \theta_i^{(p+1)} \tag{4.6}$$

where $contrib_{C_k}$ is the contribution of client C_k to $\theta_i^{(p+1)}$. Later, $Incv\_C_k$ is automatically distributed to participating clients once $\theta_i^{(p+1)}$ has been generated, thus providing a decentralized and responsive system. A summary of the BPFL framework can be found in Algorithm 1.



a) CNN model with the learning rate of 0.01



b) MLP model with the learning rate of 0.01

Figure 2. Training loss of 50 epochs for MLP and CNN models.

5. Numerical results and discussion

5.1. Implementation

This section describes the implementation of the proposed model which forms a distributed edge clustering framework to realize PFL by leveraging the merits of blockchain and EC. The modified National Institute of Standards and Technology (MNIST) [39] datasets were used with 10.000 images for the test set and 60.000 images for the training set. Our preliminary research compared the performance of IID and non-IID data distributions among clients through two different models of MNIST, i.e., a basic two-layer multilayer perceptron (MLP; also known as MNIST 2NN) and a convolutional neural network (CNN) comprising two 5×5 convolution layers. Figure 2 shows the loss from 50 epochs of training with a learning rate of 0.01 for both MLP and CNN models. The loss in the non-IID setting is higher than that of IID because of the unbalanced and highly heterogeneous nature of client data. The impact of non-IID data is further reflected in poor convergence and a decline in performance accuracy, as depicted in Figure 3. The accuracy is observed to be at its lowest at epoch 5, reaching 61.48%, as opposed to an accuracy of 95.03% in the IID setting.

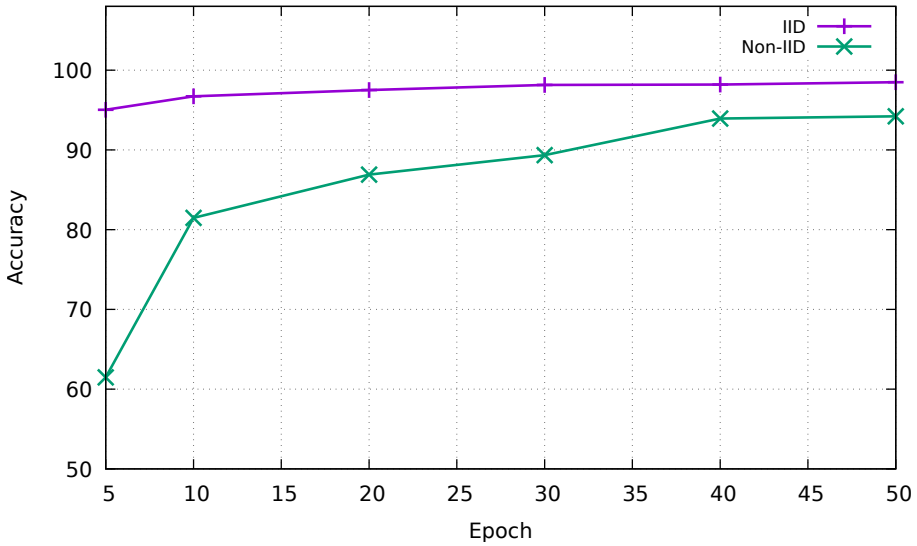


Figure 3. Accuracy performance results on IID and Non-IID setting.

To create a BPFL framework that leverages blockchain technology, we adopted a consortium setting [40] that harnesses the power of blockchain to perform decentralized PFL transactions, assess participant contributions to the global model with transparency, and establish a decentralized incentive system. The distribution of edge cluster contributions towards generating the global PFL model based on the Ethereum platform is depicted in Figure 4. We designed three distinct collaborative ECL_i edge clusters, ECL_1, ECL_2, and ECL_3, which work collaboratively to train FL models using their personalized models and local datasets. Upon creating a new personalized global model, the incentive is distributed to the edge clusters based on the recorded contribution of each cluster in the blockchain's distributed ledger. Figure 5 shows that, on average, BPFL achieves better performance accuracy than existing works. In summary, BPFL seeks to motivate clients possessing

heterogeneous data to actively participate in preserving PFL and improve system performance.

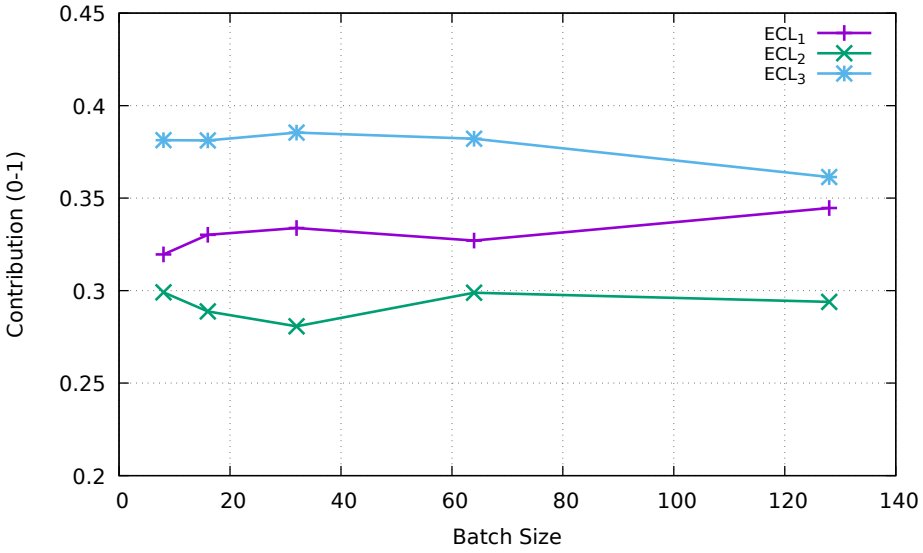


Figure 4. Distribution of edge cluster contributions towards generating the global model PFL based on Ethereum platform.

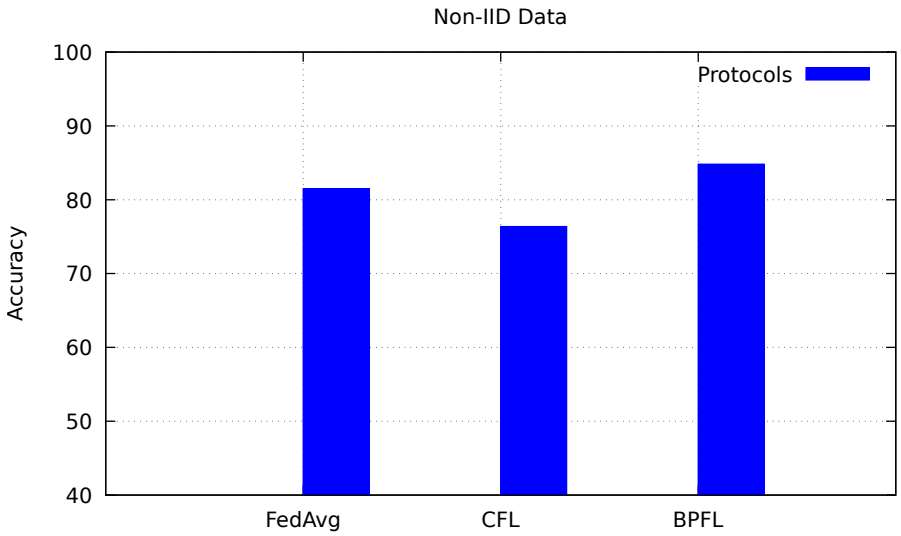


Figure 5. Performance accuracy comparison.

5.2. Discussion

The study of PFL is the latest trend addressing the issue of statistical heterogeneity of non-IID data distributions, which is one of the primary challenges of the existing FL approach. Even though the PFL approach is advantageous compared to the general FL framework of FedAvg, there are some significant challenges, especially regarding the data privacy risk of clients during the model training

process. Therefore, to enhance security and privacy in PFL, we introduced BPFL, which exploits the concept of distributed edge clusters to leverage the merits of edge computing and blockchain technology. Table 1 provides a theoretical summary of the advantages of BPFL compared to current PFL clustering techniques.

Table 1. Summary of BPFL advantages compared to current PFL techniques.

Key parameters	Smith et al. [9]	Sattler et al. [10]	This work
Computation and communication cost	High	High	Low
Works on non-convex setting	No	Yes	Yes
Multiple global model setting	Not investigated	Not investigated	Yes
Distributed edge cluster	Not applied	Not applied	Yes
Privacy guarantees	Not investigated	Yes	Yes
Asset statement	Not applied	Not applied	Yes
Incentive mechanism	Not applied	Not applied	Applied

Nevertheless, to strengthen and achieve a robust PFL technique, further work is required to investigate the potential attacks and defenses originating from the more complex protocols and structures of PFL. To protect the client's sensitive data from various threats, several privacy techniques might be leveraged, such as differential privacy, homomorphic encryption, secure multiparty computation, and a trusted execution environment. In addition, the development of a reliable incentive mechanism can be studied to maintain fairness and motivate client contributions. The scalability of blockchain is another aspect that requires attention. The main issue with a fully distributed ledger network is that every node must agree on the complete state of the ledger, which leads to challenges such as limited scalability and longer transaction delays as the network expands. However, representative datasets are essential for developing the PFL field. Datasets with more modalities (sensor signals, video, and audio) and involving a more comprehensive assortment of machine learning tasks from practical applications are required to further PFL research. Additionally, performance benchmarking is another critical aspect for the long-term expansion of the PFL research domain.

6. Conclusions

This study introduced a blockchain-enabled distributed edge cluster approach for PFL, exploiting the benefits of blockchain and edge computing. Blockchain protects client privacy and security by recording all transactions in immutable distributed ledger networks, thereby enhancing efficient client selection and clustering. Similarly, an edge computing system offers appropriate storage and computation, whereby computational processing is locally performed in the edge infrastructure to be nearer to clients. Thus, this system provides proper computation capability and improves real-time services and low-latency communication of PFL. As challenges for future study directions in the PFL field, privacy-preserving, trustworthy PFL, as well as representative datasets and benchmarks are suggested.

Acknowledgements

This research was supported by the MSIT (Ministry of Science and ICT), Republic of Korea, under the ITRC (Information Technology Research Center) support program (IITP-2022-2020-0-01797) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation) and partially supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2021R1I1A3046590).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. P. Voigt, A. Von dem Bussche, The eu general data protection regulation (gdpr), *A Practical Guide, 1st Ed., Cham: Springer International Publishing*, **10** (2017), 5510–5555. https://doi.org/10.1007/978-3-319-57959-7
2. G. J. Annas, Hipaa regulations: a new era of medical-record privacy?, *New England J. Med.*, **348** (2003), 1486. https://doi.org/10.1056/NEJMlim035027
3. A. Hard, K. Rao, R. Mathews, S. Ramaswamy, F. Beaufays, S. Augenstein, Het al., Federated learning for mobile keyboard prediction, preprint, arXiv:1811.03604.
4. N. Rieke, J. Hancox, W. Li, F. Milletari, H. R. Roth, S. Albarqouni, et al., The future of digital health with federated learning, *NPJ Dig. Med.*, **3** (2020), 119. https://doi.org/10.1038/s41746-020-00323-1
5. X. Wang, S. Garg, H. Lin, G. Kaddoum, J. Hu, M. S. Hossain, A secure data aggregation strategy in edge computing and blockchain-empowered internet of things, *IEEE Int. Things J.*, **9** (2020), 14237–14246. https://doi.org/10.1109/JIOT.2020.3023588
6. X. Wang, S. Garg, H. Lin, G. Kaddoum, J. Hu, M. M. Hassan, Heterogeneous blockchain and ai-driven hierarchical trust evaluation for 5g-enabled intelligent transportation systems, *IEEE Trans. Intell. Transp. Syst.*, **24** (2021), 2074–2083. https://doi.org/10.1109/TITS.2021.3129417
7. G. Long, Y. Tan, J. Jiang, C. Zhang, Federated learning for open banking, in *Federated Learning: Privacy and Incentive*, (2020), 240–254. https://doi.org/10.1007/978-3-030-63076-8_17
8. B. McMahan, E. Moore, D. Ramage, S. Hampson, B. A. y Arcas, Communication-efficient learning of deep networks from decentralized data, in *Artificial intelligence and statistics*, (2017), 1273–1282.
9. V. Smith, C. K. Chiang, M. Sanjabi, A. S. Talwalkar, Federated multi-task learning, in *Advances in neural information processing systems*, (2017).
10. F. Sattler, K. R. Müller, W. Samek, Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints, *IEEE Trans. Neural Networks Learning Syst.*, **32** (2020), 3710–3722. https://doi.org/10.1109/TNNLS.2020.3015958

11. Y. C. Hu, M. Patel, D. Sabella, N. Sprecher, V. Young, Mobile edge computing—a key technology towards 5g, *ETSI White Paper*, **11** (2015), 1–16.
12. E. Fazeldehkordi, T. M. Grønli, A survey of security architectures for edge computing-based iot, *IoT*, **3** (2022), 332–365. https://doi.org/10.3390/iot3030019
13. K. Cao, Y. Liu, G. Meng, Q. Sun, An overview on edge computing research, *IEEE Access*, **8** (2020), 85714–85728. https://doi.org/10.1109/ACCESS.2020.2991734
14. K. B. Letaief, Y. Shi, J. Lu, J. Lu, Edge artificial intelligence for 6g: Vision, enabling technologies, and applications, *IEEE J. Selected Areas Commun.*, **40** (2021), 5–36. https://doi.org/10.1109/JSAC.2021.3126076
15. E. Kristiani, Y. T. Tsan, P. Y. Liu, N. Y. Yen, C. T. Yang, Binary and multi-class assessment of face mask classification on edge ai using cnn and transfer learning, *Human Centric Comput. Inf. Sci.*, **12** (2022). https://doi.org/10.22967/HCIS.2022.12.053
16. M. Babar, M. S. Khan, U. Habib, B. Shah, F. Ali, D. Song, Scalable edge computing for iot and multimedia applications using machine learning, *Human centric Comput. Inf. Sci.*, **11** (2021). https://doi.org/10.22967/hcis.2021.11.041
17. B. He, T. Li, An offloading scheduling strategy with minimized power overhead for internet of vehicles based on mobile edge computing, *J. Inf. Process. Syst.*, **17** (2021), 489–504. https://doi.org/10.3745/JIPS.01.0077
18. Y. He, Z. Tang, Strategy for task offloading of multi-user and multi-server based on cost optimization in mobile edge computing environment, *J. Inf. Process. Syst.*, **17** (2021), 615–629. https://doi.org/10.3745/JIPS.01.0078
19. Z. Zhou, X. Chen, E. Li, L. Zeng, K. Luo, J. Zhang, Edge intelligence: Paving the last mile of artificial intelligence with edge computing, *Proc. IEEE*, **107** (2019), 1738–1762. https://doi.org/10.1109/JPROC.2019.2918951
20. X. Zhu, H. Li, Y. Yu, Blockchain-based privacy preserving deep learning, in *International Conference on Information Security and Cryptology*, (2018), 370–383. https://doi.org/10.1007/978-3-030-14234-6_20
21. T. S. Brisimi, R. Chen, T. Mela, A. Olshevsky, I. C. Paschalidis, W. Shi, Federated learning of predictive models from federated electronic health records, *Int. J. Med. Inf.*, **112** (2018), 59–67. https://doi.org/10.1016/j.ijmedinf.2018.01.007
22. S. Samarakoon, M. Bennis, W. Saad, M. Debbah, Federated learning for ultra-reliable low-latency v2v communications, in *2018 IEEE Global Communications Conference (GLOBECOM)*, (2018), 1–7. https://doi.org/10.1109/GLOCOM.2018.8647927
23. M. Chen, R. Mathews, T. Ouyang, F. Beaufays, Federated learning of out-of-vocabulary words, preprint, arXiv:1903.10635.
24. S. P. Karimireddy, S. Kale, M. Mohri, S. J. Reddi, S. U. Stich, A. T. Suresh, Scaffold: Stochastic controlled averaging for federated learning, in *International Conference on Machine Learning*, (2020), 5132–5143.

25. T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, V. Smith, Federated optimization in heterogeneous networks, in *Proceedings of Machine learning and systems*, (2020), 429–450.
26. J. Wang, Q. Liu, H. Liang, G. Joshi, H. V. Poor, Tackling the objective inconsistency problem in heterogeneous federated optimization, in *Advances in neural information processing systems*, (2020), 7611–7623.
27. M. Firdaus, K. H Rhee, A joint framework to privacy-preserving edge intelligence in vehicular networks, in *Information Security Applications: 23rd International Conference, WISA 2022*, Cham: Springer Nature Switzerland, (2023), 156–167. https://doi.org/10.1007/978-3-031-25659-2_12
28. J. S. P. G. M. Nam, J. G. Shon, A blockchain-based cheating detection system for online examination, *KIPS Trans. Software Data Eng.*, **11** (2022), 267–272. https://doi.org/10.3745/KTSDE.2022.11.6.267
29. A. Z. Tan, H. Yu, L. Cui, Q. Yang, Towards personalized federated learning, *IEEE Trans. Neural Networks Learn. Syst.*, (2022), 1–17. https://doi.org/10.1109/TNNLS.2022.3160699
30. K. C. Sim, P. Zadrazil, F. Beaufays, An investigation into on-device personalization of end-to-end automatic speech recognition models, preprint, arXiv:1909.06678.
31. V. Kulkarni, M. Kulkarni, A. Pant, Survey of personalization techniques for federated learning, in *2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)*, (2020), 794–797. https://doi.org/10.1109/WorldS450073.2020.9210355
32. Y. Jiang, J. Konečnỳ, K. Rush, S. Kannan, Improving federated learning personalization via model agnostic meta learning, preprint, arXiv:1909.12488.
33. Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, V. Chandra, Federated learning with non-iid data, preprint, arXiv:1806.00582.
34. Y. Chen, X. Qin, J. Wang, C. Yu, W. Gao, Fedhealth: A federated transfer learning framework for wearable healthcare, *IEEE Intell. Syst.*, **35** (2020), 83–93. https://doi.org/10.1109/MIS.2020.2988604
35. Y. Tian, T. Li, J. Xiong, M. Z. A. Bhuiyan, J. Ma, C. Peng, A blockchain-based machine learning framework for edge services in iiot, *IEEE Trans. Ind. Inf.*, **18** (2021), 1918–1929. https://doi.org/10.1109/TII.2021.3097131
36. J. Weng, J. Weng, J. Zhang, M. Li, Y. Zhang, W. Luo, Deepchain: Auditable and privacy-preserving deep learning with blockchain-based incentive, *IEEE Trans. Dependable Secure Comput.*, **18** (2019), 2438–2455. https://doi.org/10.1109/TDSC.2019.2952332
37. Y. Huang, L. Chu, Z. Zhou, L. Wang, J. Liu, J. Pei, et al., Personalized cross-silo federated learning on non-iid data, in *Proceedings of the AAAI Conference on Artificial Intelligence*, (2021), 7865–7873. https://doi.org/10.1609/aaai.v35i9.16960
38. X. Wang, S. Garg, H. Lin, M. J. Piran, J. Hu, M. S. Hossain, Enabling secure authentication in industrial iot with transfer learning empowered blockchain, *IEEE Trans. Ind. Inf.*, **17** (2021), 7725–7733. https://doi.org/10.1109/TII.2021.3049405

39. Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, *Proc. IEEE*, **86** (1998), 2278–2324. https://doi.org/10.1109/5.726791

40. M. Firdaus, H. T. Larasati, K. H. Rhee, A secure federated learning framework using blockchain and differential privacy, in *2022 IEEE 9th International Conference on Cyber Security and Cloud Computing (CSCloud)/2022 IEEE 8th International Conference on Edge Computing and Scalable Cloud (EdgeCom)*, (2022), 18–23.

AIMS Press