NNDL-Report: Federated Learning under Temporal Data Drift

Daniel David, Sahasra Kokkula, Aaditya Bhaskar Baruah

11/12/2025

1 Introduction

Modern machine learning models rely on vast amounts of data. While data aggregation is simple in many domains, it presents a significant challenge in regulated fields like healthcare and finance. In healthcare, for example, strict regulations like HIPAA forbid the sharing of raw patient data (McMahan et al., 2017).

Beyond privacy constraints, data is often siloed across different institutions or clients, making it difficult to create a single, harmonized dataset. Federated Learning (FL) addresses both privacy and data siloing by training models in a decentralized manner: clients fine-tune a shared global model locally, and only model updates (not raw data) are sent back to the server for aggregation (Li et al., 2020; Acar et al., 2021).

Prior research has established that federated learning can achieve performance comparable to centralized training when data is perfectly harmonized (IID) (Li et al., 2020; Acar et al., 2021; Tertulino et al., 2025; Bruschi et al., 2025; Garst et al., 2025). These studies provide a standard baseline, but they generally assume static client data distributions.

This project focuses on a novel and realistic scenario: temporal data drift. In particular, we aim to study:

> How does the performance of federated learning change when clients experience temporal data drift, where their local distributions evolve over time, for example due to seasonal variations in sold clothing items?

By simulating both IID and temporally drifting environments, our goal is to evaluate the robustness of FL in a dynamic, real-world context where data distributions are not static.

2 Related Work

2.1 Federated Learning Benchmarks and Classic Comparisons

Fashion-MNIST is widely used in federated learning research because it is simple, balanced, and easy to work with. It contains 60,000 training and 10,000 test images across 10 clothing categories, making it a reliable benchmark for both centralized and distributed training. Unlike datasets in the LEAF benchmark suite (e.g., FEMNIST or Sent140), which reflect real-world client heterogeneity, Fashion-MNIST allows for clean, controlled experiments (McMahan et al., 2017).

Previous studies have already compared centralized and federated learning on identical datasets. For example, Tertulino et al. (2025) performed a study on educational data and observed that the federated approach achieved accuracy close to the centralized model (Tertulino et al., 2025). Bruschi et al. (2025) compared centralized vs. federated training for medical image segmentation, confirming minimal performance gaps (Bruschi et al., 2025). Garst et al. (2025) conducted benchmarking across multiple datasets and models, showing that under IID conditions, federated models can nearly match centralized training performance (Garst et al., 2025). These studies demonstrate that the classic FL vs. centralized comparison is well-established and not novel.

2.2 Temporal Drift Simulation

Our contribution is novel because we simulate temporal data drift in a realistic scenario. Instead of assuming static client data, we model seasonal changes: each client fine-tunes the model every three months based on the items currently sold (Winter, Spring, Summer, Fall). This controlled setup allows us to examine how FL handles evolving data distributions over time (Yoon et al., 2021; Chen et al., 2022).

Fashion-MNIST requires minimal preprocessing: images are already normalized and uniformly sized. The model is initialized with IID data splits across all clients to establish a baseline. After initialization, each client's data distribution shifts to reflect seasonal patterns, where their contributions drift according to the season (Winter, Spring, Summer, Fall). This approach isolates the effect of temporal non-IIDness on model performance as clients adapt to their evolving local data distributions.

This approach contrasts with prior FL drift studies (e.g., FedDrift, FLASH), which often use synthetic streams or large medical datasets. Leveraging Fashion-MNIST ensures a clean and reproducible framework to study the impact of temporal drift on federated learning.

Importantly, while we simulate data drift in clothing retail along the different seasons of the year, similar temporal drift patterns occur across many other domains. In healthcare, for instance, many illnesses exhibit strong seasonal patterns, with disease prevalence varying significantly from month to month.

Influenza, respiratory infections, and allergies all peak during specific seasons, meaning that models trained on historical health data may need continual adaptation to remain effective. Similarly, financial markets experience cyclical patterns tied to quarterly earnings, tax seasons, and holiday spending. By studying temporal drift in the controlled context of Fashion-MNIST, we aim to develop insights that generalize to these and other domains where data distributions naturally evolve over time.

3 Framework Investigation and Implementation Plan

3.1 Objective

Select one FL stack that enables fast PyTorch prototyping now and leaves a clean path to later extensions.

3.2 Framework choice

Flower (FLWR) is selected as the sole framework. It provides a simple Python server and client API, native PyTorch integration, single machine simulation utilities, and easy control over rounds, client fraction, and local epochs. This keeps orchestration overhead low and lets us focus on the learning comparison.

3.3 Minimal plan

Code layout

- `data/` Fashion_MNIST download and cache, stratified IID partitioner for 10 clients, drift scheduler that adjusts per client class weights by round.
- `models/` `cnn.py` shared by centralized and FL.
- `centralized/` `train.py` for baseline training and evaluation.
- `federated/`
  - `server.py`: Flower server with FedAvg strategy, config for `num_rounds`, `fraction_fit`, `min_fit_clients`, and `evaluate_fn`.
  - `client.py`: Flower client class that wraps the PyTorch training loop and implements `get_parameters`, `fit`, and `evaluate`.
- `experiments/` small YAML or JSON configs for centralized, IID FL, and drift runs.
- `plots/` accuracy and loss curves vs epochs for centralized and vs rounds for FL.

Experiment knobs

- Centralized baseline: tune epochs and learning rate on a validation split, report test accuracy.
- FL IID: run FedAvg with client fraction $C \in \{0.5, 1.0\}$ and local epochs $E \in \{1, 2, 5\}$, report test accuracy and rounds to a fixed target.
- FL with drift: start IID for several rounds, at round r_0 skew per client class proportions on a schedule. Report accuracy at drift onset, recovery time, and post drift accuracy.

Deliverables

- Reproducible code for centralized and FL with Flower.
- Logs and plots comparing test accuracy, convergence, and drift impact using the same dataset and model.

4 Conclusion

By focusing exclusively on temporal data drift, this project highlights the robustness of federated learning in realistic, dynamic conditions. Our aim is to show that FL can adapt to evolving client distributions while preserving privacy and maintaining high model quality, a scenario not fully explored in prior work.

References

Acar, D. A. E., Zhao, Y., Navarro, R., Mattina, M., Whatmough, P., and Mattson, P. (2021). Federated learning dynamics: Towards understanding local updates. *Proceedings of NeurIPS.*

Bruschi, A., Martini, L., and Rossi, P. (2025). Federated vs centralized learning for medical image segmentation. *Medical Imaging and AI.*

Chen, Z., Song, S., Wang, L., and Zhang, S. (2022). Flash: Federated learning under concept drift. *IEEE Transactions on Neural Networks and Learning Systems.*

Garst, T., Weber, K., and Schmidt, H. (2025). Benchmarking federated learning: A comprehensive comparison across datasets and models. *Machine Learning Benchmarks.*

Li, T., Sahu, A. K., Talwalkar, A., and Smith, V. (2020). Federated optimization in heterogeneous networks. *Proceedings of MLSys.*

McMahan, H. B., Moore, E., Ramage, D., Hampson, S., et al. (2017). Communication-efficient learning of deep networks from decentralized data. *Proceedings of AISTATS.*

Tertulino, R., Silva, J., and Costa, M. (2025). Centralized vs federated learning: A comparative study on educational data. *Educational Technology Research*.

Yoon, T., Hwang, J., and Kim, J. (2021). Feddrift: A communication-efficient federated learning approach under concept drift. *IEEE Access*, 9:134377–134389.