Check for updates

RESEARCH ARTICLE

Federated learning for heterogeneous edge environments_ A multi-domain evaluation on object detection and energy forecasting

[version 1; peer review: awaiting peer review]

Petrina Troulitaki[1], Eirini Bageorgou[1], Sevasti Politi[1], Nikos Oikonomou[1], Raúl López-Blanco[2], Albano Carrera[2], Marta Plaza-Hernandez[2], Evangelos Tsipas[3,4], Panagiotis Dimitrakis[3]

[1]Information Technology for Market Leadership, Athens, Greece
[2]Bisite Research Group, University of Salamanca, Salamanca, Spain
[3]NCSR "Demokritos", Patr. Gregoriou E' & 27 Neapoleos Str., Agia Paraskevi, 153 41, Greece
[4]Democritus University of Thrace, Building A, Kimmeria Campus, Xanthi, 67100, Greece

v1 **First published:** 11 Aug 2026, **6**:290
https://doi.org/10.12688/openreseurope.24462.1
Latest published: 11 Aug 2026, **6**:290
https://doi.org/10.12688/openreseurope.24462.1

Abstract

Federated Learning (FL) has increasingly been promoted as a practical way to train models over distributed edge data while avoiding raw-data centralization, yet most published evaluations have remained confined to a single task and therefore have provided limited evidence about how FL behaves across substantially different workloads. In this paper, a multi-domain empirical study of FL is presented in two representative edge settings: vision-based object detection for connected vehicles and time-series energy forecasting for smart-grid assets. A Flower-based system with explicit client, server, and coordination layers is implemented, and multiple data-partitioning strategies ranging from artificial to deployment-realistic heterogeneity are evaluated. Centralized baselines, isolated local training, federated training, and privacy-enhanced federated training are combined in the study. Across the evaluated domains, three recurring patterns are identified, especially under deployment-aligned client partitions. First, robustness to unseen data is improved by FL in the evaluated settings: in object detection, a 2.7% higher mAP 50 and a 3.4% higher mAP 50–95 are achieved by the federated global model compared with the average independently trained client on hidden validation data. Second, the value of FL is shown to depend strongly on how clients are formed: the benefit of collaboration can be obscured by random partitions, whereas substantial forecasting gains

Open Peer Review

Approval Status *AWAITING PEER REVIEW*

Any reports and responses or comments on the article can be found at the end of the article.

are produced by realistic seasonal and per-turbine partitions, including 16.6%–52.9% MSE reductions for the PV task. Third, much of the collaborative advantage is preserved by adaptive server-side clipping and noise addition in the evaluated PV forecasting scenario, with PV seasonal-partition MSE increased by 21.7% relative to non-private FL while still outperforming isolated training. A more precise picture is therefore provided of when FL is beneficial in heterogeneous edge environments and which evaluation choices most strongly affect the observed outcome.

Keywords
federated learning, edge computing, object detection, energy forecasting, differential privacy, non-IID data

HE This article is included in the Horizon Europe gateway.

Corresponding author: Raúl López-Blanco (raullb@usal.es)

Author roles: Troulitaki P: Conceptualization; **Bageorgou E**: Conceptualization; **Politi S**: Conceptualization; **Oikonomou N**: Conceptualization; **López-Blanco R**: Investigation, Methodology, Resources, Software, Supervision, Validation; **Carrera A**: Data Curation, Formal Analysis, Methodology, Software, Supervision; **Plaza-Hernandez M**: Project Administration, Supervision; **Tsipas E**: Methodology, Resources, Software, Writing – Review & Editing; **Dimitrakis P**: Conceptualization, Resources, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This research has been supported by the project "A catalyst for EuropeaN ClOUd Services in the era of data spaces, high- performance and edge computing (NOUS)", Grant Agreement Number 101135927. Funded by the European Union, views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2026 Troulitaki P *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The author(s) is/are employees of the US Government and therefore domestic copyright protection in USA does not apply to this work. The work may be protected under the copyright laws of other jurisdictions when used in those jurisdictions.

How to cite this article: Troulitaki P, Bageorgou E, Politi S *et al.* **Federated learning for heterogeneous edge environments_ A multi-domain evaluation on object detection and energy forecasting [version 1; peer review: awaiting peer review]** Open Research Europe 2026, **6**:290 https://doi.org/10.12688/openreseurope.24462.1

First published: 11 Aug 2026, **6**:290 https://doi.org/10.12688/openreseurope.24462.1

Introduction

The proliferation of edge devices and the increasing volume of distributed data have created both opportunities and challenges for machine learning systems. Traditional centralized training approaches require aggregating data at a central server, which raises significant concerns regarding data privacy, communication costs, and regulatory compliance with frameworks such as General Data Protection Regulation (GDPR).[1] Federated Learning (FL), introduced by McMahan et al.,[2] addresses these challenges by enabling collaborative model training without raw data leaving the edge devices.

In the FL paradigm, clients train models locally on their private data and share only model updates with a central aggregation server. This iterative process continues until model convergence while maintaining data locality. Despite extensive research on FL, most evaluations remain confined to single application domains, typically image classification or next-word prediction, limiting our understanding of FL's behavior across fundamentally different data modalities and task types.

This paper addresses this gap through a comprehensive multi-domain evaluation of FL across two contrasting use cases: (i) object detection for connected vehicles using YOLOv11 on traffic camera data, representing large-scale computer vision tasks, and (ii) energy load and renewable generation forecasting using neural networks on time-series data, representing regression tasks common in smart grid applications.

Our contributions are threefold:

- We provide a cross-domain FL evaluation spanning both dense visual perception and time-series regression, showing that the same orchestration stack can support these tasks while exposing domain-specific differences in convergence, sensitivity to partitioning, and privacy-performance trade-offs.
- We quantify the effect of client partitioning strategies (random, row-wise, seasonal, and per-turbine) and show that evaluation realism is decisive: deployment-aligned non-independent and identically distributed (non-IID) partitions reveal stronger FL benefits than artificial random splits.
- We assess adaptive server-side clipping and noise addition in a realistic forecasting scenario and show that this privacy-enhancing mechanism remains compatible with clear utility gains over isolated local training.

The remainder of the paper is organized as follows. Section 2 reviews related work on FL, object detection, energy forecasting, and privacy-enhancing mechanisms. Section 3 presents the proposed FL system architecture. Section 4 describes the experimental setup, including datasets, models, partitioning strategies, and DP configuration. Section 5 reports the empirical results for object detection and energy forecasting. Section 6 discusses the main findings, limitations, and threats to validity. Finally, Section 7 concludes the paper and outlines future work.

Related work

This section reviews the main research areas that frame the proposed study. First, foundational FL methods are summarized to establish the aggregation and non-IID learning context. Then, prior work on FL for object detection and energy forecasting is discussed to position the two evaluated application domains. Finally, differential privacy in FL is reviewed as background for the privacy-enhancing configuration evaluated in the forecasting use case.

Federated learning fundamentals

The Federated Averaging (FedAvg) algorithm[2] established the foundational FL training paradigm where clients perform multiple local stochastic gradient descent steps before aggregation. At communication round t, the server updates the global model by computing a data-size-weighted average of the participating client models:

$$w_{t+1} = \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^{(k)},$$

where K denotes the number of participating clients, n_k is the number of local training samples held by client k, $n = \sum_{k=1}^{K} n_k$ is the total number of samples across participating clients, and $w_{t+1}^{(k)}$ represents the model parameters obtained after local training on client k. Subsequent work has addressed challenges including non-IID data distributions,[3,4] communication efficiency,[5] and personalization.[6,7]

FL for object detection

Federated learning for object detection presents unique challenges due to large model architectures and complex loss functions. Recent work has explored FL with YOLO variants,[8,9] though most studies use simulated partitions of benchmark datasets. Our work extends this by evaluating on realistic traffic perception datasets combining KITTI and UA-DETRAC data with real-world class distributions.

FL for energy forecasting

Smart grid applications have increasingly adopted FL for load forecasting[10,11] and renewable generation prediction. These applications naturally involve geographically distributed data sources (buildings, solar parks, wind farms)[12] with heterogeneous patterns, making them ideal FL candidates.

Differential privacy in FL

Differential privacy (DP) provides formal privacy guarantees by adding calibrated noise to model updates.[13] In FL, DP can be applied at the client or server level. Adaptive clipping approaches dynamically adjust the clipping threshold to balance privacy and utility, which we evaluate in our energy forecasting use case.[14]

System architecture

This section describes the system architecture used to support the multi-domain FL experiments. The goal is to present the main functional layers of the framework, the responsibilities assigned to clients and server components, and the training workflow followed during federated execution.

Overall FL framework

Our FL system follows a three-tier architecture built on the Flower framework,[15] which provides framework-agnostic support for both PyTorch and TensorFlow. The architecture, illustrated in Figure 1, comprises:

Client nodes

Each FL client node incorporates three core modules:

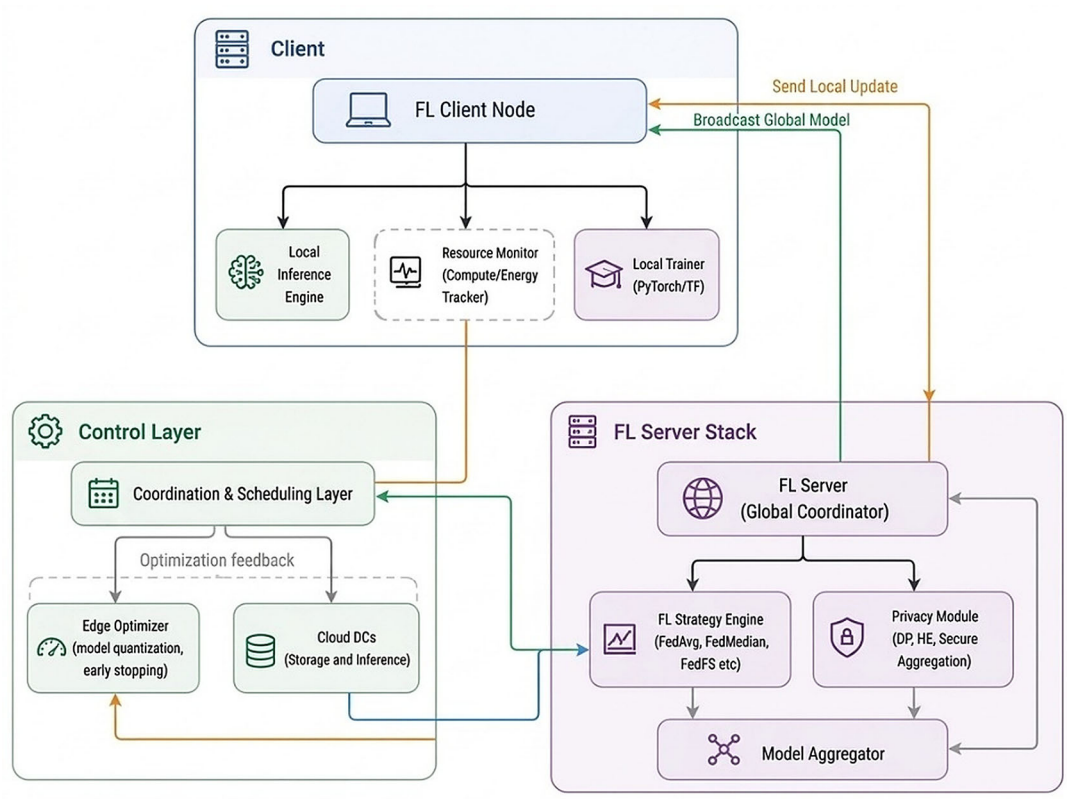

Figure 1. Three-tier federated learning architecture showing client nodes, server stack with privacy module, and control layer.

- **Local Training Module**: Implements on-device training using PyTorch or TensorFlow, supporting both full model fine-tuning and head-only (partial) training.
- **Local Inference Engine**: Performs predictions on unseen data using the current local or global model.
- **Resource Monitoring Module**: Monitors runtime metrics, including CPU utilization, GPU utilization, memory usage, and latency, to optimize edge performance.

Server stack
The FL server stack manages the global training lifecycle:

- **FL Strategy Engine**: Implements aggregation strategies, currently using Federated Averaging (FedAvg), with support for extensions toward adaptive approaches (Q-FedAvg, FedFS).
- **Privacy Module**: Enables privacy-enhancing mechanisms through centralized adaptive clipping[14] and server-side noise addition.
- **Coordination and Scheduling Layer**: Manages configuration distribution, client registration, and feedback incorporation from client-side monitoring.

Edge optimization
An edge optimization module introduces techniques including model quantization and early stopping mechanisms to enhance efficiency on resource-constrained devices.

Training workflow
The FL training workflow proceeds as follows:

1. Server initialization and client registration with unique identifiers
2. Identity verification and secure channel establishment (TLS/gRPC)
3. FL strategy selection (FedAvg) and privacy mechanism configuration
4. Global model initialization and broadcast to all clients
5. Local training on each client's private data
6. Model update transmission to server
7. Server-side aggregation (with optional DP clipping and noise)
8. Convergence check: if not converged, redistribute and repeat from step 5
9. Procedure terminates with finalized global model

Experimental setup
This section describes the experimental design used to evaluate FL across the two application domains. It summarizes the datasets, model configurations, training protocols, client partitioning strategies, and privacy-enhancing configuration used to compare centralized, isolated local, and federated training.

Use case 1: Object detection for connected vehicles
The first use case evaluates FL in a connected-vehicle perception scenario, where visual data are assumed to be collected by distributed roadside units, traffic cameras, or vehicle-associated sensing nodes. In such deployments, raw images may be difficult to centralize because of privacy, bandwidth, ownership, and operational constraints. FL is therefore used as a mechanism for collaboratively improving an object detector while keeping image data local to each client. The task is challenging because traffic scenes contain heterogeneous illumination, viewpoints, object scales, and class frequencies,

making it a representative edge workload for testing whether federated training can improve generalization beyond isolated local fine-tuning.

Dataset

The kitti_ua_detrac_filtered dataset is a curated object-detection benchmark formed by combining KITTI[16] and UA-DETRAC,[17] with the final label space focused on common road-scene categories such as car, pedestrian, cyclist, bus, van, and truck. In the local dataset copy, it contains 11,270 training images and 1,032 validation images, making it a compact multi-source dataset for traffic perception and autonomous-driving experiments. Relative to the original non-filtered kitti_ua_detrac merge, the filtered version is more selective: the non-filtered set contains 15,301 training images and 1,996 validation images, whereas the filtered set removes a substantial portion of KITTI frames while keeping the UA-DETRAC portion unchanged. As a result, kitti_ua_detrac_filtered can be described as a cleaner and more focused version of the merged dataset, designed to reduce redundancy and better align the two source datasets around shared traffic-object categories. For the client-based experiments, the training split was further partitioned using a random equal split into four client subsets, so each client received a near-balanced share of the 11,270 training images, while the 1,032-image validation split remained common for benchmark evaluation.

Model architecture

We employ YOLOv11n (nano variant),[18] which consists of three main components:

- **Backbone**: Feature extraction from input images, producing multi-scale feature maps through convolutional layers.
- **Neck**: Feature fusion across scales, combining semantic information from deep layers with spatial detail from shallow layers.
- **Head**: Detection head generating bounding box coordinates and class probabilities from fused feature maps.

Training configurations

The implemented FL pipeline supports three training modes:

1. **Head-only fine-tuning**: Backbone and neck weights frozen; only detection head parameters updated and shared.
2. **Full fine-tuning**: All model parameters updated and shared (partial results due to infrastructure constraints).
3. **Centralized training**: Traditional training on aggregated data as the upper bound.

FL configuration

The centralized dataset was split into training (80%) and validation (20%), then randomly partitioned into 4 client subsets of approximately equal size. FL training proceeded for 5 rounds with 20 local epochs per round. For comparison, naïve clients trained independently on their local data for 100 epochs (matching centralized training epochs) with early stopping. In practice, the implemented FL pipeline supported three sharing modes: (i) full fine-tuning with full-parameter exchange, (ii) full fine-tuning with head-only exchange, and (iii) head-only fine-tuning with head-only exchange. Due to infrastructure limitations, the experiments reported in this paper focus on the third configuration, which offered the best balance between feasibility and comparability with the centralized benchmark.

Evaluation protocol

Table 3 summarizes the main experimental configuration used across the reported settings. The training data were partitioned across clients, while the held-out validation set was not used for local model updates. The FL global model and the independently trained naïve client models were evaluated on the same hidden validation set, ensuring that the reported generalization comparison is performed under identical unseen-data conditions. The evaluation was designed to separate **in-domain training performance** from **generalization to unseen data**. For UC1, we therefore report: (i) a no-fine-

Table 1. Benchmarking Results on Combined Dataset (YOLOv11n, No Fine-tuning).

Split	Precision	Recall	mAP 50	mAP 50-95
Validation	0.303	0.404	0.299	0.175

Table 2. Energy Forecasting Model Configurations.

Parameter	PG Model	PV Model	Wind Model
Input features	9	20	40
Hidden neurons	128	256	128
Lookback window	10	10	10
Batch size	1	4	1

tuning benchmark to establish task difficulty, summarized in Table 1, (ii) centralized fine-tuning as an upper bound, (iii) the final federated round metrics observed on participating clients and the global model, and (iv) a hidden validation-set comparison between the federated global model and independently trained naïve clients. For UC2, the same logic is followed using mean squared error (MSE): each partitioning strategy is evaluated with centralized training, naïve local training, and FL, and the impact of differential privacy is measured against the non-private FL baseline rather than against centralized training.

Use Case 2: Energy forecasting

Datasets and models

We evaluated three forecasting tasks with configurations summarized in Table 2: active power/load forecasting (PG), photovoltaic generation forecasting (PV), and wind-turbine power forecasting. The task definitions and network sizes were provided by the project partner NCSRD and integrated into the Flower-based FL stack. In all cases, the prediction target is continuous and performance is measured with MSE, making the use case complementary to the classification-style metrics of UC1.

Data Partitioning Strategies

We systematically evaluated four partitioning strategies designed to represent different degrees of data heterogeneity:

- **Random partition**: Data randomly assigned to clients without preserving temporal order. Results in non-IID but unrealistic distributions.
- **Row-wise partition**: Consecutive rows assigned to clients iteratively. Preserves time-series but creates IID distributions (unrealistic for FL).

Table 3. Experimental Configuration Summary.

Category	UC1 Object Detection	UC2 Energy Forecasting
FL framework	Flower	Flower
Model	YOLOv11n	PG, PV, Wind forecasting models
Clients	4	2–4 depending on partition
FL rounds	5	3
Local epochs	20	Task-specific
Main metric	Precision, Recall, mAP 50, mAP 50–95	MSE
Deployment	Local Docker containers	Local Docker containers
Training mode	Head-only FL fine-tuning	Federated regression training
Baselines	Centralized, naïve local, FL	Centralized, naïve local, FL
DP setting	Not applied	Adaptive clipping and noise addition

- **Seasonal partition**: Data split by meteorological season, simulating clients in different geographical regions. Preserves time-series and creates realistic non-IID distributions.

- **Per-turbine partition**: Each wind turbine assigned to a separate client. IID but realistic for multi-turbine wind farm scenarios.

This taxonomy is important for interpretation. Random and row-wise partitions are useful stress tests, but they are weak proxies for real deployments. By contrast, seasonal and per-turbine partitions map naturally to geographic or physical asset boundaries and therefore provide a more credible estimate of the utility practitioners can expect from FL in production settings.

Differential privacy configuration

We implemented centralized adaptive clipping with server-side noise addition following.[14] The clipping threshold is dynamically adjusted after each training round based on the observed update norms. Because this version does not report a complete privacy accountant with final ϵ and δ values, this setting is treated as a privacy-enhancing configuration rather than as a complete formal DP guarantee.

Results

Object detection performance

Centralized fine-tuning baselines

Table 4 presents centralized fine-tuning results serving as performance upper bounds.

Federated learning results

Table 5 shows results after 5 FL rounds with head-only fine-tuning. The global model achieves mAP 50 of 0.630, closely approaching the centralized head-only baseline of 0.664. This corresponds to recovering roughly 94.8% of the centralized head-only mAP 50 while training without pooling raw data.

Generalization analysis

A critical evaluation of model generalization was performed by testing all models on the held-out validation dataset. As shown in Table 6, the FL global model exhibits superior generalization compared to the average of independently trained naïve models.

Table 4. Centralized Fine-tuning Results (100 Epochs).

Metric	Head-only	Full FT
Precision	0.718	0.915
Recall	0.587	0.841
mAP 50	0.664	0.903
mAP 50-95	0.466	0.724
Training time	2.74h	3.38h

Table 5. FL Head-only Results After Round 5.

Metric	C1	C2	C3	C4	Global
Precision	0.680	0.742	0.626	0.605	0.663
Recall	0.563	0.591	0.603	0.595	0.588
mAP 50	0.614	0.669	0.617	0.618	0.630
mAP 50-95	0.456	0.481	0.456	0.458	0.463

The FL global model achieves a 2.7% higher mAP 50 and 3.4% higher mAP 50-95 compared to the naïve client average, despite both approaches having access to the same total training data. In absolute terms, the improvement is modest, but it is consistent across all reported detection metrics and is therefore more informative than comparing only the training-set endpoint. This suggests that the FL aggregation process provides a regularization effect that improves generalization to unseen data.

Energy forecasting results

Impact of Data Partitioning Strategy

Figure 2 and Tables 7 and 8 demonstrate how partitioning strategy affects the relative benefit of FL over independent training.

For the random partition, the report notes that naïve clients can appear deceptively strong because samples from the same temporal process are redistributed without respecting sequence structure. Row-wise splitting restores temporal consistency but still produces nearly IID subsets, so it remains only partially representative of real FL deployments. The seasonal split is the most informative setting because it preserves temporal order *and* induces meaningful client heterogeneity associated with weather-dependent energy behavior.

The aggregated FL model for the PV seasonal partition achieved an overall MSE of 3579.5.

Key findings from the partitioning analysis:

- **Random partitions** (non-IID, no temporal preservation): FL provides limited or no benefit over naïve training, as the unrealistic data distribution fails to represent practical FL scenarios.

Table 6. Validation set performance comparison.

Metric	FL Global	Naïve Avg	Δ
Precision	0.643	0.634	+0.009
Recall	0.549	0.545	+0.004
mAP 50	0.602	0.586	+**0.016**
mAP 50-95	0.420	0.406	+**0.014**

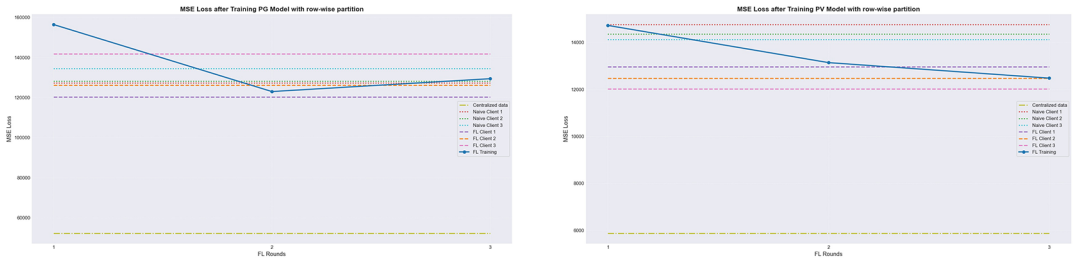

Figure 2. MSE loss evolution for PG and PV models under the row-wise partitioning strategy. Centralized and naïve client baselines are shown as horizontal reference lines, while the FL Training curve reports the average federated loss over three communication rounds.

Table 7. Partitioning strategy summary and qualitative outcome.

Strategy	Time Order	Client Shift	Observed Outcome
Random	Broken	Non-IID, artificial	FL benefit largely obscured
Row-wise	Preserved	Nearly IID	Moderate FL gains
Seasonal	Preserved	Non-IID, realistic	Strongest PG/PV gains
Per-turbine	Preserved	Asset-specific, realistic	Consistent wind gains

Table 8. Seasonal Partition Results for PG and PV Models.

Model	Client	Naïve MSE	FL MSE	Improvement
PG	Client 1	72466.88	45131.96	37.7%
PG	Client 2	49197.01	41007.10	16.6%
PG	Client 3	74940.37	53914.10	28.1%
PG	Client 4	76748.22	65296.38	14.9%
PV	Client 1	6965.10	4709.43	32.4%
PV	Client 2	5797.35	4832.98	16.6%
PV	Client 3	4435.84	2088.12	52.9%
PV	Client 4	5572.41	2687.41	51.8%

Table 9. Per-Turbine Partition: Wind Model Results.

Turbine	Naïve MSE	FL Client MSE
WT-KR01	0.0006	0.00018
WT-KR02	0.0006	0.00035
WT-KR03	0.0007	0.00023
WT-KR04	0.0007	0.00022

- **Row-wise partitions** (IID, temporal preserved): FL shows moderate improvement, with all clients benefiting from collaborative training despite similar data distributions.
- **Seasonal partitions** (non-IID, temporal preserved): FL demonstrates the strongest relative benefits, with individual clients showing 16.6% to 52.9% MSE reduction. This realistic non-IID scenario best motivates federated collaboration.
- **Per-turbine partitions** (IID, realistic): FL outperforms naïve training across the evaluated wind turbines (Table 9).

The seasonal results also show that FL gains are not uniform across clients. Clients 3 and 4 in the PV task benefit much more strongly than Clients 1 and 2, indicating that some local distributions profit more from cross-client information sharing. This heterogeneity is practically relevant because it suggests that reporting only average MSE can conceal substantial client-level variation in benefit.

The aggregated FL global model achieved an overall MSE of 0.0002.

We additionally examined a *seasonal* partition of the wind task, where each turbine model was evaluated across season-defined clients. This setting produced mixed results: KR01 and KR02 benefited from FL, whereas KR03 and KR04 showed cases where the federated client metric did not consistently improve over the naïve baseline. The report attributes this behavior to two plausible causes: first, the absolute errors are very small (often on the order of 10^{-3}), making the comparison sensitive to small fluctuations; second, seasonal splitting may not align well with the turbine-specific feature engineering used in the wind models. This contrast further supports the paper's central claim that partition realism matters as much as algorithm choice.

Seasonal partition results

As shown in Figure 3, the seasonal partition produces a clearer benefit for FL than the row-wise setting. In both PG and PV models, the average FL training loss decreases sharply across communication rounds and ends below the average naïve-client baseline, supporting the claim that deployment-aligned non-IID partitions better reveal the collaborative value of FL.

Discussion

FL's generalization advantage

Across the evaluated settings, FL global models generally exhibited better generalization than independently trained local models, especially when client partitions captured meaningful deployment heterogeneity. In UC1 (object detection), the FL global model showed a 2.7% higher mAP 50 on validation data compared to the naïve average (0.602 vs 0.586). In UC2 (energy forecasting), FL clients achieved lower MSE than their naïve counterparts mainly under realistic seasonal and per-turbine partitioning strategies.

This generalization improvement can be attributed to the implicit regularization effect of model aggregation.[3] By combining diverse gradient updates from clients with heterogeneous data distributions, the global model learns more robust features that transfer better to unseen data, avoiding overfitting to any single client's local distribution.

The critical role of data partitioning

Our systematic comparison of four partitioning strategies reveals that the choice of data distribution fundamentally determines whether FL provides value over independent training:

- Realistic non-IID partitions (seasonal) yield the strongest FL benefits, with individual client improvements ranging from 16% to 53% MSE reduction.
- Artificially created IID partitions (row-wise) show moderate FL benefits, suggesting FL adds value even when clients have similar distributions.
- Unrealistic partitions (random without temporal preservation) fail to demonstrate FL advantages, highlighting the importance of evaluating FL under realistic data heterogeneity conditions.

This finding has practical implications: when deploying FL in real-world scenarios, practitioners should carefully analyze natural data partitions (geographic, temporal, or organizational boundaries) rather than relying on random splits that may artificially inflate or deflate FL performance estimates.

Convergence characteristics

Both use cases demonstrate that FL converges relatively quickly. In UC1, most benefits were realized by round 3, with minimal improvement in subsequent rounds according to the technical report. This suggests potential for communication-efficient training with early stopping, particularly important for bandwidth-constrained edge environments. In practice, this means that careful round-budget tuning may deliver most of the achievable utility while reducing communication, orchestration overhead, and client energy consumption.

Privacy-performance trade-off

The privacy-enhanced FL configuration was evaluated on the PV forecasting task under the seasonal partitioning strategy, with the corresponding utility comparison reported in Figure 4 and Table 10. Non-private FL achieved the lowest MSE, while adaptive server-side clipping and noise addition increased the error relative to the non-private FL baseline. However, the privacy-enhanced configuration still outperformed the average independently trained naïve baseline, indicating that a substantial part of the collaborative learning benefit was preserved even after introducing noise into the aggregation process.

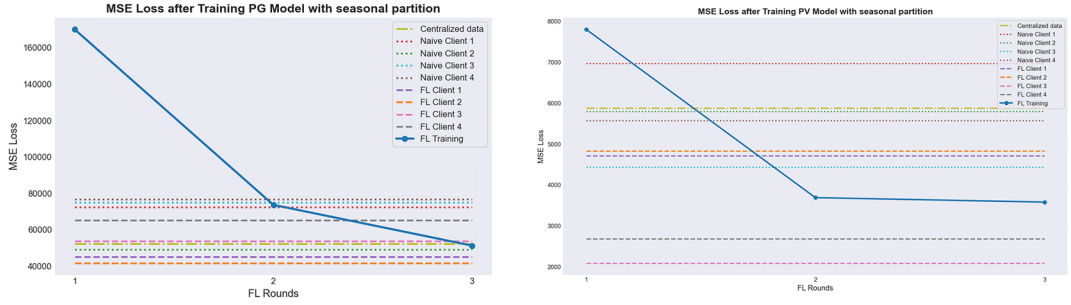

Figure 3. MSE loss comparison for PG and PV models under the seasonal partitioning strategy. Horizontal lines represent centralized, naïve, and FL client references, while the solid curve shows the average FL training loss over three communication rounds.

These results suggest that adaptive server-side clipping and noise addition can offer a promising privacy-utility trade-off in the evaluated PV forecasting scenario. Nevertheless, because this version does not report a complete privacy accountant with final epsilon and delta values, the experiment should be interpreted as an initial privacy-utility analysis rather than as a complete formal Differential Privacy guarantee.

Limitations

Several limitations should be acknowledged:

1. **Scale**: Experiments were limited to 4 clients due to infrastructure constraints. While consistent patterns emerged, larger-scale deployments with tens or hundreds of clients may exhibit different dynamics.

2. **Full model training**: Infrastructure limitations prevented comprehensive full-model FL fine-tuning for YOLOv11, restricting analysis to head-only training.

3. **DP parameter tuning**: Differential privacy experiments were exploratory; this version does not report a complete privacy accountant with final ϵ and δ values, so the results should be interpreted as an initial privacy-utility analysis rather than as a complete formal DP guarantee.

4. **Communication costs**: The study does not yet provide systematic measurements of communication cost, model update size, client-side energy consumption, or round-level latency.

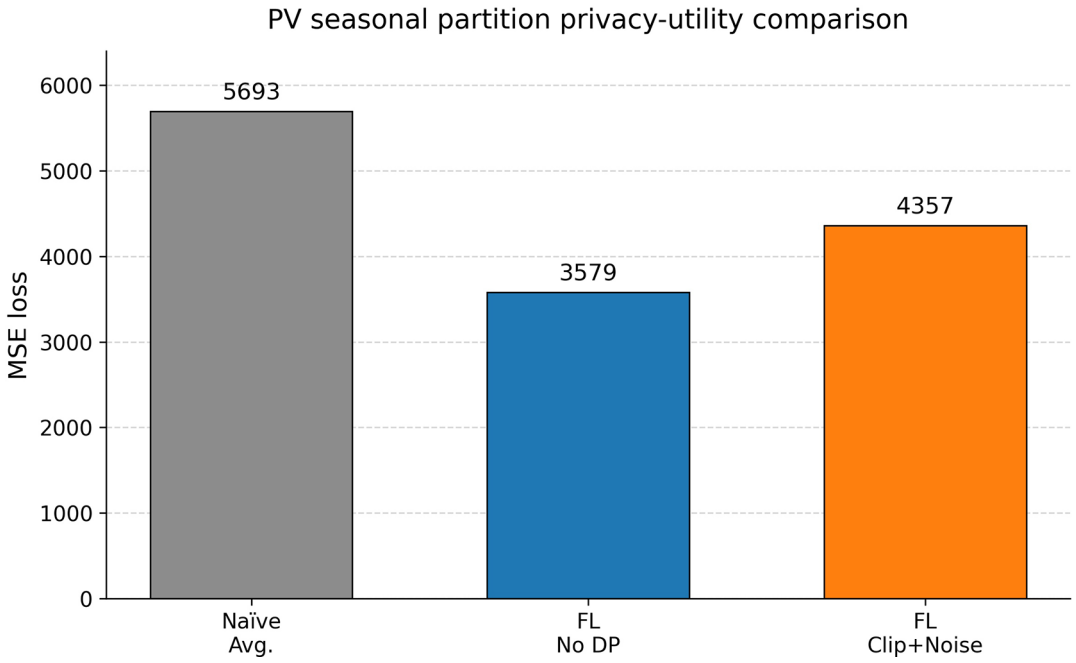

Figure 4. Privacy-utility comparison for the PV model under seasonal partitioning. Adaptive server-side clipping and noise addition increase MSE relative to non-private FL but remain below the average naïve local baseline.

Table 10. Differential Privacy Impact (PV Seasonal, 4 Clients).

Configuration	MSE	Rel. to FL
Centralized (upper bound)	Best	—
FL without DP	3579.48	1.00×
FL with clipping/noise	4357.22	1.22×
Naïve Average	∼5692.68	1.59×

Threats to validity

Several threats to validity should be considered. First, the experiments were conducted with a limited number of clients and FL rounds, which may restrict the generalizability of the observed convergence behavior. Second, the reported results are based on a limited set of runs and may be sensitive to random initialization, client partitioning, and hyperparameter choices. Third, the local Docker-based deployment provides a controlled approximation of edge environments but does not fully capture network variability, intermittent connectivity, or hardware heterogeneity in production deployments. Fourth, the DP experiment should be interpreted as an exploratory privacy-utility evaluation because a complete privacy accountant is not reported in this version. Finally, the study does not yet provide systematic measurements of communication cost, model update size, client-side energy consumption, or round-level latency.

Conclusion

This paper presented a systematic multi-domain evaluation of federated learning across object detection and energy forecasting applications. Our results demonstrate three practical findings for FL deployment in heterogeneous edge environments:

First, federated training improved model generalization in the evaluated settings, particularly when client partitions reflected meaningful deployment heterogeneity. In object detection, the FL global detector achieved 2.7% higher mAP 50 and 3.4% higher mAP 50–95 on hidden validation data compared with the average independently trained client. In energy forecasting, FL delivered substantial MSE reductions under realistic seasonal and per-turbine partitioning strategies.

Second, the choice of data partitioning strategy critically impacts FL effectiveness. Realistic non-IID partitions, such as seasonal splits reflecting geographic heterogeneity yield the strongest FL benefits, while unrealistic random partitions may mask FL's potential. This underscores the importance of evaluating FL under deployment-realistic conditions.

Third, adaptive server-side clipping and noise addition can support privacy-enhanced FL while maintaining significant collaborative learning benefits, achieving a 1.6× MSE improvement over isolated training in the evaluated PV forecasting scenario.

Future work should scale these experiments to production-grade deployments with larger client populations, explore advanced aggregation strategies (FedProx, SCAFFOLD) for extreme non-IID scenarios, and investigate personalized FL approaches that balance global knowledge sharing with local model specialization.

Ethics and consent

No ethical approval or consent was required.

Data availability

The object detection experiments are based on the publicly available KITTI and UA-DETRAC datasets. The original KITTI dataset is available through the KITTI Vision Benchmark Suite, and the original UA-DETRAC dataset is available through the UA-DETRAC benchmark website. The task-specific preprocessing script for the object detection use case (UC1), the client partition scripts, the federated learning configuration files, and the model configuration used in this study have been deposited in the Zenodo open-access repository and are openly available under a CC-BY 4.0 licence at https://doi.org/10.5281/zenodo.21066260.[19]

The energy forecasting experiments were conducted using three datasets. The PG (load forecasting) dataset is available under the conditions described by the project consortium. The PV (photovoltaic generation) and Wind datasets contain operational measurements provided by the Public Power Corporation (PPC, Greece). These datasets are not publicly available because they contain commercially sensitive operational data belonging to PPC and are subject to contractual data-sharing restrictions agreed within the NOUS project consortium.

In addition, the preprocessing pipeline for the energy forecasting use case (UC2) is not publicly released, as it is tightly coupled to the restricted PPC datasets and processing it would expose their proprietary structure. Researchers wishing to access this preprocessing pipeline may submit a request to Evangelos Tsipas at etsipas@ee.duth.gr, subject to the same confidentiality and contractual conditions that govern the underlying PPC datasets.

The study did not involve human participants, personal data, or animal subjects. Therefore, Institutional Review Board (IRB) approval was not required, and no IRB restrictions apply to data sharing. The restrictions arise solely from contractual confidentiality obligations governing proprietary operational data owned by PPC.

Researchers wishing to access the PV and Wind datasets for the purpose of reproducing the results presented in this study may submit a request to Christina Papapostolou at c.papapostolou@ppcgroup.com. Access requests will be evaluated by PPC on a case-by-case basis. If approved, access may be granted subject to the execution of an appropriate data-sharing or non-disclosure agreement and provided that the proposed use complies with the applicable contractual obligations and confidentiality requirements.

References

1. Voigt P, Von dem Bussche A: *The EU general data protection regulation (GDPR): A practical guide.* Cham: Springer International Publishing; 1st ed 2017. vol. **10**.
2. McMahan B, Moore E, Ramage D, *et al.*: **Communication-efficient learning of deep networks from decentralized data.** *Artificial intelligence and statistics.* 2017; 1273–1282.
3. Li T, Sahu AK, Talwalkar A, *et al.*: **Federated learning: Challenges, methods, and future directions.** *IEEE Signal Process. Mag.* 2020; **37**(3): 50–60.
Publisher Full Text
4. Li T, Sahu AK, Zaheer M, *et al.*: **Federated optimization in heterogeneous networks.** *Proceedings of Machine Learning and Systems.* 2020; **2**: 429–450.
5. Konečný J, McMahan HB, Yu FX, *et al.*: **Federated learning: Strategies for improving communication efficiency.** *NIPS workshop on private multi-party machine learning.* 2016.
6. Fallah A, Mokhtari A, Ozdaglar A: **Personalized federated learning with theoretical guarantees: A model-agnostic meta-learning approach.** *Advances in neural information processing systems.* 2020; pp. 3557–3568.
7. Li T, Hu S, Beirami A, *et al.*: **Ditto: Fair and robust federated learning through personalization.** *International conference on machine learning.* 2021; pp. 6357–6368.
8. Liu Y, Ma Z, Liu X, *et al.*: **Federated learning for object detection in autonomous driving: A survey.** *IEEE Trans. Intell. Transp. Syst.* 2022.
9. Zhang X, Li Y, Chen W: **FedDet: Federated learning for object detection with heterogeneous data.** *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.* 2023.
10. Saputra YM, Hoang DT, Nguyen DN, *et al.*: **Energy demand prediction with federated learning for electric vehicle networks.** *IEEE global communications conference.* 2019.
11. Taheri S, Khorasani A, Bakshi MS: **Federated learning for short-term residential load forecasting.** *IEEE Transactions on Smart Grid.* 2022.
12. Chen X, Li H, Liu J: **Federated learning for renewable energy forecasting: A privacy-preserving approach.** *Appl. Energy.* 2021; **304**: 117732.
13. Abadi M, *et al.*: **Deep learning with differential privacy.** *Proceedings of the ACM SIGSAC conference on computer and communications security.* 2016; pp. 308–318.
14. Andrew G, Thakkar O, McMahan B, *et al.*: **Differentially private learning with adaptive clipping.** *Advances in neural information processing systems.* 2021.
15. Beutel DJ, *et al.*: **Flower: A friendly federated learning research framework.** *arXiv preprint arXiv:2007.14390.* 2020.
16. Geiger A, Lenz P, Urtasun R: **Are we ready for autonomous driving? The KITTI vision benchmark suite.** *IEEE conference on computer vision and pattern recognition.* 2012.
17. Wen L, *et al.*: **UA-DETRAC: A new benchmark and protocol for multi-object detection and tracking.** *Comput. Vis. Image Underst.* 2020; **193**: 102907.
Publisher Full Text
18. Jocher G, Chaurasia A, Qiu J: **Ultralytics YOLO.**
Reference Source 2023.
19. Troulitaki P, Bageorgou E, SEVASTI P, *et al.*: **Federated Learning for Heterogeneous Edge Environments: A Multi-Domain Evaluation on Object Detection and Energy Forecasting.** *Zenodo.* 2026.
Publisher Full Text