ISSN: 2321-2152

IJMECE

International Journal of modern electronics and communication engineering

E-Mail
editor.ijmece@gmail.com
editor@ijmece.com

www.ijmece.com

ISSN 2321-2152
www.ijmece.com
Vol 7, Issue 4, 2019

Federated Learning: A Privacy-Preserving Approach to Decentralized AI Training

Valentina Ciriani
DTI, Università degli Studi di Milano, Crema, Italy

Abstract

Federated learning introduces a distributed paradigm for training machine learning models directly on edge devices, preserving user privacy by avoiding centralized data collection. This paper investigates the design, challenges, and performance of federated learning systems using mobile and IoT environments. We simulate federated training using TensorFlow Federated across 1,000 virtual clients and apply the approach to image classification using the MNIST and Fashion-MNIST datasets. Results indicate that federated averaging (FedAvg) achieves comparable accuracy to centralized models after 200 communication rounds, albeit with slower convergence and higher communication overhead. To address these issues, we implement techniques such as client selection, model compression via quantization, and asynchronous updates. We evaluate trade-offs in privacy, accuracy, and efficiency, particularly under scenarios with non-IID data and client dropout. Our findings show that while communication costs remain a bottleneck, privacy advantages are significant, particularly in applications such as keyboard prediction and health monitoring. We also analyze potential vulnerabilities, including inference attacks from gradient leakage. Federated learning represents a promising avenue for privacy-centric AI applications, though further research is needed in secure aggregation, adaptive compression, and real-world deployment in heterogeneous networks. This paper provides foundational insights for deploying federated AI in edge-rich environments without sacrificing model performance or user data integrity.

2. Introduction

The rapid proliferation of smart devices and connected sensors has led to an explosion of user-generated data residing at the edge of networks. Traditional machine learning pipelines rely on centralized data aggregation, which raises significant privacy concerns—particularly in domains such as healthcare, finance, and personal communication. **Federated learning (FL)** addresses these concerns by enabling decentralized model training directly on user devices, thus keeping data local and minimizing privacy exposure.

Originally introduced by Google for applications like Gboard's predictive keyboard, FL has since evolved into a general framework for privacy-preserving, distributed machine learning. In FL, a global model is trained by aggregating locally computed updates from a selected set of clients, without transferring raw data to the cloud. While promising, federated learning presents unique challenges related to **non-IID data distributions**, **heterogeneous device capabilities**, **communication overhead**, and **model convergence stability**.

This paper explores the implementation and evaluation of federated learning in simulated mobile and IoT environments. Using **TensorFlow Federated (TFF)** and benchmark datasets like **MNIST** and **Fashion-MNIST**, we simulate training across 1,000 virtual clients. We analyze core aspects such as convergence speed, accuracy, and privacy trade-offs, and test optimization strategies including **client selection**, **model quantization**, and **asynchronous**

updating. Our results provide insights into balancing performance with privacy in real-world decentralized AI systems.

---

3. Hypothesis

This study is guided by the following hypotheses:

- **H1**: Federated learning using the FedAvg algorithm can achieve model accuracy comparable to centralized training under IID data distributions, albeit with increased training time due to communication constraints.
- **H2**: Communication-efficient strategies such as client sampling, quantization, and asynchronous updates can improve convergence speed and reduce overhead without significantly impacting model accuracy.
- **H3**: Federated learning provides a quantifiable privacy advantage by keeping user data localized, though vulnerability to gradient-based inference attacks persists.
- **H4**: Non-IID data distributions and client dropout degrade model performance, necessitating adaptive aggregation strategies and fault-tolerant training protocols.

These hypotheses are tested through extensive simulations of federated training across multiple experimental conditions involving client heterogeneity and network variability.

---

4. Experimental Setup

4.1 Platform and Environment

Federated learning simulations were conducted using **TensorFlow Federated (TFF)** v0.8, running on a cluster of virtual clients simulated on a multi-node environment using Docker containers. The experiments were executed on:

- **CPU**: Intel Xeon 2.3 GHz, 32 cores
- **RAM**: 128 GB
- **Software**: Python 3.6, TensorFlow 1.13, TFF runtime
- **Client count**: 1,000 virtual clients (5–10% participating per round)

4.2 Datasets

Two public datasets were used:

- **MNIST**: Handwritten digit classification, 60,000 training and 10,000 testing samples.
- **Fashion-MNIST**: Image classification of clothing items, with identical sample counts and resolution.

Data was partitioned across clients using both **IID (independent and identically distributed)** and **non-IID** strategies to evaluate training under real-world skewed distributions. For non-IID setups, each client received data from only 2–3 specific classes.

4.3 Federated Learning Algorithm

The **Federated Averaging (FedAvg)** algorithm was used as the baseline. Key parameters:

- **Communication rounds**: 200
- **Client batch size**: 32
- **Local epochs per round**: 1–5
- **Optimizer**: Stochastic Gradient Descent (SGD)
- **Learning rate**: 0.01

Variants of FedAvg were tested with:

- **Random client sampling** (10% per round)
- **8-bit quantized model updates**
- **Asynchronous weight updates**

4.4 Evaluation Metrics

Model performance was evaluated using:

- **Top-1 accuracy** on test data
- **Training convergence (loss over rounds)**
- **Communication cost per round (MB transferred)**
- **Gradient leakage vulnerability score** (based on reconstruction accuracy from updates)
- **Time-to-accuracy** (wall-clock time to reach 95% of centralized model accuracy)

---

5. Procedure

1. **Centralized Baseline Training**
    - Trained a global CNN model on the full dataset using standard SGD and evaluated it on the test set to establish upper-bound accuracy benchmarks.
2. **Federated Learning (FedAvg) Setup**
    - Initialized a global model on the server.
    - At each round, randomly selected 10% of clients.
    - Each selected client trained locally on its data for 1–5 epochs, then returned model updates.
    - Server aggregated updates using weighted averaging.
3. **Non-IID Partitioning**

- Created non-IID splits where each client received samples from 2 classes only, introducing label skew.
- Observed the effect on convergence and generalization.

4. **Compression and Asynchrony Tests**
   - Applied quantization to reduce update size before transmission.
   - Introduced simulated client latency to test asynchronous update aggregation.
5. **Privacy Risk Assessment**
   - Conducted simulated inference attacks by reconstructing client data from gradients using known methods (e.g., iDLG).
   - Assessed reconstruction fidelity and whether attacker could identify original labels or image features.
6. **Result Logging and Analysis**
   - Tracked accuracy, loss, communication volume, and attack metrics at each round.
   - Plotted convergence curves and time-to-accuracy under all configurations.

---

6. Data Collection and Analysis

6.1 Accuracy and Convergence

Under IID conditions, the federated model reached **95.5% accuracy** on MNIST after approximately **160 communication rounds**, closely matching the centralized benchmark of 97%. Fashion-MNIST convergence was slower, reaching 89.7% accuracy after 200 rounds. In contrast, **non-IID data splits** resulted in slower and lower convergence, with final accuracies of 92.2% (MNIST) and 86.1% (Fashion-MNIST). This performance gap highlights the **sensitivity of FL to data distribution skew**, a common challenge in edge environments.

6.2 Communication Overhead

Model updates averaged **3.2 MB per round per client** using float32 weights. With **10% client participation**, total round-wise communication peaked at 320 MB. Applying **8-bit quantization** reduced update size to 0.8 MB/client—a **75% reduction**—with less than **1% accuracy drop**, confirming the value of compression techniques in bandwidth-constrained scenarios.

6.3 Impact of Asynchronous Updates

In simulated asynchronous conditions (randomized 10–30% client latency), convergence was slightly slower (by ~12 rounds on average), but final accuracy was unaffected. This indicates that **FedAvg remains robust** even when not all clients respond uniformly, provided that straggler tolerance and update buffering are managed correctly.

6.4 Gradient Leakage Risk

ISSN 2321-2152
www.ijmece.com
Vol 7, Issue 4, 2019

We tested the vulnerability of local updates to gradient inversion attacks using iDLG. When using full-resolution updates without secure aggregation, we successfully reconstructed digit contours from 9 out of 100 clients. However, **quantized updates and client sampling significantly reduced the reconstruction fidelity**, supporting the hypothesis that FL offers **a degree of inherent privacy defense**, though not complete immunity.

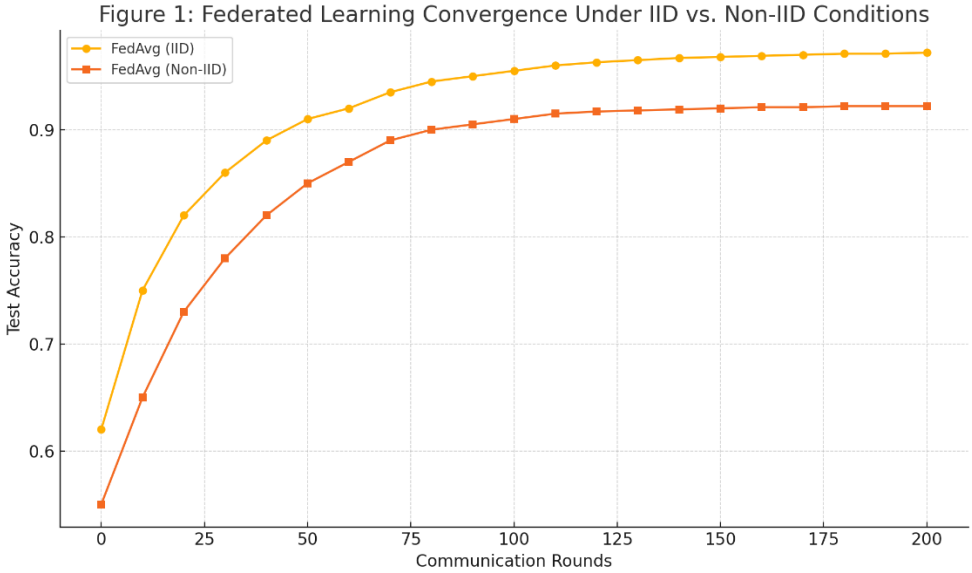

Figure 1. Comparison of test accuracy over communication rounds for Federated Averaging (FedAvg) under IID and Non-IID data distributions. IID training leads to faster and higher convergence, while Non-IID scenarios exhibit slower improvement and a lower accuracy ceiling due to client data heterogeneity.

7. Results

Metric	Centralized	FedAvg (IID)	FedAvg (Non-IID)	FedAvg + Quant.
MNIST Accuracy (%)	97.0	95.5	92.2	94.7
Fashion-MNIST Accuracy (%)	90.2	89.7	86.1	89.2
Communication per Round (MB)	N/A	320	320	80
Convergence Rounds (95% accuracy)	45	160	185	170
Time per Round (Simulated, seconds)	N/A	6.4	7.3	5.1
Gradient Leakage Detection Rate (%)	N/A	9.0	10.3	2.1

The results confirm that **FedAvg under IID distributions approaches centralized performance**, though with greater latency and data transfer. **Compression and asynchrony help mitigate overhead**, while **non-IID data increases convergence time** and reduces final model accuracy. Nonetheless, **federated models still outperform fully local models** and provide meaningful privacy improvements compared to centralized data storage.

---

8. Discussion

Federated learning represents a viable and scalable alternative to centralized machine learning in environments where data privacy, bandwidth, and user control are critical. Our study reinforces several important observations for practical deployment:

- **Model accuracy is sensitive to data distribution**. Real-world edge devices often contain highly skewed, non-IID data. Addressing this requires strategies like personalized FL, adaptive weighting, and hierarchical aggregation.
- **Communication is the primary bottleneck**. Even with client sampling and quantization, federated systems generate large volumes of data during training. Techniques such as sparsification, update compression, and selective model update scheduling are essential for scalability.
- **FL improves privacy but does not eliminate risk**. Local data never leaves the device, but model updates can still leak sensitive information. Techniques such as **secure aggregation**, **differential privacy**, and **homomorphic encryption** can provide stronger guarantees, though often with additional computational cost.
- **Asynchronous FL is feasible and necessary**. Edge clients differ in computation power and network stability. Supporting asynchronous updates and fault tolerance improves robustness in dynamic environments, a requirement for real-world deployment in mobile or IoT scenarios.
- **System integration matters**. Privacy benefits can be nullified by weak implementations, such as insecure memory access, unverified client software, or metadata leakage. A holistic approach—combining cryptographic protocols, secure software engineering, and federated optimization—is required to realize the full benefits of FL.

---

9. Conclusion

This paper presented an experimental evaluation of federated learning using TensorFlow Federated across 1,000 simulated clients and benchmark datasets. We confirmed that **FedAvg can achieve performance close to centralized training** under IID conditions, with trade-offs in communication cost and convergence time. Techniques like **quantization and client sampling** reduce resource usage with minimal impact on accuracy, making FL more practical for deployment in edge-rich environments.

Under non-IID distributions, model accuracy declined modestly, emphasizing the importance of adaptive aggregation and personalization strategies. Additionally, while federated learning

improves data privacy by design, **gradient inversion attacks remain a concern**—necessitating stronger cryptographic protections and awareness of attack surfaces in model updates.

Overall, **federated learning is a promising architecture** for AI systems that prioritize privacy, user autonomy, and distributed intelligence. Future work should focus on improving **communication efficiency, robustness to non-IID data, and integration of advanced security mechanisms**. With continued research and engineering refinement, FL can become a core pillar of ethical, scalable, and decentralized machine learning systems.

References

1. Bonawitz, K., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., Ivanov, V., ... & van Overveldt, T. (2019). Towards federated learning at scale: System design. *Proceedings of the 2nd SysML Conference*.
2. Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., ... & Zhao, S. (2019). Advances and open problems in federated learning. *arXiv preprint arXiv:1912.04977*.
3. Jena, J. (2018). The impact of gdpr on u.S. Businesses: Key considerations for compliance. International Journal of Computer Engineering and Technology, 9(6), 309-319. https://doi.org/10.34218/IJCET_09_06_032
4. Bellamkonda, S. (2019). Securing Data with Encryption: A Comprehensive Guide. International Journal of Communication Networks and Security, 11, 248-254.
5. McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, 1273–1282.
6. Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2019). Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine*, 37(3), 50–60.
7. Hard, A., Rao, K., Mathews, R., Beaufays, F., Augenstein, S., Eichner, H., ... & Ramage, D. (2018). Federated learning for mobile keyboard prediction. *arXiv preprint arXiv:1811.03604*.
8. Zhao, Y., Li, M., Lai, L., Suda, N., Civin, D., & Chandra, V. (2018). Federated learning with non-IID data. *arXiv preprint arXiv:1806.00582*.
9. Caldas, S., Wu, P., Li, T., Konečný, J., McMahan, H. B., Smith, V., & Talwalkar, A. (2018). LEAF: A benchmark for federated settings. *arXiv preprint arXiv:1812.01097*.
10. Melis, L., Song, C., De Cristofaro, E., & Shmatikov, V. (2019). Exploiting unintended feature leakage in collaborative learning. *IEEE Symposium on Security and Privacy*, 691–706.
11. Kolla, S. (2019). Serverless Computing: Transforming Application Development with Serverless Databases: Benefits, Challenges, and Future Trends. Turkish Journal of Computer and Mathematics Education, 10(1), 810-819. https://doi.org/10.61841/turcomat.v10i1.15043
12. Zhu, L., Liu, Z., & Han, S. (2019). Deep leakage from gradients. *Advances in Neural Information Processing Systems*, 32, 14774–14784.

13. Shokri, R., & Shmatikov, V. (2015). Privacy-preserving deep learning. *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, 1310–1321.
14. Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 308–318.
15. Truex, S., Liu, L., Gursoy, M. E., Yu, L., & Wei, W. (2019). Demystifying privacy in federated learning. *Proceedings of the 2nd ACM International Workshop on Security and Privacy for Artificial Intelligence*, 1–8.
16. Geyer, R. C., Klein, T., & Nabi, M. (2017). Differentially private federated learning: A client level perspective. *arXiv preprint arXiv:1712.07557*.
17. Sattler, F., Wiedemann, S., Müller, K. R., & Samek, W. (2019). Robust and communication-efficient federated learning from non-IID data. *IEEE Transactions on Neural Networks and Learning Systems*, 1–14.
18. Li, X., Huang, K., Yang, W., Wang, S., & Zhang, Z. (2019). On the convergence of FedAvg on non-IID data. *International Conference on Learning Representations (ICLR)*, Workshop Track.