

# ARCHITECTURAL BLUEPRINT FOR HETEROGENEITY-RESILIENT FEDERATED LEARNING

*Satwat Bashir, Tasos Dagiuklas, Kasra Kassai, Muddesar Iqbal*

<sup>1</sup>Department of Computer Science, London South Bank University, London, UK

\*E-mail: {bashis11, tdagiuklas, kasra.kassai, m.iqbal}@lsbu.ac.uk

**Keywords:** FEDERATED LEARNING, EDGE COMPUTING, CONVERGENCE, NON-IID, MULTI-GLOBAL MODELS

## Abstract

This paper proposes a novel three-tier architecture for federated learning to optimize edge computing environments. The proposed architecture addresses the challenges associated with client data heterogeneity and computational constraints. It introduces a scalable, privacy-preserving framework that enhances the efficiency of distributed machine learning. Through experimentation, the paper demonstrates the architecture's capability to manage non-IID data sets more effectively than traditional federated learning models. Additionally, the paper highlights the potential of this innovative approach to significantly improve model accuracy, reduce communication overhead, and facilitate broader adoption of federated learning technologies.

## 1 Introduction

In the digital era, the proliferation of data from the Internet of Things (IoT) and smart devices presents unprecedented opportunities and challenges. Edge computing has become an essential paradigm, processing data proximal to its origin, thereby reducing latency and bolstering privacy. Nonetheless, the integration of Federated Learning (FL) with edge computing encounters significant obstacles, such as high communication costs and data/model heterogeneity, especially with non-Independently and Identically Distributed (non-IID) data [1–3].

FL emerges as a revolutionary method, processing data locally on devices while sharing model updates instead of raw data, thus significantly diminishing bandwidth demands and fortifying privacy. Yet, FL confronts notable challenges, including substantial communication costs and the complexity of data/model heterogeneity across disparate devices [4]. The data collected from various devices often exhibit non-IID characteristics, meaning that the data points are neither independent nor identically distributed across the network. This variation poses significant challenges for traditional learning models, using edge-cloud computing, where data are processed close to its source [5, 6]. Therefore, underscores the need for innovative solutions tailored to the demands of edge computing environments [7].

To address the nuanced challenges presented by edge computing and non-IID data, a novel three-layered architecture for FL has been introduced. This design seamlessly integrates *clients*, *edge layers*, and *fedge layer* to enhance data processing efficiency and model efficacy. By defining clear roles and interactions among the layers, the proposed architecture provides a focused strategy for addressing data and computational heterogeneity. The architecture introduces aggregation across

two layers and multi-global models, allowing for the management of distinct models and streamlining aggregation to better address the concerns of data diversity [8, 9].

This paper is structured as follows: Section 2 provides a background on FL, discussing the multifaceted nature of heterogeneity and its challenges. Section 3 presents the theoretical considerations for improved FL frameworks, introducing our novel three-tier architecture. Section 4 delves into empirical findings on FL with non-IID data, showcasing the architecture's performance through various scenarios. Section 5 concludes the study with a discussion on the implications of our findings and outlines future directions for research. Throughout, we aim to highlight the innovative aspects of our approach and its potential impact on the field of FL in edge computing environments.

## 2 Background on Federated Learning

The advent of FL represents a monumental shift towards decentralized model training across a myriad of devices, pivotal in the era dominated by IoT. This approach allows for the harnessing of vast amounts of data directly at the source, thus preserving user privacy and leveraging computational capabilities without the need for data centralization [10]. Despite its potential, the practical deployment of FL is fraught with challenges, predominantly due to data heterogeneity, model variance, and the diverse capabilities of devices, which collectively demand innovative solutions for effective global model development [3].

### 2.1 The Multifaceted Nature of Heterogeneity

In FL, heterogeneity manifests in several critical areas: data, model, and device, each introducing unique challenges to the learning process. Data heterogeneity, particularly with

non-Independently and Identically Distributed (non-IID) data across devices, leads to significant discrepancies in local model accuracy and global model performance [11, 12]. This variance underscores the difficulty of creating models that generalize well across diverse data distributions, a fundamental goal of FL. Model heterogeneity, stem from the diverse computational capabilities of devices, influencing the choice of model architectures and thus impacts the coherence of aggregated global models [13]. Furthermore, device heterogeneity, is characterized by disparities in computational power, memory, and connectivity, directly affects FL's efficiency and scalability [14, 15].

Adding to these complexities, are the challenges posed by communication heterogeneity, which impacts the efficiency and reliability of data transmission, crucial for collaborative learning in distributed environments. These multifaceted aspects of heterogeneity highlight the intricate challenges FL faces, necessitating comprehensive strategies that address not only the technical hurdles, but also the ethical considerations inherent in deploying FL in real-world scenarios [16, 17].

## 2.2 Tackling Non-IID Data Challenges

Non-IID data significantly impedes FL's model performance and learning efficiency. To counteract this, innovative approaches such as Personalized Federated Learning (PFL) have been proposed, focusing on architecture-based and similarity-based personalization [18]. Techniques like parameter decoupling and knowledge distillation aim to tailor models to individual device capabilities while drawing insights from the global model [19, 20]. Despite these advancements, achieving equitable and efficient learning in federated settings remains a formidable challenge, highlighting the need for further innovation [21].

## 2.3 Critical Analysis of Aggregation Methods

At the heart of FL lies the model aggregation process, with FedAvg being the prototypical method [22–24]. While FedAvg has shown effectiveness under conditions of homogeneity, its performance significantly diminishes in the face of data heterogeneity, leading to a loss of precision in local models [18, 25].

In the study referenced as [26], the authors evaluate the impact of utilizing pre-trained models on local devices, focusing on the aggregation of shared parameters. They find that employing pre-trained models and sharing a subset of parameters from the start can enhance convergence speed, improve performance, and lower communication costs. Similarly, the work cited as [27] introduces pFedGP, which utilizes a shared kernel function to increase the accuracy of local models. However, this approach is constrained by the size of the dataset due to the limitations of the shared kernel function. To tackle the issue of communication overhead in personalized model training, the study [28] applies soft quantization and transfer learning techniques, achieving higher accuracy for local devices and enhancing communication efficiency. Likewise,

[29] employs transfer learning to categorize devices with similar data and facilitate collaborative training for varied data sets.

Other research, such as [30], focuses on incorporating personalized layers through fine-tuning techniques to train the global model's shallow layers on each device. This method involves sharing only the foundational layers during the aggregation process, but it faces challenges related to storing the personalized model on each local device. Some studies, like [31], address the issue of data heterogeneity by framing it as a regularization problem, utilizing Moreau Envelopes to improve the global model's convergence. However, a notable limitation of these approaches is their reliance on numerous adjustable parameters.

Despite these innovative strategies, limitations remain, particularly in handling complex real-world data distributions, ensuring communication efficiency, and accommodating diverse scenarios. The pursuit of more adaptable and robust solutions continues, reflecting the evolving landscape of FL research [32, 33].

## 2.4 Towards Comprehensive Solutions

The exploration of aggregation methods and personalization strategies within FL reveals a concerted effort to address the challenges posed by heterogeneity and non-IID data as demonstrated in studies [10, 21]. However, the limitations of current approaches underscore the pressing need for adaptable, sophisticated solutions capable of navigating the intricacies of FL environments. The shift towards advanced edge-cloud models, the employment of cutting-edge algorithms, and the development of multi-model systems illustrate the ongoing quest for enhanced FL frameworks [34]. Our findings lay the groundwork for a transformative FL ecosystem, characterized by its robustness, efficiency, and unparalleled scalability across myriad applications. Looking ahead, we explore advanced machine learning algorithms that could further refine the proposed architecture's ability to manage extreme data heterogeneity. Additionally, we aim to investigate the implications of our framework in real-world scenarios, such as IoT and mobile computing, where FL's potential can be fully unleashed.

## 3 Theoretical Considerations for Improved FL Frameworks

In the rapidly evolving field of FL, identifying architectures capable of handling the complexities of distributed data processing is critical. Traditional FL models have pioneered the way forward, but often grapple with limitations due to non-IID data distributions and significant computational demands on client devices. In response to these challenges, this paper proposes a novel three-tier FL architecture designed to enhance data privacy, computational efficiency, and scalability across distributed networks.

### 3.1 Rationale for a Multi-Global Model Framework

The proposed architecture significantly diverges from traditional FL paradigms by implementing a multi-global model strategy within a hierarchical framework. This approach is analyzed for its potential to facilitate personalized and efficient learning across heterogeneous devices and data distributions. The operational mechanics of each layer are shown in Figure 1, illustrating how they collectively contribute to model convergence and scalability. Preliminary simulations underscore the adaptability and performance of our architecture when compared to standard FL frameworks, particularly under non-IID conditions.

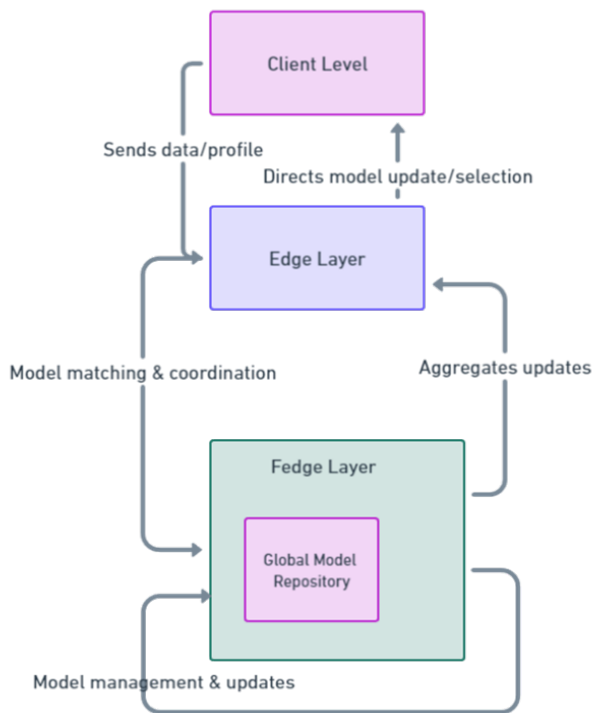


Fig. 1 Proposed three-layered architecture, delineating the interactions between the client level, edge layer, and fedge layer

**The Client Layer**, profiling ensures that each participant begins with a model specifically tailored to its unique data set. This enables effective local training, optimizing computational load and aligning model evolution with individual data characteristics. Empirical evidence demonstrates that this alignment results in improved initial accuracy and faster convergence rates.

**The Edge Layer** acts as a critical coordinator, analyzing metadata from client training sessions to assess model states and requirements. If a suitable model match for a client is not found within its repository, the edge layer queries the fedge layer. This process ensures that clients are always provided with the most appropriate model for their data, enhancing the learning process’s efficiency and personalization.

**The Fedge Layer** is responsible for storing multiple distinct global models, managing them to ensure they are up-to-date and accurately reflect the diverse data landscapes they represent. When requested by the edge layer, the fedge Layer searches for a suitable model match. If no match is found, it informs the edge layer, prompting the creation of a new model for the client. This new model is then saved as a global model, enriching the fedge Layer’s repository. Furthermore, the fedge Layer periodically updates its global models based on inputs from the edge layer, performing aggregation and updates of similar models to maintain a current and effective suite of global models.

This structured approach tackles computing issues and helps create better learning environments. We have carefully followed principles of technical precision in our work, making sure our approach, experiments, and results are shared accurately and clearly.

### 3.2 Innovation

The core innovation of the proposed architecture is its deployment of multiple global models, a strategy unprecedented in current FL frameworks. This supports personalized learning for diverse client datasets and introduces a novel method of model sharing and aggregation that significantly reduces computational overhead. The empirical findings affirm this architecture’s potential to redefine FL paradigms, offering a scalable solution for real-time data analysis.

## 4 Empirical Findings on Federated Learning with Non-IID Data

This section delves into the performance of FL, specifically focusing on the Federated Averaging (FedAvg) method, under non-IID data conditions. This investigation is supported by an experimental setup, revealing insights into FL’s adaptability to varied data distributions.

### 4.1 System Architecture and Experimental Setup

**Dataset:** The analysis employs the MNIST dataset, a benchmark for evaluating machine learning models through digit recognition tasks, featuring 70,000 grayscale images across ten classes [35].

### 4.2 Scenario Design

We aim to explore the multifaceted nature of non-IID distributions in a controlled yet challenging environment. For this, we designed two distinct non-IID scenarios using the MNIST dataset to test the FL architecture’s handling of data heterogeneity: By simplifying the scenario to these two configurations, we not only highlight the limitations of traditional FL model in handling such disparities, but also establish a clear, proof-of-concept framework for demonstrating the superior adaptability and performance of our proposed three-layer architecture. This approach allows to systematically assess the architecture’s efficacy in managing non-IID data, ensuring that

the findings have practical implications for enhancing FL in heterogeneous environments.

- **Generalizable Non-IID Scenario:** The MNIST dataset has been divided among three clients to simulate a non-IID but generalized scenario. An equal number of images has been allocated to each client. Such allocation provides restriction to distinct labels, ensuring diversity in data distribution while maintaining a level of generality across the dataset.
- **Non-Generalizable Non-IID Scenario:** For a more challenging setup, we have created another non-IID scenario by dividing the MNIST dataset between two clients. One client has received images from only two labels with a robust dataset size for training, while in the second client has been allocated images from the remaining eight labels, presenting a challenging test for the architecture’s efficiency.

**Client-Server and Three-Tier FL Models:** We compare the traditional client-server FL model with our proposed three-tier FL model, highlighting the latter’s enhanced data privacy, computational efficiency, and scalability.

### 4.3 Methodology and Model Implementation

**FL Algorithms and Model Details:** The FedAvg algorithm benchmarks the standard FL approach, while the proposed three-tier FL model introduces a specialized algorithm to better manage non-IID data through a multi-global model framework. The dynamic interactions and the structural foundation of the proposed three-tier FL model is shown in Figure 2. This sequence diagram elucidates the multi-global model strategy implemented within a hierarchical framework, showcasing the model’s innovative approach to managing non-IID data.

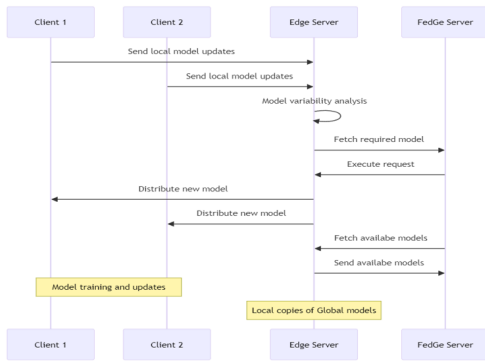


Fig. 2 The sequence diagram illustrates the dynamic interactions within the proposed architecture, focusing on the multi-global model strategy within a hierarchical framework

**Model Specification:** A Simple Convolutional Neural Network (SimpleCNN) has been chosen for consistency in evaluating both FL models against the MNIST dataset.

### 4.4 Parameters and Data Handling

**Parameters and Configurations:** The experimental setup involves multiple communication rounds, a learning rate of 0.01, momentum of 0.5, and a batch size of 64.

**Data Preprocessing:** Standard normalization processes has been applied to ensure uniform data scaling across clients.

**Model Serialization:** We detail the serialization process for both FL models, emphasizing the efficient transmission of model updates between clients, edge layers, and the global model repository.

### 4.5 Findings, Behaviors, and Insights

In the first scenario, which tested the models’ response to label skew with a balanced data distribution, the standard FL model showed a progressive increase in accuracy. Client 1’s accuracy began at 96.22% and rose to 99.36%. For Client 2, the accuracy started at 96.93% and reached 99.14%, and Client 3 improved from 94.82% to 98.90%. These trends are depicted in Figure 3.

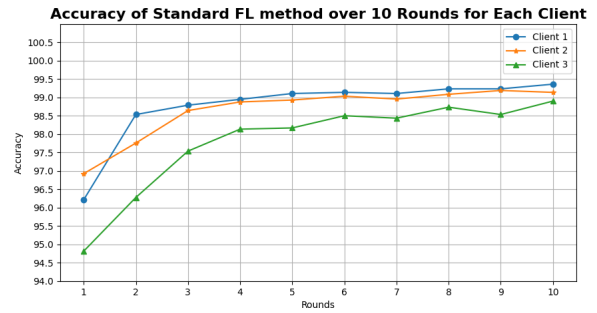


Fig. 3. Accuracy of Standard FL for non-IID Scenario 1

The three-tier FL model demonstrated a consistent high accuracy across all clients from the start, with Client 1 reaching 99.62%, Client 2 ending at 99.38%, and Client 3 at 99.47% by the tenth round. These findings indicate the three-tier model’s ability to quickly achieve and maintain high accuracy, as shown in Figure 4.

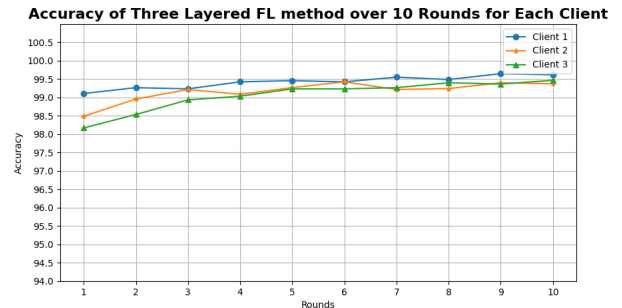


Fig. 4. Accuracy of Three Layered FL for non-IID Scenario 1

In the second scenario, designed to present a more complex non-IID situation with both label and data skew, the standard FL model and the three-tier FL model were subjected to a challenging test. For the standard FL model, Client 1 managed to maintain an accuracy of nearly 99.86%, while Client 2, with less representative data and a broader label distribution, started at a lower 14.5% accuracy, gradually increasing to 71.38% as shown in Figure 5. This growth trajectory signifies the model’s gradual adaptation to the non-IID conditions.

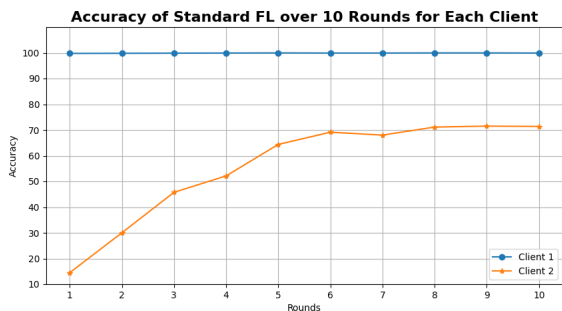


Fig. 5. Accuracy of Standard FL for non-IID Scenario 2

Conversely, the three-tier FL model excelled, with Client 1 maintaining a near-perfect accuracy of approximately 99.95%, and Client 2, despite the sparse data representation and more extensive label range, exhibited a substantial improvement, escalating from 29.25% to 84.88% accuracy. This remarkable progress is graphically represented in Figure 6.

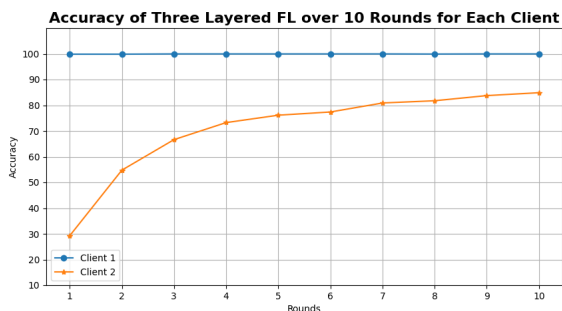


Fig. 6. Accuracy of Three Layered FL for non-IID Scenario 2

These results validate the standard FL model’s capacity to achieve generalization over time in balanced scenarios, but they also spotlight the three-tier FL model’s advanced capability to handle various degrees of non-IID data efficiently. The superior performance of the three-tier model, especially for Client 2 in the second scenario, underlines its effectiveness in real-world applications, where data distributions and label availability can significantly differ. The model’s robustness in enhancing learning under such complex conditions attests to its potential for broad deployment and underscores the value of the proposed FL technique.

## 5 Conclusion

This study introduced a novel three-layered architecture for FL, designed to address the significant challenges of integrating FL with edge computing, especially under conditions of non-IID data. Our architecture demonstrates a promising potential to enhance FL’s applicability in edge computing environments by optimizing training and processing efficiencies. By introducing an intermediary layer, the fedge layer, and employing a multi-global model framework, we provided a focused strategy for managing client and computational heterogeneity. Empirical evaluations, utilizing the MNIST dataset under various non-IID scenarios, have shown that our proposed architecture outperforms traditional FL frameworks in terms of accuracy, scalability, and efficiency.

However, it’s important to acknowledge the limitations of our current work. The proposed architecture, while innovative, serves as a foundational proof of concept. It is important to acknowledge that the architecture’s effectiveness, as demonstrated under controlled experimental conditions, may vary in complex real-world scenarios. The adaptation to different types of non-IID data, the management of communication overhead, and the practical deployment in edge environments with varied device capabilities and network conditions are aspects that require further exploration and validation.

In conclusion, this paper contributes a significant step forward in the pursuit of more efficient, privacy-preserving, and scalable FL solutions tailored for edge computing. The proposed three-layered architecture, with its emphasis on addressing non-IID data challenges and client heterogeneity, lays the groundwork for future advancements in the field. By continuing to refine and expand upon this framework, we can move closer to realizing the full potential of FL in the edge computing paradigm.

## 6 Future Directions

In future work, we plan to further refine and validate the effectiveness of our proposed three-layer architecture by testing it alongside the latest and most innovative federated learning approaches. This comparative analysis will provide valuable insights into where our model stands in the current research landscape and identify areas for improvement. Additionally, experimenting with other benchmark datasets beyond MNIST will allow us to assess the versatility and robustness of our architecture across different types of data and learning scenarios. This step is crucial for ensuring that our model can be effectively applied in various domains and with diverse data characteristics.

Furthermore, we aim to enhance the functionalities and operation of the fedge layer to better suit real-world scenarios. This includes optimizing the fedge layer for more efficient model management and updating processes, which are vital for handling the dynamic and diverse nature of data in practical applications. Another critical area of development will be the refinement of the client selection method. By designing more sophisticated criteria for client selection, we can ensure

that our architecture is adaptable to real and broader contexts, improving its applicability and efficiency in distributed learning environments.

By focusing on these areas, we anticipate not only strengthening the foundation of our proposed architecture but also significantly advancing its potential for practical implementation in federated learning systems.

## References

- [1] B. Tang, Y. Xiao, L. Zhang, B. Cao, M. Tang, and Q. Yang, "AFL-HCS: asynchronous federated learning based on heterogeneous edge client selection," *Cluster Computing*, 2024. [Online]. Available: <https://link.springer.com/article/10.1007/s10586-024-04314-9>
- [2] F. Zhou, S. Hu, X. Du, X. Wan, and J. Wu, "A Lightweight Neural Network Model for Disease Risk Prediction in Edge Intelligent Computing Architecture," *Future Internet*, vol. 16, no. 3, p. 75, 2024.
- [3] M. Ye, X. Fang, B. Du, P. C. Yuen, and D. Tao, "Heterogeneous federated learning: State-of-the-art and research challenges," *ACM Computing Surveys*, vol. 56, no. 3, pp. 1–44, 2023.
- [4] B. T. Hasan and A. K. Idrees, "Federated Learning for IoT/Edge/Fog Computing Systems," *arXiv e-prints*, 2024. [Online]. Available: <https://arxiv.org/abs/2402.13029>
- [5] Z. Zhang, X. Pei, and Y. Zhang, "Cost-efficient hierarchical federated edge learning for satellite-terrestrial Internet of Things," *Research Square*, 2024. [Online]. Available: <https://www.researchsquare.com/article/rs-3962040/latest.pdf>
- [6] P. Consul, I. Budhiraja, D. Garg, and N. Kumar, "Resource Allocation Approach For Digital Twin-Empowered UAV-Assisted MEC Network Using Federated Reinforcement Learning For Future Wireless Network," *IEEE Transactions*, 2024. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/10443270/>
- [7] M. Sharma and P. Kaur, "Fog-based Federated Time Series Forecasting for IoT Data," *Journal of Network and Systems Management*, 2024. [Online]. Available: <https://link.springer.com/article/10.1007/s10922-024-09802-2>
- [8] Z. Yang, S. Zhang, C. Li, M. Wang, and H. Wang, "Efficient knowledge management for heterogenous federated continual learning on resource-constrained edge devices," *Future Generation Computer Systems*, 2024.
- [9] H. Li, Y. Yang, Y. Liu, and W. Pei, "Federated Dueling DQN Based Microgrid Energy Management Strategy in Edge-Cloud Computing Environment," *Sustainable Energy, Grids and Networks*, 2024.
- [10] Y. N. Tan, V. P. Tinh, P. D. Lam, N. H. Nam, and T. A. Khoa, "A transfer learning approach to breast cancer classification in a federated learning framework," *IEEE Access*, vol. 11, pp. 27 462–27 476, 2023.
- [11] Q. Xia, W. Ye, Z. Tao, J. Wu, and Q. Li, "A survey of federated learning for edge computing: Research problems and solutions," *High-Confidence Computing*, vol. 1, no. 1, p. 100008, 2021.
- [12] R. Liu, M. Hu, Z. Xia, J. Xia, P. Zhang, Y. Huang, Y. Liu, and M. Chen, "Adapterfl: Adaptive heterogeneous federated learning for resource-constrained mobile computing systems," *arXiv preprint arXiv:2311.14037*, 2023.
- [13] L. Huang, A. L. Shea, H. Qian, A. Masurkar, H. Deng, and D. Liu, "Patient clustering improves efficiency of federated machine learning to predict mortality and hospital stay time using distributed electronic medical records," *Journal of Biomedical Informatics*, vol. 99, p. 103291, 2019.
- [14] C. Briggs, Z. Fan, and P. Andras, "Federated learning with hierarchical clustering of local updates to improve training on non-iid data," in *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020, pp. 1–9.
- [15] A. Ghosh, J. Hong, D. Yin, and K. Ramchandran, "Robust federated learning in a heterogeneous environment," *arXiv preprint arXiv:1906.06629*, 2019.
- [16] 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.
- [17] Y. Kim, E. Al Hakim, J. Haraldson, H. Eriksson, J. M. B. da Silva, and C. Fischione, "Dynamic clustering in federated learning," in *ICC 2021-IEEE International Conference on Communications*. IEEE, 2021, pp. 1–6.
- [18] A. Z. Tan, H. Yu, L. Cui, and Q. Yang, "Towards personalized federated learning," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.
- [19] M. Stallmann and A. Wilbik, "Towards federated clustering: A federated fuzzy c-means algorithm (ffcm)," *arXiv preprint arXiv:2201.07316*, 2022.
- [20] Q. Duan, J. Huang, S. Hu, R. Deng, Z. Lu, and S. Yu, "Combining federated learning and edge computing toward ubiquitous intelligence in 6g network: Challenges, recent advances, and future directions," *IEEE Communications Surveys & Tutorials*, 2023.
- [21] Y. J. Wong, M.-L. Tham, B.-H. Kwan, E. M. A. Gnana-muthu, and Y. Owada, "An optimized multi-task learning model for disaster classification and victim detection in federated learning environments," *IEEE Access*, vol. 10, pp. 115 930–115 944, 2022.
- [22] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," *arXiv preprint arXiv:1610.05492*, 2016.
- [23] 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*. PMLR, 2017, pp. 1273–1282.
- [24] S. Banabilah, M. Aloqaily, E. Alsayed, N. Malik, and Y. Jararweh, "Federated learning review: Fundamentals, enabling technologies, and future applications," *Information processing & management*, vol. 59, no. 6, p. 103061, 2022.
- [25] A. Shamsian, A. Navon, E. Fetaya, and G. Chechik, "Personalized federated learning using hypernetworks," in

- International Conference on Machine Learning*. PMLR, 2021, pp. 9489–9502.
- [26] K. Pillutla, K. Malik, A.-R. Mohamed, M. Rabbat, M. Sanjabi, and L. Xiao, “Federated learning with partial model personalization,” in *International Conference on Machine Learning*. PMLR, 2022, pp. 17 716–17 758.
- [27] I. Achituve, A. Shamsian, A. Navon, G. Chechik, and E. Fetaya, “Personalized federated learning with gaussian processes,” *Advances in Neural Information Processing Systems*, vol. 34, pp. 8392–8406, 2021.
- [28] K. Ozkara, N. Singh, D. Data, and S. Diggavi, “Quped: Quantized personalization via distillation with applications to federated learning,” *Advances in Neural Information Processing Systems*, vol. 34, pp. 3622–3634, 2021.
- [29] J. Zhang, S. Guo, X. Ma, H. Wang, W. Xu, and F. Wu, “Parameterized knowledge transfer for personalized federated learning,” *Advances in Neural Information Processing Systems*, vol. 34, pp. 10 092–10 104, 2021.
- [30] 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.
- [31] C. T. Dinh, N. Tran, and J. Nguyen, “Personalized federated learning with moreau envelopes,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 21 394–21 405, 2020.
- [32] Q. Xia, W. Ye, Z. Tao, J. Wu, and Q. Li, “A survey of federated learning for edge computing: Research problems and solutions,” *High-Confidence Computing*, vol. 1, no. 1, p. 100008, 2021.
- [33] 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,” *Foundations and Trends® in Machine Learning*, vol. 14, no. 1–2, pp. 1–210, 2021.
- [34] K. Kassai, S. Ghosh, and A. Dagiuklas, “Evolution of orchestration towards 5g,” *Journal of Communication*, vol. 14, no. 12, 2019.
- [35] HojjatK, “Mnist dataset,” <https://www.kaggle.com/datasets/hojjatk/mnist-dataset>, Year of Access, accessed: date-of-access.