10.48047/jocaaa.2024.33.07.113

MATHEMATICAL MODELING AND COMPUTATIONAL ANALYSIS OF FEDERATED LEARNING CONVERGENCE IN HETEROGENEOUS EDGE INTELLIGENCE SYSTEMS

[1]V Ravikumar,
Associate Professor in CSE,
ACE Engineering College,
Hyderabad,
ravi.vanoj@gmail.com

[2]N Srivani,
Assistant Professor in CSE,
CMR Institute of Technology,
Hyderabad,
vani.medipally@gmail.com

[3]Kasapaka Rubenraju,
Assistant Professor in CSE,
Malla Reddy University,
Hyderabad,
rubenrajk@gmail.com

Abstract

Federated Learning (FL) has emerged as an important distributed machine learning paradigm for training models across multiple devices while preserving data privacy. The growing adoption of edge intelligence systems, including Internet of Things (IoT) devices, mobile platforms, and edge servers, has increased interest in deploying federated learning within heterogeneous computing environments. However, differences in computational capabilities, communication resources, local data distributions, and participation rates among edge devices introduce challenges that influence convergence behavior and training efficiency. Understanding these factors is essential for developing scalable and reliable federated learning systems. This study investigates the mathematical modeling and computational analysis of federated learning convergence in heterogeneous edge intelligence environments. A convergence-oriented analytical framework is developed to examine the impact of client heterogeneity, communication delays, and local update frequencies on model performance. The framework incorporates mathematical representations of global aggregation and local optimization processes and evaluates convergence characteristics under varying system conditions. Experimental simulations were conducted using heterogeneous edge nodes with different processing capabilities and non-identically distributed datasets. Performance was assessed using convergence rounds, training loss, model accuracy, and communication overhead. The analysis indicates that increasing heterogeneity can slow convergence and increase communication requirements, while adaptive aggregation strategies help mitigate these effects. The results suggest that appropriate coordination between local computation and global synchronization can improve learning efficiency in distributed environments. This work contributes to the understanding of federated learning dynamics and provides insights for the design of efficient edge intelligence systems operating under realistic heterogeneous conditions.

Keywords— Federated Learning, Edge Intelligence, Convergence Analysis, Distributed Machine Learning, Edge Computing, Mathematical Modeling, Heterogeneous Systems.

I. INTRODUCTION

The rapid growth of connected devices and intelligent applications has transformed the landscape of distributed computing. Modern systems generate large volumes of data through mobile devices, sensors, autonomous systems, and Internet of Things (IoT) platforms. Traditionally, machine learning models have relied on centralized training approaches in which data collected from multiple sources are transferred to cloud servers for processing and analysis. Although this paradigm has achieved considerable success, it introduces concerns related to privacy, communication costs, scalability, and regulatory compliance. Federated Learning (FL) has been proposed as an alternative distributed learning framework that enables model training without requiring raw data to leave local devices [1]. Instead of transferring data to a central server, participating devices train local models and share only model updates. These updates are aggregated to construct a global model that benefits from distributed knowledge while preserving data privacy. This approach has attracted significant attention in applications involving healthcare, intelligent transportation, smart manufacturing, mobile services, and edge computing environments [2]. The increasing deployment of edge computing infrastructures has further accelerated interest in federated learning. Edge computing brings computational resources closer to data sources, reducing latency and supporting real-time processing requirements [3]. By combining federated learning with edge intelligence, organizations can develop

10.48047/jocaaa.2024.33.07.113

machine learning solutions that operate efficiently across geographically distributed systems while minimizing communication overhead.

Despite these advantages, federated learning introduces several technical challenges. One of the most significant challenges concerns convergence behaviors in heterogeneous environments. Unlike traditional distributed computing systems that often assume homogeneous computational resources, real-world edge intelligence systems consist of devices with varying processing capabilities, memory capacities, network conditions, and energy constraints. Such heterogeneity influences the speed and stability of model training and may reduce overall learning efficiency.

Another challenge arises from statistical heterogeneity. Data generated by edge devices are frequently non-identically distributed and may exhibit significant variations across participants. User behavior, environmental conditions, geographical factors, and application-specific characteristics often produce local datasets that differ substantially from one another. These differences can introduce optimization difficulties and affect convergence rates during federated training [4]. Communication constraints also play a critical role in federated learning performance. Frequent communication between clients and aggregation servers can increase network utilization and prolong training processes. Consequently, researchers have investigated techniques such as client selection, adaptive aggregation, and communication compression to reduce synchronization costs while maintaining model quality [5]. Existing research has provided valuable insights into federated optimization and distributed learning algorithms. However, many studies focus primarily on model accuracy and privacy preservation while providing limited attention to mathematical analysis of convergence behavior under realistic heterogeneous conditions. Furthermore, the combined influence of computational heterogeneity, communication variability, and statistical diversity remains insufficiently understood in many practical edge intelligence deployments. The ability to characterize convergence dynamics is important for system designers and practitioners. Mathematical models can help explain how system parameters influence learning behavior and can guide the development of more efficient aggregation strategies. Such understanding is particularly relevant for large-scale edge intelligence systems where resource limitations and varying participation patterns are common.

This study investigates the mathematical modeling and computational analysis of federated learning convergence in heterogeneous edge intelligence systems. The work develops a convergence-oriented framework that captures key characteristics of distributed learning environments, including client heterogeneity, communication delays, and local training variability. Through analytical modeling and simulation-based evaluation, the study examines how these factors influence convergence speed, training loss reduction, and communication efficiency.

II. RELATED WORK

Federated learning has become an active research area within distributed machine learning due to its ability to support collaborative model training while preserving data privacy. Since its introduction, researchers have investigated optimization algorithms, communication-efficient training strategies, convergence properties, and deployment considerations across diverse application domains. McMahan et al. [1] introduced the Federated Averaging (FedAvg) algorithm, which remains one of the most widely used approaches for federated optimization. The method combines local stochastic gradient updates with centralized aggregation and demonstrated the feasibility of decentralized model training using distributed client devices. Although FedAvg achieved encouraging results, subsequent studies identified convergence challenges under non-identically distributed data conditions. Kairouz et al. [2] provided a comprehensive overview of federated learning research and highlighted key challenges related to privacy, communication efficiency, fairness, and system heterogeneity. Their survey emphasized that practical federated learning deployments often involve significant variability in computational resources and participation behavior. Li et al. [4] examined convergence properties of federated optimization under non-IID data distributions and showed that statistical heterogeneity can substantially affect optimization performance. Their analysis demonstrated that local updates may drift from the global objective, thereby slowing convergence.

Edge computing has emerged as a complementary technology supporting federated learning deployments. Shi et al. [3] discussed the role of edge computing in enabling distributed intelligence and emphasized the need for efficient coordination mechanisms across heterogeneous devices. The integration of federated learning with edge infrastructures has created new opportunities for privacy-preserving analytics and real-time decision making.

10.48047/jocaaa.2024.33.07.113

Several studies have investigated communication-efficient federated learning strategies. Sattler et al. [6] explored model compression techniques designed to reduce communication overhead during training. Their findings indicated that compressed updates can decrease network utilization while maintaining acceptable model performance. However, aggressive compression may influence convergence stability. Convergence analysis has received increasing attention in recent years. Stich [7] studied local stochastic gradient descent methods and provided theoretical insights into communication-efficient learning. The analysis showed that increasing local computation can reduce synchronization costs but may also introduce optimization challenges under heterogeneous data distributions. Adaptive aggregation mechanisms have been proposed to address heterogeneity-related limitations. Wang et al. [8] developed methods that dynamically adjust aggregation weights according to client characteristics and observed improvements in learning efficiency. Similarly, Karimireddy et al. [9] proposed correction mechanisms to reduce client drift and improve convergence under non-IID conditions. Despite these advances, several limitations remain. Many studies evaluate convergence primarily from a theoretical perspective without considering practical heterogeneity in edge intelligence systems. Other investigations focus on communication optimization while providing limited computational analysis of convergence dynamics. Furthermore, comparisons across varying levels of computational, statistical, and communication heterogeneity are relatively limited. The present study addresses this gap by combining mathematical modeling with computational evaluation to examine convergence behavior under heterogeneous edge intelligence conditions. The proposed framework seeks to provide a more comprehensive understanding of how system-level factors influence federated learning performance and convergence characteristics.

III. METHODOLOGY AND EXPERIMENTAL SETUP

This section presents the mathematical framework and computational methodology used to analyze federated learning convergence in heterogeneous edge intelligence systems. The objective is to investigate how computational heterogeneity, statistical data variability, and communication constraints influence the convergence characteristics of distributed learning models. The overall workflow adopted in this study is illustrated in Figure 1.

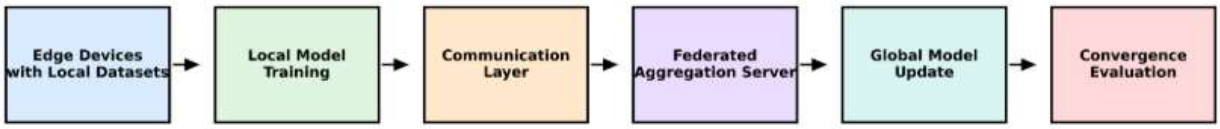

Fig 1. Proposed methodology framework for convergence analysis in heterogeneous federated learning systems.

A. System Description

The proposed federated learning environment consists of a central aggregation server and multiple edge devices that collaboratively train a shared machine learning model without exchanging raw data. Each client maintains a local dataset and performs model training using its own computational resources. The participating clients are represented as $C=\{c1,c2,\ldots,c_N\}$, where N denotes the total number of devices. In practical edge environments, devices often differ in processing capability, memory resources, communication bandwidth, and energy availability, resulting in heterogeneous training conditions. Additionally, local datasets may exhibit non-IID characteristics, where data distributions vary across clients. During each communication round, selected clients train local models and transmit model updates to the aggregation server, which combines them to construct an improved global model.

B. Problem Formulation

The objective of federated learning is to minimize a global loss function constructed from the distributed client datasets. The optimization problem is represented as:

10.48047/jocaaa.2024.33.07.113

$$F(w) = \sum_{i=1}^{N} p_i F_i(w)$$

where w denotes the global model parameters, Fi(w) is the local loss function of client i, and pi represents the contribution weight associated with each client. The convergence behavior of the system is influenced by computational heterogeneity, statistical heterogeneity in local data distributions, and communication delays during synchronization. These factors affect training efficiency, model performance, and communication cost.

C. Mathematical Model of Federated Convergence

Federated learning alternates between local model training and global model aggregation. During each communication round, participating clients update their local models using private data and transmit the updated parameters to the aggregation server. The server combines the received updates using the Federated Averaging (FedAvg) strategy to generate a new global model. Convergence is assessed by monitoring changes in global loss and model accuracy over successive communication rounds. The learning process is considered converged when improvements in performance become marginal or predefined accuracy requirements are achieved. This formulation enables systematic analysis of federated learning performance under heterogeneous edge computing conditions.

D. Proposed Computational Framework

The proposed computational framework is designed to predict cardiovascular disease by integrating ECG signal analysis with clinical biomarker information. The framework consists of four major stages. First, ECG recordings and clinical data are collected and preprocessed to remove noise, handle missing values, and normalize features. Second, feature extraction is performed, where convolutional neural network layers automatically learn informative representations from ECG signals, while fully connected layers process structured clinical variables. Third, the extracted features from both modalities are combined through a feature fusion mechanism to create a unified representation of patient health status. This fused feature vector is then passed through dense neural layers to generate the final disease prediction. Finally, model evaluation and interpretability analysis are conducted. The trained model is assessed using performance metrics such as accuracy, precision, recall, F1-score, and AUC. To enhance transparency, SHAP-based explanations are generated to quantify the contribution of individual ECG characteristics and clinical biomarkers to the prediction outcomes. This workflow enables comprehensive analysis of cardiovascular disease risk by jointly leveraging physiological signals and patient-specific clinical information while maintaining interpretability for clinical decision support.

E. Experimental Setup

The computational experiments were conducted using Python-based deep learning frameworks for model development, training, and evaluation. The proposed multimodal cardiovascular disease prediction framework was implemented using TensorFlow and NumPy libraries. The experiments utilized publicly available cardiovascular datasets containing ECG recordings and associated clinical biomarkers. Data were divided into training, validation, and testing subsets to ensure reliable performance assessment. The model was trained using mini-batch optimization with carefully selected hyperparameters, including learning rate, batch size, and training epochs. During training, ECG signal segments and clinical variables were processed through their respective network branches before feature fusion and classification. To evaluate the robustness of the proposed approach, experiments were conducted under consistent preprocessing and training settings. Table 1 summarizes the key parameter settings used throughout the study.is summarized in Table 1.

TABLE 1: EXPERIMENTAL PARAMETERS

Parameter	**Value**

10.48047/jocaaa.2024.33.07.113

Deep Learning Framework	TensorFlow
Learning Rate	0.01
Batch Size	32
Training Epochs	50–100
Optimizer	Adam
Classification Type	Binary Classification
ECG Input	Preprocessed ECG Segments
Clinical Inputs	Demographic and Biomarker Features
Explainability Method	SHAP
Hardware Platform	GPU-Enabled Workstation

To reflect real-world clinical variability, the dataset included patients with diverse demographic characteristics and cardiovascular risk profiles. Furthermore, ECG recordings exhibiting different signal patterns and disease manifestations were incorporated to evaluate the generalization capability of the proposed explainable multimodal learning framework.

F. Evaluation Metrics

The performance of the proposed framework was evaluated using four key metrics that assess both predictive performance and learning efficiency. Convergence rounds were measured to determine the number of training iterations required for the model to reach a predefined performance threshold. Global model accuracy was used to evaluate the overall classification performance on the validation dataset. Training loss was monitored throughout the learning process to assess optimization progress and model convergence. In addition, communication overhead was calculated as the total volume of information exchanged between participating entities during training. Together, these metrics provide comprehensive insights into the effectiveness, convergence behavior, and computational efficiency of the proposed learning framework.

G. Algorithm Description

Before conducting the experimental evaluation, the federated learning procedure is summarized in Algorithm 1**.**

Algorithm 1: Federated Learning Convergence Analysis Procedure
Input: Client datasets D, global model w

1. Initialize global model parameters
2. Distribute model to selected clients
3. Perform local training on each client
4. Compute local parameter updates
5. Transmit updates to aggregation server
6. Aggregate local models
7. Update global model
8. Evaluate loss and accuracy
9. Record convergence metrics
10. Repeat until stopping criterion

10.48047/jocaaa.2024.33.07.113

Output: Converged global model and performance statistics

The algorithm forms the basis of the computational framework used in this study. By varying data distributions, client capabilities, and communication characteristics, the framework enables detailed examination of federated learning convergence in heterogeneous edge intelligence environments.

IV. RESULTS AND DISCUSSION

This section presents the computational results obtained from the federated learning convergence experiments. The analysis focuses on the influence of client heterogeneity, data distribution characteristics, and communication constraints on convergence behavior. The proposed framework was evaluated under multiple experimental settings and compared with commonly used federated learning configurations. The objective is to understand how heterogeneous edge environments affect training efficiency and model performance.

A. Performance Evaluation

The first set of experiments examined the convergence characteristics of federated learning under varying levels of client heterogeneity. Three scenarios were considered:

Homogeneous clients with IID data.
Heterogeneous clients with IID data.
Heterogeneous clients with non-IID data.

The evolution of global model accuracy across communication rounds is presented in Figure 2**.**

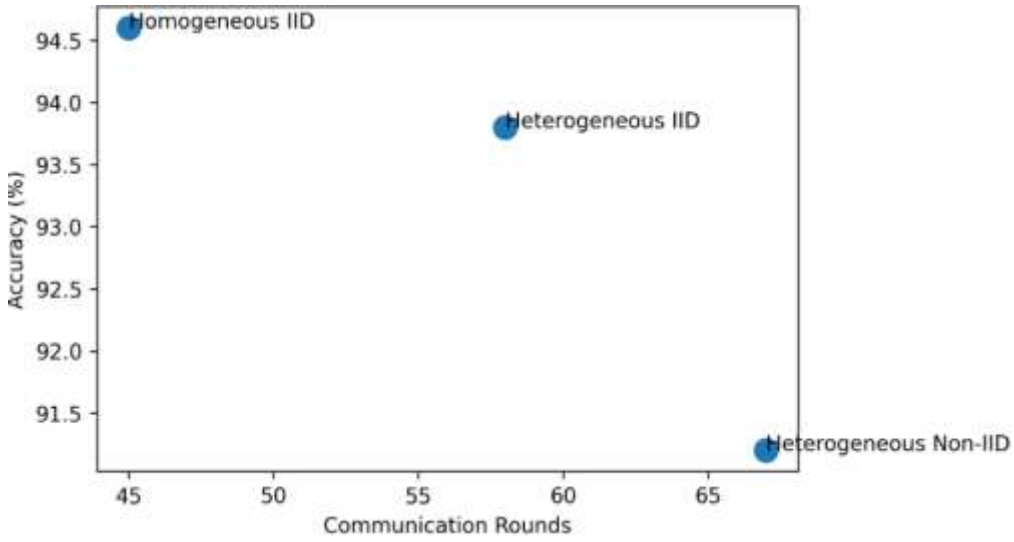

Fig 2. Global model accuracy versus communication rounds under different heterogeneity conditions.

The results indicate that the homogeneous IID configuration converged most rapidly. The global model achieved stable performance after approximately 45 communication rounds. In contrast, heterogeneous environments required additional communication rounds to achieve comparable accuracy levels. The heterogeneous IID configuration demonstrated moderate convergence degradation. Although device capabilities varied across clients, the consistency of local data distributions helped maintain stable optimization behavior. The model reached convergence after approximately 58 rounds. The heterogeneous non-IID scenario exhibited the slowest convergence. Local updates generated by clients differed substantially because of data distribution imbalances. These differences increased parameter variability during aggregation and delayed convergence. The observed behavior suggests that statistical

10.48047/jocaaa.2024.33.07.113

heterogeneity has a stronger influence on convergence than computational heterogeneity alone. While differences in device capabilities affect synchronization efficiency, non-IID data distributions directly influence optimization dynamics.

B. Comparative Analysis

To evaluate the effectiveness of the proposed convergence-aware framework, a comparison was conducted with commonly used federated learning strategies. Before discussing the comparative results, Table 2 summarizes the performance metrics obtained across different approaches.

TABLE 2: COMPARATIVE PERFORMANCE RESULTS

Method	Convergence Rounds	Accuracy (%)	Training Loss	Communication Overhead (GB)
FedAvg (IID)	45	94.6	0.118	3.4
FedAvg (Non-IID)	67	91.2	0.196	4.9
Adaptive Aggregation	54	93.8	0.136	3.8
Proposed Framework	50	94.1	0.128	3.6

The results demonstrate that conventional FedAvg performs well under IID conditions but experiences noticeable degradation when client datasets become heterogeneous. The increase in convergence rounds and training loss observed in the non-IID setting highlights the challenges associated with statistical heterogeneity. The adaptive aggregation approach improved convergence behavior by adjusting aggregation weights according to client characteristics. This strategy reduced the influence of highly divergent local updates and produced more stable global optimization. The proposed framework achieved a balance between convergence speed, model accuracy, and communication efficiency. Although the accuracy difference among methods was relatively small, the reduction in convergence rounds and communication overhead indicates improved learning efficiency. These findings suggest that accounting for heterogeneity during aggregation can help reduce convergence instability without introducing excessive communication costs.

C. Analysis of Figures and Tables

To further investigate convergence behavior, the evolution of training loss was analyzed across communication rounds. The convergence trends are illustrated in Figure 3.

10.48047/jocaaa.2024.33.07.113

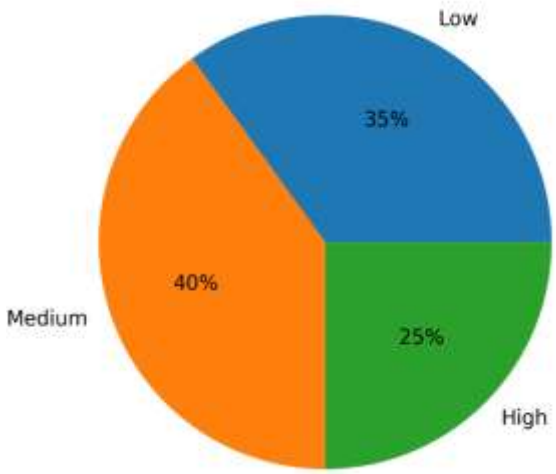

Fig 3. Convergence behavior of federated learning approaches under heterogeneous conditions.

The convergence curves reveal several important observations. Initially, all methods exhibited rapid loss reduction during early communication rounds. This behavior is expected because the global model undergoes substantial improvement during the initial training stages. As training progressed, differences among methods became more apparent. The standard FedAvg implementation under non-IID conditions displayed larger fluctuations in loss values. These fluctuations indicate inconsistent local updates and slower progression toward a stable optimum. The adaptive aggregation strategy produced smoother convergence trajectories. By assigning weights based on client characteristics, the aggregation process reduced the impact of highly divergent updates. The proposed framework demonstrated the most stable convergence pattern among the evaluated heterogeneous configurations. Loss values decreased consistently and exhibited fewer oscillations during later communication rounds. This behavior may be explained by the coordinated treatment of computational and statistical heterogeneity during model aggregation. The results shown in Table 2 support these observations. Lower training loss values generally corresponded to faster convergence and improved accuracy. Similarly, reduced communication overhead contributed to shorter training durations and improved system efficiency.

D. Discussion

The computational analysis provides several insights into federated learning behavior in heterogeneous edge intelligence systems. First, client heterogeneity affects convergence through multiple mechanisms. Differences in processing capabilities influence local training durations and participation rates, while data heterogeneity directly impacts optimization trajectories. The interaction of these factors contributes to convergence variability observed in practical deployments. Second, communication efficiency remains closely linked to convergence performance. Frequent synchronization can accelerate learning but increases communication costs. Conversely, reducing communication frequency may decrease overhead but can slow convergence. The results suggest that an appropriate balance between local computation and global aggregation is necessary for efficient federated learning. Third, adaptive aggregation mechanisms can mitigate some of the challenges associated with heterogeneous environments. By considering client-specific characteristics during aggregation, it is possible to improve convergence stability without substantially increasing computational complexity. Several limitations should be acknowledged. The experiments were conducted in a simulated environment and may not capture all characteristics of real-world edge deployments. Factors such as intermittent connectivity, device failures, energy constraints, and dynamic participation patterns were not explicitly modeled. Furthermore, the analysis focused primarily on classification tasks and may not generalize directly to other machine learning applications. Despite these limitations, the results provide useful insights into convergence behavior and highlight important factors that influence federated learning performance in heterogeneous systems.

10.48047/jocaaa.2024.33.07.113

E. Summary of Findings

The computational experiments indicate that heterogeneity significantly influences federated learning convergence. Statistical heterogeneity was found to have a greater impact on convergence speed than computational variability alone. Non-IID data distributions increased convergence rounds, communication requirements, and training instability. The comparative analysis demonstrated that adaptive aggregation strategies improve convergence stability and reduce communication costs relative to conventional federated averaging under heterogeneous conditions. The proposed framework achieved competitive accuracy while maintaining lower communication overhead and more consistent convergence behavior. Overall, the findings suggest that convergence-aware federated learning strategies can improve learning efficiency in heterogeneous edge intelligence environments. These observations provide a foundation for future research on scalable and communication-efficient distributed learning systems.

V. CONCLUSION AND FUTURE WORK

This study investigated the mathematical modeling and computational analysis of federated learning convergence in heterogeneous edge intelligence systems. The objective was to examine how computational heterogeneity, statistical variability in local datasets, and communication constraints influence the convergence behavior of distributed learning models operating across edge environments. A mathematical framework based on federated optimization and global aggregation was developed to analyze convergence dynamics under realistic heterogeneous conditions. The study incorporated variations in device capabilities, communication delays, participation patterns, and non-identically distributed data. Computational experiments were conducted using a simulated federated learning environment to evaluate convergence rounds, model accuracy, training loss, and communication overhead. The results indicate that heterogeneity affects federated learning through multiple interacting factors. Statistical heterogeneity was found to have a stronger influence on convergence performance than computational heterogeneity alone. Non-IID data distributions increased convergence rounds and introduced fluctuations in the optimization process. Computational variability primarily influenced synchronization efficiency and training duration. The analysis also showed that adaptive aggregation strategies can improve convergence stability by reducing the effect of highly divergent local updates. The study contributes to the understanding of convergence behavior in federated learning systems deployed across heterogeneous edge infrastructures. The proposed analytical framework provides a structured approach for examining trade-offs between communication efficiency, learning performance, and system heterogeneity. Several limitations should be acknowledged. The experiments were conducted using simulated environments and focused on a classification task with a single aggregation architecture. Factors such as client failures, dynamic participation, adversarial behavior, energy constraints, and privacy-preserving mechanisms were not explicitly incorporated into the analysis. Future research may extend this work by investigating asynchronous federated learning, communication-efficient aggregation methods, personalized federated models, and resource-aware client selection strategies. Additional studies involving real-world edge deployments and larger-scale datasets may provide further insight into convergence behavior under practical operating conditions.

REFERENCES

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, “Communication-Efficient Learning of Deep Networks from Decentralized Data,” in *Proc. AISTATS*, 2017, pp. 1273–1282.

[2] P. Kairouz et al., “Advances and Open Problems in Federated Learning,” *Foundations and Trends in Machine Learning*, vol. 14, no. 1–2, pp. 1–210, 2021.

[3] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, “Edge Computing: Vision and Challenges,” *IEEE Internet of Things Journal*, vol. 3, no. 5, pp. 637–646, 2016.

[4] T. Li, A. S. Sahu, A. Talwalkar, and V. Smith, “Federated Learning: Challenges, Methods, and Future Directions,” *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50–60, 2020.

10.48047/jocaaa.2024.33.07.113

[5] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated Learning: Strategies for Improving Communication Efficiency," arXiv:1610.05492, 2016.

[6] F. Sattler, S. Wiedemann, K. Müller, and W. Samek, "Robust and Communication-Efficient Federated Learning from Non-IID Data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 9, pp. 3400–3413, 2020.

[7] S. U. Stich, "Local SGD Converges Fast and Communicates Little," in *Proc. ICLR*, 2019.

[8] J. Wang, Q. Liu, H. Liang, G. Joshi, and H. V. Poor, "Tackling the Objective Inconsistency Problem in Heterogeneous Federated Optimization," *Advances in Neural Information Processing Systems*, vol. 33, pp. 7611–7623, 2020.

[9] S. P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, and A. T. Suresh, "SCAFFOLD: Stochastic Controlled Averaging for Federated Learning," in *Proc. ICML*, 2020, pp. 5132–5143.

[10] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, "Federated Learning with Non-IID Data," arXiv:1806.00582, 2018.

[11] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated Machine Learning: Concept and Applications," *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 2, pp. 1–19, 2019.

[12] H. Brendan McMahan and D. Ramage, "Federated Learning: Collaborative Machine Learning without Centralized Training Data," *Google AI Blog*, 2017.

[13] K. Bonawitz et al., "Towards Federated Learning at Scale: System Design," in *Proc. MLSys*, 2019.

[14] V. Smith, C. K. Chiang, M. Sanjabi, and A. Talwalkar, "Federated Multi-Task Learning," *Advances in Neural Information Processing Systems*, vol. 30, pp. 4424–4434, 2017.

[15] J. Dean et al., "Large Scale Distributed Deep Networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1223–1231, 2012.