Risk Assessment and Management Decisions

www.ramd.reapress.com

Risk. Assess. Manage. Decis. Vol. 1, No. 2 (2024) 277–283.

Paper Type: Original Article

Comparative Federated Algorithms for Solving Non- IID Data Challenges

Alireza Asl Nemati[1,*], Mohammad Hassan Sadreddini[1], Mohammad Mahdizade[1]

[1] Department of Computer Engineering, Imam Khomeini International University, Qazvin, Iran; alireza.aslnemati@edu.ikiu.ac.ir; mohammad.sadreddini@edu.ikiu.ac.ir; mohammad.mahdizade@edu.ikiu.ac.ir.

Citation:

Received: 05 July 2024 *Revised: 25 August 2024* *Accepted: 09 October 2024*	Asl Nemati, A., Sadreddini, M. H., & Mahdizade, M. (2024). Comparative federated algorithms for solving non-IID data challenges. *Risk assessment and management decisions, 1*(2), 277-283.

Abstract

This study evaluates three Federated Learning (FL) algorithms—FedAvg, Federated Proximal (FedProx), and MOON—by assessing their performance in Independent and Identically Distributed (IID) and non-IID settings. We found that FedAvg performs best in IID scenarios, offering quick convergence and high accuracy. However, MOON stood out as the top performer in non-IID settings, thanks to its contrastive learning method, providing better stability and accuracy across heterogeneous data. FedProx improved over FedAvg in handling non-IID data but was less effective than MOON. Our findings suggest that for environments with IID data, FedAvg is ideal, while MOON is more suitable for non-IID cases. We also highlight the need for further research into personalized FL, regularization techniques, and multimodal data integration.

Keywords: Federated learning, FedAvg, FedProx, MOON, Data heterogeneity.

1 | Introduction

1.1 | Overview of Federated Learning

Federated Learning (FL) is a decentralized machine learning paradigm that allows multiple devices or institutions to train a global model collaboratively without sharing raw data. Unlike traditional machine learning, where data is centralized on a single server, FL enables participants to train models locally and only share model updates, ensuring privacy and data security [1]. This approach has gained traction in fields like healthcare, finance, and edge computing, where privacy concerns and data regulations (e.g., General Data Protection Regulation (GDPR), Health Insurance Portability and Accountability Act (HIPAA)) limit data sharing [1]. For example, hospitals can collaboratively build diagnostic models without exposing patient records.

1.2 | Importance of FL in Distributed Machine Learning and Privacy

FL is significant because it balances privacy with collaborative learning. By keeping data localized and sharing only aggregated model updates, FL minimizes the risk of data breaches and ensures compliance with privacy regulations [3]. However, FL introduces new challenges, particularly when client data is non-Independent And Identically Distributed (non-IID). Real-world datasets often vary across users due to demographic differences, user behavior, or regional variations. This data heterogeneity leads to model bias, slower convergence, and degraded performance [2].

Corresponding Author: alireza.aslnemati@edu.ikiu.ac.ir

https://doi.org/10.48314/ramd.v1i2.56

1.3 | Motivation for Reviewing and Comparing FL Algorithms

Researchers have proposed several state-of-the-art FL algorithms to tackle the challenges posed by non-IID data. Federated Averaging (FedAvg), the baseline FL algorithm, averages client model updates. Although efficient, it faces difficulties dealing with non-IID data, leading to inconsistent local updates [3], [1]. II. Federated Proximal (FedProx), an enhancement of FedAvg, incorporates a proximal term to prevent local updates from deviating too far from the global model, which helps improve convergence in non-IID scenarios [3]. Model Contrastive Federated Learning (MOON) represents a more advanced approach that uses contrastive loss to align local and global models, ensuring stability and faster convergence, particularly in non-IID environments [2]. While these algorithms are designed to address non-IID challenges, no consensus exists on which one performs best under varying conditions [4], [5]. This uncertainty highlights the need for a comprehensive comparison of these algorithms.

1.4 | Research Objectives

The primary goal of this survey is to analyze and compare the performance of FedAvg, FedProx, and MOON under both IID and non-IID settings. Using the Modified National Institute of Standards and Technology (MNIST) dataset as a benchmark, this paper aims to:

I. Evaluate performance: compare accuracy, convergence speed, and communication efficiency across different FL algorithms.
II. Assess robustness: Examine how each algorithm handles data heterogeneity, including label distribution skew, feature imbalance, and quantity variations [3], [6].
III. Identify Trade-Offs: Highlight the strengths, weaknesses, and computational costs associated with each approach, providing insights for future FL deployments.

2 | Background and Preliminaries

Understanding FL and its challenges is crucial for evaluating how different algorithms, like FedAvg, FedProx, and MOON, address non-IID data issues. This section provides an overview of FL fundamentals, the impact of data heterogeneity, and the mathematical formulation behind distributed learning.

2.1 | Federated Learning Basics

FL is a decentralized machine learning paradigm that enables multiple clients—such as mobile devices, hospitals, or financial institutions—to collaboratively train a shared global model without sharing raw data. Unlike traditional centralized learning, where all data is transferred to a central server for training, FL keeps data local, preserving privacy and reducing the risk of data breaches [1].

2.1.1 | Centralized vs. decentralized learning

In centralized learning, a server collects data from multiple sources, consolidates it into a single dataset, and trains a model. While effective, this approach raises significant privacy and security concerns, especially when dealing with sensitive information like medical records or financial transactions [3]. In contrast, decentralized FL trains models locally on client devices. Each client trains its model using local data and sends only the model updates (gradients or parameters) to a central server, aggregating them to update the global model. This approach enhances privacy and reduces the need for large-scale data transfers, making it more efficient for distributed systems [4].

2.1.2 | Privacy concerns and communication efficiency

While FL improves privacy by keeping data local, it introduces new risks. Attackers can infer sensitive information from model updates, even if raw data is not shared. Methods like differential privacy and Secure Multi-Party Computation (SMPC) can mitigate these risks but often come at the cost of increased computational overhead [1]. Communication efficiency is another challenge. In FL, clients must frequently exchange model updates with the server. This can strain network resources, especially in environments with limited bandwidth or unstable connections [6]. Techniques like model compression, quantization, and sparse updates can help reduce communication costs while maintaining model performance [5].

2.1.3 | Federated Averaging as a Baseline

The FedAvg algorithm, introduced by [1], is the most widely used baseline in FL. It works as follows:

I. Each client trains a local model using its data for multiple epochs.
II. Clients send their updated model parameters to the server.
III. The server aggregates these updates by averaging the parameters and updates the global model.

Mathematically, the global model w at each round t is updated as

$$w^{t+1} = \sum_{i=1}^{n} \frac{n_i}{n} wi^t,$$

where wi^t represents the local model of client i, n_i is the dataset size for client i, and n is the total number of samples across all clients [3]. While FedAvg performs well in IID settings, it struggles with non-IID data because clients' local updates are biased by their unique data distributions, causing the global model to drift away from an optimal solution [3].

2.2 | Non-IID Data Challenges in Federated Learning

One of the biggest challenges in FL is dealing with non-IID data, where client datasets differ in class distribution, feature space, and quantity. This statistical heterogeneity causes model divergence, slower convergence, and reduced accuracy [3], [4]. The key non-IID challenges include.

2.2.1 | Statistical heterogeneity: Imbalanced class distributions

In real-world applications, clients often have skewed class distributions. For example, a smartphone used by a teenager may generate app usage data that differs from that of an elderly user. This label distribution skew leads to biased model updates, making it hard for the global model to generalize across all clients [3]. FedProx addresses this issue by adding a proximal term to the loss function, preventing local models from drifting too far from the global model [3].

2.2.2 | System heterogeneity: Different device capabilities

FL operates across devices with varying computational power, memory, and battery life. High-performance servers can process complex models, while low-power IoT devices may struggle with the same workload [1]. To address this, some FL frameworks, like MOON, adjust model complexity based on device capacity, ensuring weaker devices can still participate effectively [3].

2.2.3 | Communication constraints: Bandwidth and network fluctuations

Efficient communication is critical in FL, as clients frequently exchange model updates with the server. However, limited bandwidth, network congestion, and intermittent connectivity can delay updates and slow convergence [2]. Approaches like compressed updates, sparsification, and quantization reduce the size of transmitted data while maintaining model performance [4].

2.3 | Mathematical Formulation

FL aims to minimize a global loss function across distributed clients while accounting for non-IID data. Suppose there are N clients, each with a local dataset D_i, and the objective is to minimize the following global loss function:

$$\min F(W) = \sum_{i=1}^{N} \frac{|D_i|}{D} F_i(W),$$

where:

I. F(w) is the global loss function.
II. F_i(w) represents the local loss function of client i.
III. D_i is the size of the dataset for client i.
IV. D is the total number of data points across all clients.

2.3.1 | Data distribution differences in IID vs. Non-IID settings

In IID settings, each client's dataset follows the same underlying distribution:

$P_i(X, Y) = P_j(X, Y)$, for all $_{i,j}$.

However, in non-IID settings, client data distributions differ:

$$P_i(X, Y) \neq P_j(X, Y). \quad (1)$$

This distribution mismatch causes local models to diverge, making it challenging to aggregate them into a cohesive global model [3], [6].

2.4 | Summary

In summary, while FL offers significant advantages for privacy-preserving distributed learning, it faces critical challenges when client data is non-IID. The FedAvg algorithm serves as a baseline but struggles with statistical heterogeneity, system constraints, and communication bottlenecks. Advanced approaches like FedProx and MOON attempt to mitigate these challenges by regularizing updates, personalizing models, and optimizing communication [4], [5]. The next section will delve deeper into how these algorithms perform under different non-IID conditions, comparing their strengths, limitations, and trade-offs.

3 | Related Work

3.1 | Comparison of Federated Learning Algorithms

FL has gained significant attention due to its ability to train models across decentralized devices while preserving data privacy. One of the most widely used FL algorithms is FedAvg, introduced by McMahan et al. [5]. In this approach, each client trains a model locally and sends updates to the server, which are averaged to form a global model. FedAvg works well in IID settings, where data across clients is similar. However, when data is non-IID and distributed unevenly across clients, FedAvg experiences slower convergence and reduced accuracy. To overcome these challenges, FedProx was introduced by Sahu et al. [4]. FedProx includes a proximal regularization term in the optimization process to reduce discrepancies between local and global models. This adjustment allows FedProx to perform better in non-IID scenarios, achieving faster convergence and more accurate models than FedAvg. MOON, proposed by Hartmann et al. [8], takes a different approach by incorporating contrastive learning, which helps reduce the discrepancies between client models. MOON has been shown to outperform both FedAvg and FedProx in non-IID environments, providing faster convergence and better generalization.

3.2 | Benchmarks and datasets used in federated learning research

Various datasets are commonly used to evaluate the performance of FL algorithms. MNIST, Federated Extended Modified National Institute of Standards and Technology (FEMNIST), and CIFAR-10 are some of the most widely adopted benchmarks in FL research. MNIST, a classic dataset of handwritten digits, is often used in experiments to compare FedAvg, FedProx, and MOON, especially in both IID and non-IID settings [5], [6]. FEMNIST, an extension of MNIST with handwritten characters from a broader range of people, is used to test FL algorithms in more complex non-IID environments, where data distributions are skewed and imbalanced [7]. CIFAR-10, a dataset containing 60,000 color images across 10 classes, is more complicated than MNIST and is often used to evaluate the scalability and efficiency of FL algorithms, particularly in image classification tasks [10]. These datasets allow researchers to test the generalization ability of FL algorithms across various types of data and client distributions.

3.3 | Performance of FedAvg, FedProx, and MOON in Different Settings

The performance of FL algorithms varies between IID and non-IID settings. FedAvg works well in IID scenarios, where data is evenly distributed but struggles with convergence in non-IID environments. FedProx, with its proximal regularization, is more robust in non-IID settings, converging faster and achieving higher accuracy. Using contrastive learning, MOON outperforms FedAvg and FedProx in non-IID environments, offering the fastest convergence and best accuracy, making it the most robust for handling data heterogeneity.

3.4 | Discussion on Algorithm Performance in IID vs. Non-IID Settings

The comparison of FedAvg, FedProx, and MOON highlights the primary challenge of FL: Managing non-IID data. FedAvg is effective when the data across clients is similar (IID), but its performance drops significantly when it becomes more heterogeneous (non-IID). FedProx and MOON offer more robust solutions for non-IID data, with FedProx being particularly suitable for scenarios where client data distributions are imbalanced but still share some similarities. However, MOON stands out as the best performer in the most challenging non-IID environments, providing faster convergence and better generalization due to its innovative contrastive learning approach

4 | Methodology

This experiment uses the MNIST dataset of 60,000 training and 10,000 test images to evaluate FL algorithms (FedAvg, FedProx, and MOON) under both IID and non-IID conditions. The dataset is evenly distributed across clients in the IID setup, while in the non-IID setup, data is unevenly distributed among clients. The experiments are run using the FL-bench simulator, with support for both serial and parallel execution and GPU acceleration for faster training. The evaluation metrics include accuracy, loss, convergence rate, and communication efficiency. Each client trains for 20 local epochs and sends updates to the server, aggregating them and broadcasting the updated global model. The process repeats for 100 global rounds. For hyperparameters, FedAvg uses a learning rate of 0.01, an SGD optimizer, a batch size of 32, 20 local epochs, and 50 global epochs. FedProx follows similar settings but includes a proximal regularization term (μ) of 0.1. MOON uses a learning rate of 0.01, a Tau value of 0.5, Mu of 5, 20 local epochs, and 100 global epochs. The experiment involves 50 clients, each training for 20 local epochs, with global model updates over 50 global rounds. The data distribution varies between the IID and non-IID setups.

5 | Experiments and Results

In this section, we present the experimental results comparing the performance of the FedAvg, FedProx, and MOON algorithms in both IID and non-IID settings. We analyze the algorithms' convergence behavior, accuracy, and loss and provide insights into the factors influencing their performance. The experiments were conducted on the MNIST dataset.

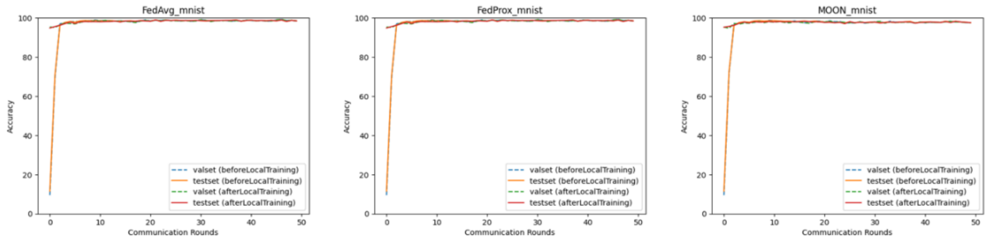
5.1 | IID Results

In the IID scenario, where data is evenly distributed across all clients, all three algorithms—FedAvg, FedProx, and MOON—show impressive performance. The training curves for all three algorithms indicate rapid convergence, with the accuracy reaching nearly 100% within the initial communication rounds.

I. Federated averaging: The FedAvg algorithm achieves a quick and smooth increase in accuracy, with both the validation and test accuracies closely following each other. The model stabilizes early, with minimal fluctuations. This result is expected, as FedAvg is well-suited for IID settings, where data is uniformly distributed across clients. The lack of significant fluctuations indicates that FedAvg performs efficiently in this scenario.

II. Federated proximal: FedProx shows similar performance to FedAvg, with a rapid rise in accuracy. However, the test accuracy curve is slightly more stable than FedAvg, suggesting that the regularization term in FedProx helps maintain model consistency even when minor variations occur in the local updates. This behavior demonstrates FedProx's ability to reduce fluctuations in accuracy slightly.

III. Model contrastive federated learning: MOON performs exceptionally well in IID settings, with a smooth and fast convergence to near 100% accuracy. The accuracy curves for both the validation and test sets show minimal fluctuation, indicating that the contrastive learning technique used in MOON effectively handles model alignment. Consistent validation and test accuracy performance reinforces MOON's robustness in IID settings.



a. b. c.

Fig. 1. Accuracy of FedAvg, FedProx, and MOON under IID; a. FedAvg under IID scenario, b. FedProx under IID scenario, c. moon under IID scenario.

5.2 | Non-IID Results

In the non-IID scenario, where data is distributed unevenly across clients, we observe more significant differences in the algorithms' performance.

I. Federated averaging: FedAvg shows slower convergence in the non-IID scenario compared to the IID setting. The accuracy fluctuates significantly, especially in the test set, reflecting the algorithm's difficulty in dealing with the non-IID data distribution. These fluctuations occur because FedAvg does not have mechanisms to address local data imbalances, causing inconsistencies between the local and global models.

II. Federated proximal: FedProx performs better than FedAvg in non-IID settings, showing smoother convergence and fewer fluctuations in accuracy. The addition of the proximal regularization term helps reduce the discrepancies between local models and the global model. FedProx achieves more stable performance, but still faces some fluctuations in accuracy, especially during the early communication rounds.

III. Model contrastive federated learning: MOON exhibits the best performance in the non-IID scenario. The contrastive learning technique significantly reduces accuracy fluctuations, resulting in more stable convergence compared to both FedAvg and FedProx. MOON achieves higher accuracy with minimal fluctuation, demonstrating that its approach effectively mitigates the challenges posed by non-IID data.

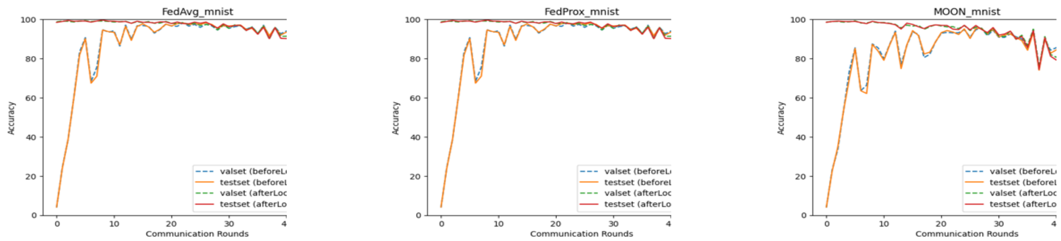


Fig. 2. Accuracy of FedAvg, FedProx, and MOON under non-IID; a. FedAvg under non-IID scenario, b, FedProx under non-IID scenario, c. moon under non-IID scenario.

5.3 | Class Distribution in IID and Non-IID Settings

This section compares the class distribution in IID and non-IID settings. In the IID setup, the data is evenly distributed across clients, ensuring a balanced representation of all classes and facilitating easier training and better algorithm performance. In contrast, the non-IID setup has an uneven data distribution, with some clients receiving only a subset of the classes, creating imbalance and challenges for the algorithms. This uneven distribution can slow convergence and affect model performance as the algorithms struggle to generalize across diverse data distributions.

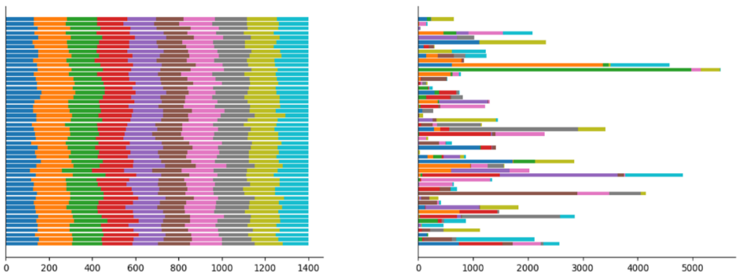


Fig. 3. Comparison of experimental data across different ranges; a. Class distribution in IID, b. class distribution in non-IID.

6 | Discussion

The performance differences between FedAvg, FedProx, and MOON in IID and non-IID settings stem from several factors. In IID scenarios, all algorithms perform well, with FedAvg showing the quickest convergence. However, in non-IID environments, FedAvg struggles due to imbalanced data, while FedProx and MOON better handle local data variations, with MOON providing the most stable performance. Additionally, non-IID setups increase communication costs, as frequent updates are needed to address discrepancies between local and global models. Both FedProx and MOON manage this more efficiently, with MOON offering the best balance of communication efficiency and performance. In terms of convergence, FedProx and MOON demonstrate more stable and faster convergence in non-IID settings, with MOON excelling in minimizing model divergence.

7 | Conclusion

This study compared the performance of three FL algorithms—FedAvg, FedProx, and MOON—across both IID and non-IID settings. In IID scenarios, where data is evenly distributed, FedAvg showed the fastest convergence and high accuracy, but MOON and FedProx were more stable, with MOON ultimately achieving the best performance. In non-IID settings, FedAvg faced slower convergence and instability, while FedProx, through regularization, improved the handling of data heterogeneity. MOON outperformed FedAvg and FedProx, offering superior convergence and accuracy thanks to its contrastive learning approach.

FedAvg is the most efficient for IID environments, but MOON is recommended for non-IID cases due to its robustness and generalization. FedProx is also a viable option for moderate data heterogeneity. Future research in FL should focus on enhancing regularization techniques, exploring personalized learning for specific client needs, and improving communication efficiency in large-scale systems. Additionally, multimodal FL, integrating diverse data types, could enhance model robustness and adaptability. In conclusion, while FedAvg is best for IID, MOON stands out for its effectiveness in non-IID scenarios, and advancements in FL will further broaden its real-world applications. In conclusion, FedAvg is well-suited for IID environments, while FedProx and MOON excel in non-IID settings. MOON provides the most effective solution for handling data heterogeneity. Further advancements in FL, especially in personalization and communication efficiency, will expand its applicability to real-world scenarios.

References

[1] Abdulrahman, S., Tout, H., Ould-Slimane, H., Mourad, A., Talhi, C., & Guizani, M. (2021). A survey on federated learning: The journey from centralized to distributed on-site learning and beyond. *IEEE internet of things journal*, *8*(7), 5476–5497. https://doi.org/10.1109/JIOT.2020.3030072

[2] Huang, C., Huang, J., & Liu, X. (2022). *Cross-silo federated learning: Challenges and opportunities*. https://doi.org/10.48550/arXiv.2206.12949

[3] Li, Q., Diao, Y., Chen, Q., & He, B. (2022). Federated learning on non-iid data silos: An experimental study. *2022 IEEE 38th international conference on data engineering (Icde)* (pp. 965–978). IEEE. https://doi.org/10.1109/ICDE53745.2022.00077

[4] Dashan Gao, Xin Yao, Q. Y. (2022). *A survey on heterogeneous federated learning*. https://doi.org/10.48550/arXiv.2210.04505

[5] Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., & Smith, V. (2020). Federated optimization in heterogeneous networks. *Proceedings of machine learning and systems* (Vol. 2, pp. 429–450). MLSys. https://B2n.ir/qp5832

[6] Che, L., Wang, J., Zhou, Y., & Ma, F. (2023). Multimodal federated learning: A survey. *Sensors*, *23*(15), 6986. https://doi.org/10.3390/s23156986

[7] McMahan, B., Moore, E., Ramage, D., Hampson, S., & Arcas, B. A. y. (2017). Communication-efficient learning of deep networks from decentralized data. *Proceedings of the 20th international conference on artificial intelligence and statistics* (Vol. 54, pp. 1273–1282). PMLR. https://proceedings.mlr.press/v54/mcmahan17a.html

[8] Hartmann, M., Danoy, G., & Bouvry, P. (2024). FedPref: Federated learning across heterogeneous multi-objective preferences. *ACM transactions on modeling and performance evaluation of computing systems*. Association for Computing Machinery. https://doi.org/10.1145/3708984

[9] McMahan, H., Moore, E., & Ramage, D. (2016). *Federated learning of deep networks using model averaging*. https://b2n.ir/md1975

[10] Sun, L., & Wu, J. (2023). A scalable and transferable federated learning system for classifying healthcare sensor data. *IEEE journal of biomedical and health informatics*, *27*(2), 866–877. https://doi.org/10.1109/JBHI.2022.3171402

[11] Zhao, Y., Li, M., Lai, L., Suda, N., Civin, D., & Chandra, V. (2018). *Federated learning with non-IID data*. http://arxiv.org/abs/1806.00582