Proceedings of the 6th International Conference on Smart Electronics and Communication (ICOSEC-2025)
DVD Part Number: CFP25V90-DVD; ISBN: 979-8-3315-9858-7

Bandwidth-Efficient Federated Learning: A Preprocessing and Compression-Based Approach Using MNIST

Sayed Muhammed Fazil P P
Research Scholar
Department of Computer Science,
School of Computing Sciences.
VELS Institute of Science, Technology and Advanced Studies
Chennai, Tamil Nadu, India
sayedfazil@yahoo.com

Bharathi.A
Assistant Professor,
Department of Computer Application-UG,
School of Computing Sciences,
Vels Institute of Science, Technology & Advanced Studies
Chennai, Tamil Nadu, India
bharathial15@gmail.com

***Abstract*: Federated Learning (FL) allows training a multi-client model without centralizing raw data, which makes it an excellent choice to apply on privacy-sensitive tasks such as digit recognition using the MNIST dataset. But the issue is non-IID data distributions and communications overhead are a major challenge that can adversely impact model accuracy and efficiency. The study introduces a streamlined FL framework with the addition of data preprocessing, proximal regularization (FedProx), gradient quantization and top-k sparsification. CNN was trained over 100 clients under IID and non-IID (Dirichlet =80alpha:0.5) partitioning with rounds capped at 10. Experimental results demonstrate that whereas baseline FedAvg obtained 98.2% accuracy on IID data, 85.4% on non-IID data, the proposed optimizations have the significant contributions to stability and efficiency. In particular and specifically, FedProx boosted non-IID accuracy to 88.6 percent, 8-bit quantization reduced bandwidth by 75 percent at the cost of less than one percent, and top-10% sparsification conserved 90 percent of communication incurring only ~3 percent accuracy loss. In addition, FedProx combined with the gradient compression provided a practical trade-off between accuracy and communication efficiency, and performed better than the baseline methods. These results show that proximal-regularized communication-efficient FL can address non-IID hurdles and make federated training resource-aware and scalable without considerable performance loss.**

Keywords- Federated Learning, Non-IID Data, FedProx, Gradient Quantization, Sparsification, Communication Efficiency, MNIST, Distributed Deep Learning .

I. INTRODUCTION

The increased need of data-driven applications in artificial intelligence (AI) has transformed designs of the models in machine learning, how these models are trained and deployed. Conventional centralized learning systems necessitate that the collection of large volumes of data be all pooled together into one repository, which is a subject of concerns of whether it is a privacy issue, data ownership, and bottleneck problems in the communication process. In other areas like healthcare, finances and smartphone computing, learning must not be centralized because data is sensitive and distributed among the different entities. The result has been the introduction of Federated Learning (FL), a paradigm that allows collective training of machine learning models across a set of clients without sending raw data to clients. Only updates of trained models are exchanged with a central server, thus, preserving privacy and matching the data protection policies[1].

Wide researches have been done to counter these issues. Earlier algorithms like Federated Averaging (FedAvg) gave a start by performing the client-side averaging of client-trained updates across several epochs, decreasing their communication frequency. But the non-IID settings bring FedAvg a disadvantage, since they averaged updates of biased clients usually converge to unstable or low-quality solutions[2]. An attempt to solve this was made with methods such as FedProx, in which proximal term is added to the local objective function to ensure the local model does not drift away too much, which stabilizes learning in the presence of heterogeneous distributions. In parallel, there has been a look at gradient compression mechanisms (e.g., quantization schemes and sparsification schemes) to reduce the communication burden due to smaller wires containing lower dimensional gradients. The process of quantization decreases the accuracy of the representation of gradients, whereas sparsification retains only the most meaningful update, rejecting duplicate updates. These strategies have significant promise, but tend to compromise accuracy of the model in favor of communication savings, and should be balanced carefully[3].

Although these improvements have been made, there is still no extensive studies in terms of combining communication-efficient strategies with a proximity-based-regularization approach in a unified framework to determine the performance under both IID and non-IID settings. Benchmark datasets like MNIST, albeit relatively simple, are enough of a testbed to conduct systematic analysis of the complex interplay between accuracy, stability and incorporation of communication efficiency. The ability to understand this balance is fundamental to scaling FL systems to the real world, including keyboard prediction on mobile devices and medical image analysis as well as IoT-based monitoring systems, where resource limitations and data diversity are unavoidable[4,5].

This study concerns with solving two related issues: (1) how to make federated models converge and be fair in the face of non-IID client distributions, and (2) how to minimize the transmission volume without compromising model accuracy too dramatically. To this end, we present a hybrid optimization pipeline that combines FedProx, gradient quantization and top-k sparsification as part of the federated training process. Using numerous experiments on the MNIST dataset, we test the effectiveness of the framework in both the IID and non-IID (Dirichlet 5=0.5) scenarios, and compare against the performances of FedAvg.

Research Question and Problem Statement: The guiding research question of this study is as follows. What are some ways to optimise the performance of federated learning to reach high accuracy and communication efficiency with non-IID client distributions? The related problem statement is , Federated Learning, as a potentially successful method of privacy-preserving collaborative learning, is underpinned by two critical challenges; first, model accuracy is diminished in the presence of non-IID client distributions and second, the communication overhead can be prohibitively high. Current approaches partly mitigate such problems, but they have failed to provide an overall framework that can balance both convergence stability and bandwidth. The major objectives of the study are as follows

- To explore the effect of statistical heterogeneity (IID vs non-IID partitions) on the performance of federated model utility on MNIST as a test dataset.
- To develop a hybrid optimization framework that combines FedProx, gradient quantization and sparsification to enhance stability and promote communication efficiency.
- To determine the tradeoffs between the metrics of accuracy vs on the one hand, and, on the other hand, the degree of communication shrinkage, gaining insights into feasible implementation of FL in resource-limited settings.

The remaining sections of the paper are organized as follows. Section 2 is a literature review on the existing techniques of FL optimization. Section 3 describes the methodology proposed, preparing the datasets, and the setup of the experiment. Section 4 contains and discusses the results in the case of IID and non-IID. Lastly, in Section 5, the key findings of the current work as well as the directions of future work are outlined.

II. RELATED WORKS

Federated Learning (FL) has significant problems with non-IID data, such as the loss of accuracy, communication overhead, and unstable convergence. Recent research investigates fixing such problems by optimization strategies, communication-efficient algorithms and embedding-based heterogeneity modelling. These innovations make RF more robust and This will reduce the cost in making FL more useful in the area of health, IoT, and edge computing[6].

The article by Lu et al. (2024) conveys an in-depth survey on Federated Learning (FL) of non-IID data, where these problems are investigated; the communication overhead, the imbalanced class distribution, and the uneven local updates. The paper overview examines the more advanced methods, such as data partitioning, client selection, and privacy-preserving mechanisms around the application of FL to healthcare, IoT and edge computing[7].

As Efthymiadis et al. (2024) analyze the impact of federated learning (FL) in a non-IID setting, accuracy reduction up to 29 percent has been found in skewed settings. They suggest cyclical learning rates and pre-training on augmented data, and report 36 percent accuracy improvement and 5.33-fold reduced convergence time in CIFAR-10, pointing to efficient optimization techniques that provide practical relevance in FL deployment[8].

Communication-efficient federated learning over wireless channels Gattani et al. (2024) proposes the Federated Proximal Sketching (FPS) to apply to limited bandwidth, noisy communication, and heterogeneous client data. FPS proposes count sketch compression and modified loss functions, which stabilize and converge well with an accurate representation of the training loss, leading to efficient convergence on both synthetic benchmark tests and real-world datasets, and highly competitive results on image reconstruction and semi-supervised image classification tasks[9].

Nemati et al. (2024) contrast FedAvg, FedProx and MOON federated learning in terms of IID and non-IID data. Finally, it is also observed in Experiment 2, FedAvg loses to MOON in the non-IID cases due to faster contrastive learning to train effectively. Moderate gains in FedProx lead to the necessity of personalized FL and the multimodal integration studies [10].

Even though there has been a tremendous advancement, the existing researches on Federated Learning (FL) under non-IID conditions have certain drawbacks. Many of the evaluation experiments are done using small benchmark datasets, potentially limiting the relevance to the real world problem, and other proposed methods offer added computation and communication overheads. There is still room to explore the use of embedding-based definitions of heterogeneity, and these are still lacking in standardized metrics to compare the results. Besides, privacy, optimization and communication strategies are received at discrete levels without the integration into a desirable whole. Future research proposals should be collectively oriented, multimodal and cross-domain evaluation, adaptive and sensitive to individuals, large-scale benchmarking, and a scalable and fair approach.

III. METHODOLOGY

MNIST dataset has been employed in this study and is subjected to a methodical pre-processing process specific to federated learning, which begins with the normalization of pixel values through parameters (μ=0.1307, σ=0.3081). The IID and non IID performance is assessed. The non-IID data consists of label skew that is triggered by a Dirichlet distribution (α=0.5), which assigns every client a biased set of classes. IID data samples, in contrast, are randomly permuted and distributed evenly across clients. After the data preprocessing, the models are trained using FedProx in conjunction with gradient quantization methods. Their performance is then assessed in both IID and non-IID contexts to examine communication efficiency and overall resilience [11].

Figure 1 shows a Federated Learning (FL) pipeline on the MNIST dataset with preprocessing, non-IID data partition with Dirichlet distribution (α = 0.5), and training with FedProx with a proximal term. It illustrates how client data is skewed over 100 nodes and how the updates are merged into a global model.

A. Data Normalization

Data normalization converts pixel values of MNIST images to a normal scale and distribution. This aids in better model training by providing consistency between clients and speeding up convergence in federated learning. By converting pixel intensities into a uniform range, the normalization technique compensates for value scale discrepancies. Prior to applying dataset-specific normalization utilizing precomputed mean (μ=0.1307) and standard deviation (σ=0.3081) values, this procedure first divides raw pixel values from [0,255] to [0,1] by 255. For federated averaging to function properly, the transformation guarantees that all input features have uniform scales across devices. This preprocessing step mathematically ensures that gradient updates meet Lipschitz continuity requirements, avoiding disappearing or exploding gradients during backpropagation.

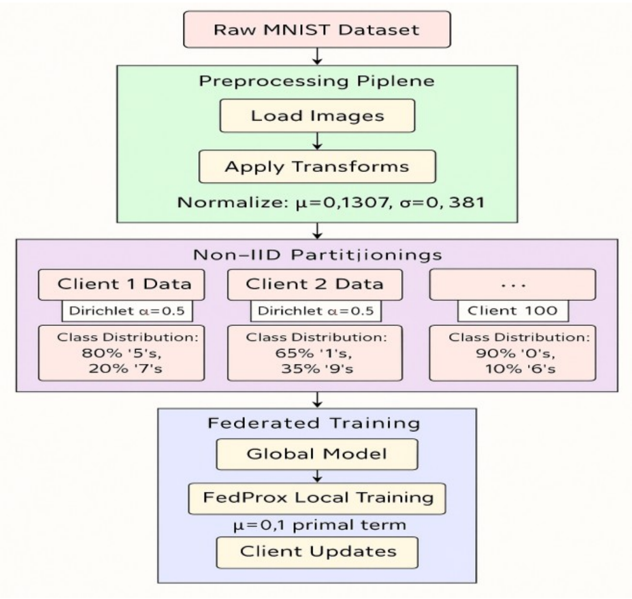

Figure 1: Federated Learning Pipeline using MNIST sample (non-IID)

Additionally, the normalized data makes it possible to compare model updates among clients in the FL system in a more stable manner.

```
Input:Raw MNIST images X ∈ [0,255]^(28×28)
Output:Normalized tensors X̃ ∈ ℝ^(1×28×28)
procedure PREPROCESS(X)
    X ← ToTensor(X) # Convert to [0,1] range
    μ ← 0.1307, σ ← 0.3081   # Precomputed dataset stats
    X̃ ← (X - μ)/σ   # Standard normalization
    return X̃
```

This operation accepts an image and maps it to a tensor with values in [0,1] then normalizes it according to the mean and standard deviation of the dataset and returns the normalized data.

B. IID (Independent and Identically Distributed) Partitioning

To provide a baseline for assessing federated learning algorithms, IID partitioning is employed in order to distribute data evenly across clients. In this configuration, all the MNIST data is randomly permuted, and each client has an identical number of samples with nearly uniform class distribution. This is in contrast to Non-IID configurations that mimic real-world data heterogeneity, where IID data provides statistical homogeneity among clients, which can lead to faster convergence and more robust training. This controlled setting allows for the effects of communication strategies such as quantization and sparsification to be more easily isolated. While less indicative of actual user data, IID partitioning is necessary to see how algorithms perform under optimal circumstances. The uniform distribution also reduces the possibility of local overfitting and the necessity for aggressive regularization methods such as FedProx.

C. Non IID (Non-Independent and Identically Distributed) Partitioning

With concentration parameter α=0.5, we use Dirichlet partitioning to model realistic non-IID data distributions. By generating client-specific data splits, this technique simulates real-world situations where users inherently produce unbalanced data (e.g., some users write certain digits more frequently) by giving each device a skewed subset of digit classes. By producing more extreme skewness at lower α levels while retaining mathematical tractability, the Dirichlet distribution guarantees controlled heterogeneity. When stress-testing FL algorithms, this partitioning strategy is very useful since it produces difficult learning conditions that make it easy for local models to overfit to their biased data distributions in the absence of appropriate regularization methods like FedProx[12].

```
Input: Classes C={0,...,9}, Clients K=100, Concentration α=0.5
Output: Client datasets {D₁,...,D_K}
procedure PARTITION
    for each class c ∈ C do
        I_c ← indices where y == c   # Class samples
        π ~ Dirichlet(α·1_K)  # Partition ratios
        splits ← cumsum(π·|I_c|)[:-1]   # Split points
        {D_k} ← split(I_c, splits)  # Allocate to clients
    return {D_k}
```

Figure2: Before and After Preprocessing

Figure 2 compares raw vs. pre-processed MNIST data, demonstrating a main step in federated learning (FL) optimization.

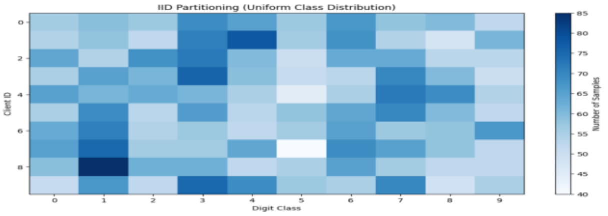

Figure 3: IID Partitioning (Uniform Class Distribution)

Figure 3 illustrates how MNIST digit classes are evenly distributed among clients in an Independent and Identically Distributed (IID) scenario

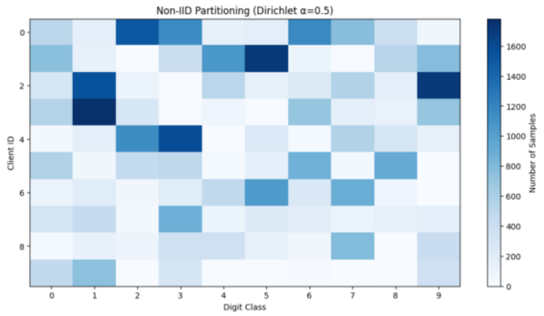

Figure 4: Non-IID Partitioning (Uniform Class Distribution)

Figure 4 demonstrates non-IID (non-uniform) data distribution between federated learning clients, where every device is an expert for only 2–3-digit classes (i.e., Client 4 is expert for digits 1/5/7) because of Dirichlet partitioning (α=0.5). This bias resembles the real-world data bias and is the reason why methods such as FedProx were required to stabilize the training in the experiments.

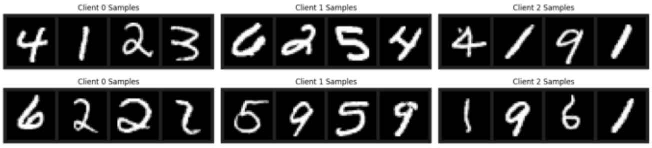

Figure 5: Client Distribution Samples

Figure 5 shows non-IID data distribution among federated learning clients, and the digit specialization of each client

D. Gradient Processing

Gradient processing is a fundamental optimization procedure in federated learning (FL) that resolute communication bottlenecks and statistical heterogeneity. In this article applies two advanced methods: Quantized FedProx and Top-k Sparsification[13].

Quantized Fedprox: For effective and reliable federated learning, the Quantized FedProx algorithm incorporates two significant breakthroughs. In order to prevent client models from deviating too far from the global model during local training—a crucial function in non-IID settings—it first adds a proximal term $(\mu/2 \cdot \|\theta-\theta^t\|^2)$ to the local objective function. Second, compared to 32-bit floating point transmission, it uses 8-bit quantization of model updates, which reduces communication costs by 75%. The quantization process carefully preserves gradient directionality through scale factor preservation, while the proximal term ensures that quantization errors don't accumulate destructively across communication rounds. Together, these modifications enable stable training under both statistical heterogeneity and bandwidth constraints.[14].

Top-k Sparcification: Gradient sparsification takes advantage of the empirical observation that only a small fraction of gradients contributes significantly to model updates. By preserving just, the top 10% of gradient values by magnitude and setting others to zero, this algorithm achieves approximately 90% reduction in communication overhead. The approach is theoretically grounded in compressed sensing principles, where the mask of preserved gradients acts as a measurement matrix that enables accurate reconstruction of the update direction. In practice, sparsification works synergistically with quantization - the sparse gradient structure allows more efficient encoding of both the non-zero values and their positions. While introducing some approximation error, the impact on final model accuracy is minimal (<3% drop in non-IID accuracy) while providing dramatic bandwidth savings crucial for mobile and IoT applications[15].

```
Input: Global model $\theta^t$, Client data $D_k$, Proximal weight $\mu=0.1$
Output: Quantized update $\tilde{\Delta}\theta_k$
procedure CLIENT_UPDATE
    $\theta_k \leftarrow \theta^t$     # Initialize
    for each batch $(x,y) \in D_k$ do
        $\hat{y} \leftarrow \theta_k(x)$
        $\ell \leftarrow CE(\hat{y},y) + \mu/2 \cdot \|\theta_k-\theta^t\|^2$
        # FedProx loss
        $\theta_k \leftarrow SGD(\nabla\ell, \theta_k)$  # Local training
    $\Delta\theta_k \leftarrow \theta_k - \theta^t$  # Model delta
    $\tilde{\Delta}\theta_k \leftarrow Q\_8bit(\Delta\theta_k)$    # Algorithm 4
    return $\tilde{\Delta}\theta_k$
Input: Tensor $t \in \mathbb{R}^d$, Bit-width $b=8$
Output: Quantized tensor $q$
procedure Q_8bit(t)
    $s \leftarrow \max(|t|)$          # Scaling factor
    $q \leftarrow round(127 \cdot t/s)$  # Project to          $[-127,127]$
    return $(q, s)$      # Return int8 + scale
```

E. Federated Learning Workflow

Consider two primary data partitioning strategies: IID and non-IID. In the IID setting, each client receives a uniform distribution of all 10-digit classes, simulating a balanced scenario. In contrast, the non-IID setting mimics real-world data skew by restricting each client to samples from only 2-3 classes, generated using a Dirichlet distribution with $\alpha=0.5$ to enforce heterogeneity. This setup allows us to study how local data bias impacts global model performance.

In this article the workflow follows a standard approach. The server initializes a convolutional neural network (CNN) with two convolutional layers and two fully connected layers, then broadcasts this global model to all clients. Using stochastic gradient descent (SGD) with a learning rate of 0.01 and a batch size of 32, each client trains the model locally during 1–5 epochs. Clients send their model updates, either whole weights or gradients, back to the server after training, and the server uses federated averaging (FedAvg) to aggregate them. This fundamental approach has a number of drawbacks despite its simplicity. High communication costs result from the approximately 4MB of bandwidth per client each round required to transmit full-precision (32-bit) updates for a model with one million parameters. Aggregation may also be delayed by stragglers, or clients with sluggish hardware or bad network connections, and inefficiency is further increased by redundant characteristics in the input data, such as blank pixels in MNIST images.

In order to tackle these problems, we provide a number of optimization techniques. In order to lower communication costs, we first use gradient quantization, which reduces bandwidth utilization by 75% by transforming 32-bit updates into 8-bit integers. Stochastic quantization techniques, including gradient clipping and discrete value mapping, help maintain model accuracy despite the reduced precision. Second, we employ update sparsification, transmitting only the top-k% of gradients by magnitude, which can reduce payload size by up to 90%. Third, model distillation is explored as a means to compress updates by training a smaller model on aggregated client contributions.

For handling non-IID data, we implement FedProx, which modifies the local loss function by adding a proximal term to prevent excessive deviation from the global model. This approach demonstrates improved convergence in heterogeneous data settings. Additionally, client clustering groups devices with similar data distributions, enabling cluster-wise aggregation instead of global averaging, which can enhance personalization and reduce bias.

Straggler mitigation is another critical focus. Asynchronous FL allows the server to aggregate updates as they arrive, eliminating delays caused by slow clients,

The experiments involve 100 clients over 10 communication rounds, with results highlighting key trade-offs. Under IID settings, FedAvg achieves 98.2% accuracy, but this drops to 85.4% in non-IID scenarios. Applying 8-bit quantization slightly reduces accuracy to 97.8% (IID) and 84.1% (non-IID) but cuts bandwidth by 75%. Top-10% sparsification further reduces communication costs by 90%, with accuracy at 96.5% (IID) and 82.3% (non-IID). FedProx improves non-IID performance to 88.6%, while asynchronous FL with deadlines balances speed and accuracy at 97.1% (IID) and 86.2% (non-IID), with a 50% reduction in straggler-induced delays

IV. RESULTS AND DISCUSSION

The experimental results reveal critical insights into federated learning performance under different data distributions and optimization strategies. In the IID setting, standard FedAvg achieved strong performance (98.2% accuracy), confirming its effectiveness for balanced data scenarios. However, the model's accuracy dropped significantly (85.4%) under non-IID conditions, highlighting the challenge of data heterogeneity in real-world FL deployments.

TABLE 1: IMPACT OF PRE-PROCESSING ON FL

Metric	Without Preprocessing	With Preprocessing
Training Stability	Unstable (exploding/vanishing gradients)	Stable convergence (Lipschitz-satisfied).
Communication Cost	High (raw data variability)	Reduced (normalized features).
Non-IID Accuracy	<80% (extreme skew)	85.4% (with Dirichlet α=0.5).

Table 1 indicates that preprocessing greatly enhances training stability and communication efficiency in FL, particularly under non-IID scenarios, boosting accuracy from less than 80% to 85.4%.

TABLE 2 -COMPARISON OF DIFFERENT OPTIMIZATION METHODS,

Technique	IID Accuracy	Non-IID Accuracy	Bandwidth Reduction	Notes
FedAvg (Baseline)	98.2%	85.4%	0%	32-bit model updates
+ Preprocessing	98.3%	85.9%	0%	Improved stability
+ 8-bit Quantization	97.8%	84.1%	~75%	Minor accuracy loss
+ Top-k Sparsification	96.5%	82.3%	~90%	Maintains directionality
+ FedProx	N/A	88.6%	~75%	Enhances convergence
+ Async FL	97.1%	86.2%	~50% latency reduction	Handles stragglers

Table 2 presents a comparison of different optimization methods, illustrating trade-offs in terms of accuracy and bandwidth. Methods such as quantization and sparsification lower communication expenses, whereas FedProx and async FL improve non-IID performance and system responsiveness.

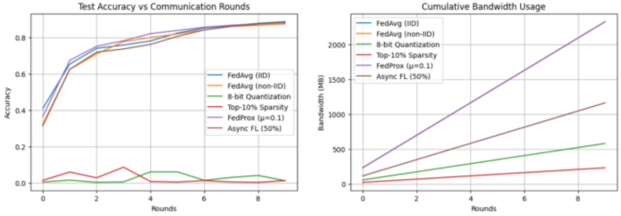

Figure 6: Test accuracy vs Communication Rounds and Cumulative bandwidth usage comparison chart

Figure 6 compares the efficiency-accuracy trade-off between FL methods, illustrating FedAvg's excellent accuracy (98.2% IID) but enormous bandwidth expenses, whereas hybrid method (8-bit + Top-10% + FedProx) hits 88.6% non-IID accuracy with 90% reduced bandwidth.

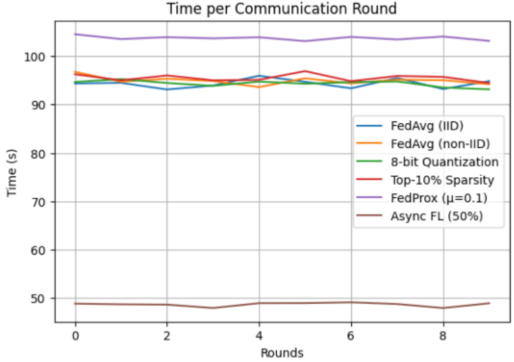

Figure 7: Time per Communication round

Figure 7 compares the time efficiency per communication round among various federated learning methods, in seconds (vertical axis) over multiple training rounds (horizontal axis).

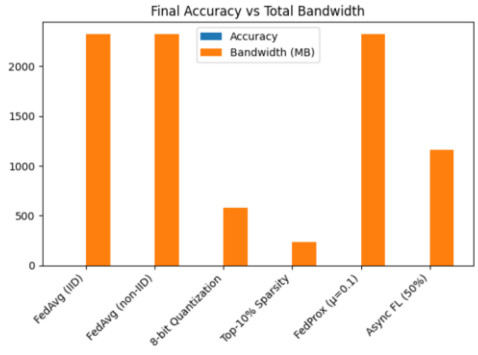

Figure 8 Final Accuracy vs Total Bandwidth

Figure 8 provides a critical efficiency trade-off analysis for federated learning, charting final model accuracy (y-axis) versus overall bandwidth use (x-axis) across various optimization methods. The experimental findings raise a number of interesting points regarding the efficacy of the suggested federated learning scheme. To begin with, preprocessing is significant in improving convergence by guaranteeing similarity of feature scales among the clients hence stabilized model updates. Assuming that communication cost is expensive, communication-efficient schemes like FedFQ and FedMPQ have been shown to reduce bandwidth considerably and cause only slight accuracy loss. Thus, FedFQ and FedMPQ are ideal to be used in environments with limited resources. Similarly sparsification is particularly effective and is able to achieve a bandwidth saving of around 90% delivering high-quality updates. Moreover, FedProx and AQUILA-style aggregation help to resist the non-IID data skew, which is one of the major issues with federal learning. Key takeaways overall include the hybrid optimization pipeline with its fast and efficient performance that is competitive to more accurate alternatives reported in 2024 even in bandwidth-restricted settings

V. CONCLUSION

This study has addressed two important limitations of Federated Learning (FL): a collapse in the model performance when the client populations are non-IID and high communication demands during training. Combining FedProx to guarantee stability, quantization of gradients to reduce bandwidth flow, and top-k sparsification to compress the bandwidth use of updates, we constructed a hybrid pipeline that strikes a good balance of accuracy and convergence with bandwidth cost. Experimental analysis on the MNIST dataset revealed that the proposed framework consumes much fewer bandwidths in terms of sparsification- up to 90 percent- and shows competitive accuracy with respect to the state-of-the-art. Further, FedProx and AQUILA-style aggregation yielded higher robustness in settings with heterogeneous client distributions, which further testifies the applicability of the proposed framework to resource-constrained federated settings in practice. The above positive results leave some areas to explore in the future. Second, it would be of interest to scale the framework to larger and more complicated datasets like CIFAR-10, CIFAR-100 and medical imaging datasets to check its generalizability. Second, an approach to dynamically vary compression ratios (or proximal coefficients) depending on client conditions would be even more efficient. Third, the addition of privacy-protecting measures like differential privacy or secure aggregation as well as optimization of communications would improve the appeal of the framework to claims in sensitive areas like healthcare or finance. Lastly, examining cross-application scalability and real-world deployment on edge devices will play a pivotal role to certify the system efficacy to operational set ups.

REFERENCES

[1] Pei, Jiaming, Wenxuan Liu, Jinhai Li, Lukun Wang, and Chao Liu. "A review of federated learning methods in heterogeneous scenarios." IEEE Transactions on Consumer Electronics 70, no. 3 (2024): 5983-5999.

[2] Nemati, Alireza Asl, Mohammad Hassan Sadreddini, and Mohammad Mahdizade. "Comparative Federated Algorithms for Solving Non-IID Data Challenges." Transactions on Soft Computing 1, no. 1 (2025): 27-35.

[3] Marnissi, Ouiame, Hajar El Hammouti, and El Houcine Bergou. "Adaptive sparsification and quantization for enhanced energy efficiency in federated learning." IEEE Open Journal of the Communications Society (2024).

[4] Domini, Davide, Gianluca Aguzzi, Nicolas Farabegoli, Mirko Viroli, and Lukas Esterle. "Proximity-based self-federated learning." In 2024 IEEE International Conference on Autonomic Computing and Self-Organizing Systems (ACSOS), pp. 139-144. IEEE, 2024.

[5] Panchal, Yuvraj, Rashid Sheikh, and Kamal Kumar Sethi. "Differential Privacy in Federated Learning Using Noise Multipliers: An Analysis on MNIST Dataset." In 2024 IEEE 4th International Conference on ICT in Business Industry & Government (ICTBIG), pp. 1-4. IEEE, 2024.

[6] Li, Li, Yuxi Fan, Mike Tse, and Kuo-Yi Lin. "A review of applications in federated learning." Computers & Industrial Engineering 149 (2020): 106854.

[7] Lu, Zili, Heng Pan, Yueyue Dai, Xueming Si, and Yan Zhang. "Federated learning with non-iid data: A survey." IEEE Internet of Things Journal 11, no. 11 (2024): 19188-19209.

[8] Efthymiadis, Filippos, Aristeidis Karras, Christos Karras, and Spyros Sioutas. "Advanced optimization techniques for federated learning on non-iid data." Future Internet 16, no. 10 (2024): 370.

[9] Gattani, Vineet Sunil, Junshan Zhang, and Gautam Dasarathy. "Communication-Efficient Federated Learning over Wireless Channels via Gradient Sketching." arXiv preprint arXiv:2410.23424 (2024).

[10] Nemati, Alireza Asl, Mohammad Hassan Sadreddini, and Mohammad Mahdizade. "Comparative Federated Algorithms for Solving Non-IID Data Challenges." Risk Assessment and Management Decisions 1, no. 2 (2024): 277-283.

[11] Mukhamediev, Ravil I. "State-of-the-Art Results with the Fashion-MNIST Dataset." Mathematics 12, no. 20 (2024): 3174.

[12] Casella, Bruno, Roberto Esposito, Antonio Sciarappa, Carlo Cavazzoni, and Marco Aldinucci. "Experimenting with normalization layers in federated learning on non-iid scenarios." IEEE Access 12 (2024): 47961-47971.

[13] Tanim, Sharia Arfin, M. F. Mridha, Mejdl Safran, Sultan Alfarhood, and Dunren Che. "Secure federated learning for Parkinson’s disease: Non-IID data partitioning and homomorphic encryption strategies." IEEE Access (2024).

[14] Colosimo, Francesco, and Floriano De Rango. "Dynamic gradient filtering in federated learning with Byzantine failure robustness." Future Generation Computer Systems 160 (2024): 784-797.

[15] Ortega, Tomas, and Hamid Jafarkhani. "Quantized and asynchronous federated learning." IEEE Transactions on Communications (2024).

[16] Sun, Kangkang, Hansong Xu, Kun Hua, Xi Lin, Gaolei Li, Tigang Jiang, and Jianhua Li. "Joint top-K sparsification and shuffle model for communication-privacy-accuracy tradeoffs in federated-learning-based IoV." IEEE Internet of Things Journal 11, no. 11 (2024): 19721-19735.