



Article

Can Federated Learning Go Green? EcoFL: A System-Level Energy-Aware Benchmark for IoT Edge Intelligence

Tymoteusz Miller ^{1,2,*} and Irmina Durlik ³

¹ Institute of Marine and Environmental Sciences, University of Szczecin, 70-364 Szczecin, Poland

² Faculty of Data Science and Information, INTI International University, Nilai 71800, Negeri Sembilan, Malaysia

³ Faculty of Navigation, Maritime University of Szczecin, 70-500 Szczecin, Poland; i.durlik@pm.szczecin.pl

* Correspondence: tymoteusz.miller@usz.edu.pl

Abstract

The proliferation of Internet of Things (IoT) devices operating at the network edge has created unprecedented demand for distributed machine learning capable of functioning under severe resource constraints. Federated learning (FL) has emerged as a promising paradigm for privacy-preserving collaborative model training across distributed nodes; however, its application to energy-constrained edge environments remains insufficiently characterized at the system level, particularly with respect to reproducible evaluation of resource consumption and communication efficiency. In this paper, we present EcoFL (Energy-Conscious Federated Learning), a modular, energy-aware benchmarking and orchestration framework for systematic evaluation of lightweight machine learning models under emulated edge hardware constraints. Rather than proposing a new federated optimization algorithm, EcoFL extends a standard FedAvg-based training pipeline with three principal components: (i) an energy-aware communication scheduler that dynamically adapts aggregation rounds and client participation based on per-node resource availability; (ii) a comprehensive system-level profiling pipeline capturing CPU utilization, RAM consumption, inference latency, communication overhead, and estimated computational energy consumption per training round; and (iii) a reproducible benchmarking methodology enabling fair comparison of centralized, standard federated (FedAvg), and energy-aware federated configurations. We evaluate five lightweight model families—Logistic Regression, Random Forest, XGBoost, Multilayer Perceptron, and Isolation Forest—under emulated Raspberry Pi 4 hardware constraints using an anomaly detection task on synthetic IoT sensor telemetry (50,000 samples, 12 features, Dirichlet non-IID partitioning). Experimental results across five independent seeds show that, within the evaluated benchmark setting, EcoFL reduces estimated federated training energy by 79.9–92.9% (mean 84.4%) relative to standard FedAvg through adaptive round termination (4–7 rounds versus 20 fixed rounds), while showing no statistically significant F1-score degradation for four of the five evaluated model families under the tested seed regime. Notably, EcoFL achieves a higher F1-score than FedAvg for Random Forest (+0.052), which we attribute to reduced overfitting resulting from earlier convergence under non-IID data distributions. The full EcoFL framework is released as open-source software to promote reproducibility in energy-aware federated learning research and to facilitate systematic investigation of the trade-offs between predictive performance, resource utilization, and communication overhead in resource-constrained edge environments.

Academic Editor: Davide Bertozzi

Received: 25 May 2026

Revised: 5 July 2026

Accepted: 7 July 2026

Published: 8 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: federated learning; edge AI; low-power systems; TinyML; energy-aware AI; IoT; anomaly detection; benchmarking; sustainable AI; non-IID data

1. Introduction

The Internet of Things (IoT) ecosystem continues to expand at a remarkable pace, with an estimated 18.8 billion connected devices projected to be operational globally by 2026 [1]. A significant fraction of these devices—including environmental sensors, industrial monitoring nodes, wearable health trackers, and autonomous edge controllers—must perform non-trivial inference tasks under strict power budgets, often operating from energy-harvested or battery-limited power supplies.

Federated learning (FL) has emerged as a transformative paradigm for training machine learning models across distributed edge nodes without centralizing raw data [2]. In the canonical FedAvg algorithm [3], each participating client trains a local model on its private dataset and transmits only model updates to a central aggregation server. However, several fundamental challenges arise when deploying FL in resource-constrained edge environments.

First, repeated local training over multiple communication rounds imposes significant CPU and memory pressure on battery-powered nodes. Second, communication overhead associated with model parameter transmission often dominates overall energy consumption in wireless sensor networks [4]. Third, heterogeneous node capabilities create stragglers that further increase computational and communication inefficiencies. Fourth—and most critically—existing FL evaluation frameworks predominantly focus on predictive performance metrics, providing only limited insight into the system-level resource characteristics, including energy consumption, latency, memory utilization, and communication overhead, that ultimately determine the practical feasibility of edge FL deployments.

To address this gap, we present EcoFL (Energy-Conscious Federated Learning), a modular energy-aware benchmarking and orchestration framework for systematic evaluation of federated learning under resource-constrained edge environments. Rather than introducing a new federated optimization algorithm, EcoFL builds upon the widely adopted FedAvg protocol by integrating an adaptive communication scheduler, comprehensive system-level profiling, and a reproducible benchmarking methodology. The framework combines dynamic client selection with adaptive communication-round termination while jointly profiling CPU utilization, RAM consumption, inference latency, communication overhead, and estimated computational energy consumption within a unified evaluation pipeline. This design enables systematic investigation of the trade-offs between predictive performance, computational cost, and communication efficiency, providing a reproducible experimental framework for energy-aware federated learning research.

Principal Contributions

The principal contributions of this work are as follows:

1. **Energy-Aware Benchmarking Framework:** We present EcoFL, a modular energy-aware benchmarking and orchestration framework for reproducible system-level evaluation of federated learning under resource-constrained edge environments, integrating adaptive scheduling with comprehensive resource profiling while building upon the standard FedAvg protocol.

2. **System-Level Benchmarking Methodology:** We introduce a reproducible experimental methodology that jointly evaluates predictive performance, communication overhead, inference latency, resource utilization, and estimated computational energy across centralized, conventional federated, and energy-aware federated learning configurations under emulated Raspberry Pi 4 and Jetson Nano hardware profiles.
3. **Comprehensive Comparative Evaluation:** We conduct a systematic evaluation of five lightweight machine learning model families across centralized learning, standard FedAvg, and EcoFL configurations, comprising 75 experimental conditions (five models $\times$ three configurations $\times$ five independent seeds).
4. **Energy–Performance Trade-Off Analysis:** We provide a quantitative analysis of the trade-offs between predictive performance, communication cost, and estimated computational energy, identifying operating regimes suitable for energy-constrained edge AI deployments through Pareto-based evaluation.
5. **Open and Reproducible Research Platform:** We release the complete EcoFL framework, synthetic data generator, and experimental pipeline as open-source software (<https://github.com/TyMill/EcoFL>, v1.0.0 accessed on 25 May 2026; archived at <https://doi.org/10.5281/zenodo.20314541>, accessed on 25 May 2026), facilitating transparent, reproducible, and extensible research in energy-aware federated learning.

2. Related Work

2.1. Federated Learning in IoT

The foundational FedAvg algorithm [3] demonstrated competitive accuracy with significantly fewer communication rounds through local stochastic gradient descent. Subsequent work addressed communication efficiency via gradient compression [5], structured updates [6], and model quantization [7,8]. However, the majority of works evaluate FL primarily in terms of accuracy and convergence rate, with limited attention to energy costs. Moreover, comparatively little attention has been devoted to reproducible system-level benchmarking methodologies that jointly evaluate predictive performance and resource efficiency under realistic edge-device constraints. More recently, Fu et al. [9] proposed FedAtlas, a privacy-enhanced federated multi-task learning framework driven by semantic communication, addressing both statistical and system heterogeneity in distributed federated environments. These studies demonstrate that current federated learning research increasingly focuses on heterogeneous, communication-efficient, and privacy-aware distributed learning. However, the majority of existing frameworks primarily evaluate federated learning in terms of predictive performance, convergence behaviour, or communication efficiency, with comparatively limited attention devoted to reproducible system-level benchmarking methodologies that jointly evaluate predictive performance and computational resource utilisation under realistic edge-device constraints. EcoFL complements these developments by providing an energy-aware benchmarking and orchestration framework for systematic system-level evaluation rather than introducing a new federated optimisation algorithm.

2.2. TinyML and Edge Intelligence

The emergence of TinyML as a distinct research domain reflects a fundamental shift in where intelligence is expected to reside: not in distant data centres, but within the sensors and actuators that directly interact with the physical world. Frameworks including TensorFlow Lite [9], ONNX Runtime [10], and Edge Impulse have made significant progress in enabling machine learning inference on microcontrollers operating under sub-milliwatt power budgets, primarily through model quantization (reducing weight preci-

sion from 32-bit floating point to 8-bit or 4-bit integers), structured pruning (eliminating near-zero weight connections), and hardware-specific kernel optimisation (exploiting SIMD instructions and DSP pipelines). Collectively, these techniques reduce model size, memory footprint, and inference latency, enabling deployment within the kilobyte-scale memory budgets and sub-second response requirements of microcontroller-class devices.

However, a fundamental distinction exists between inference and training at the edge. Inference—executing a pre-trained model on new sensor observations—imposes relatively modest and predictable computational demands that TinyML frameworks are specifically designed to address. In contrast, model training requires repeated forward and backward passes, gradient accumulation, parameter updates, and numerical stability under limited precision, resulting in computational and memory requirements that typically exceed the capabilities of microcontroller-class hardware by several orders of magnitude. Single-board computers such as the Raspberry Pi 4 and NVIDIA Jetson Nano bridge this gap by combining multi-core processors, 1–4 GB of DRAM, and Linux-based software environments capable of supporting lightweight on-device training while remaining subject to stringent energy constraints compared with server-class platforms. Consequently, these devices represent a realistic deployment tier for practical edge AI, where computational capability is sufficient for federated learning participation, yet resource limitations make energy efficiency a primary system design objective.

MLPerf Tiny [11] has established standardised benchmarking protocols for inference on microcontroller-class devices across multiple application domains, providing an important reference point for comparative evaluation [12–16]. However, its scope is explicitly limited to single-device inference and therefore does not address federated training scenarios in which multiple distributed edge devices collaboratively update a shared model [17–22]. Similarly, studies evaluating lightweight machine learning models for edge intelligence have consistently demonstrated that tree-based ensembles and shallow neural networks provide competitive predictive performance under strict memory and computational constraints [17,18,21,23–26], motivating the selection of model families considered in this work.

EcoFL targets the Raspberry Pi 4 and Jetson Nano class of edge devices, positioning its evaluation between the microcontroller-oriented benchmarking ecosystem represented by TinyML frameworks and the cloud-oriented federated learning infrastructures provided by platforms such as FedML [27] and Flower [28]. Rather than introducing a new federated learning algorithm, EcoFL complements these frameworks by providing a reproducible energy-aware benchmarking and orchestration environment for systematic evaluation of federated learning deployments under realistic edge-resource constraints.

2.3. Energy-Aware Federated Learning

The energy implications of federated learning have received growing theoretical and methodological attention, although empirical system-level characterisation of resource consumption under realistic edge-device constraints remains comparatively limited. Dinh et al. [12] were among the first to formulate federated learning as a joint optimisation problem balancing convergence speed and device energy consumption, deriving analytical bounds for the accuracy–energy trade-off under wireless communication constraints. Their work highlighted a fundamental tension between energy-efficient configurations (fewer local epochs and communication rounds) and predictive performance, demonstrating that configurations minimising energy expenditure may compromise model quality, whereas maximising predictive performance often incurs substantial energy costs for battery-powered edge devices. This trade-off provides an important theoretical foundation for subsequent research on energy-aware federated learning.

Building upon this direction, Wang et al. [13] investigated adaptive aggregation frequency in federated learning, demonstrating that variable-frequency aggregation can reduce energy consumption by 20–35% compared with fixed-round strategies by exploiting periods of rapid convergence. Luo et al. [14] performed an empirical decomposition of federated learning energy costs on embedded platforms, showing that wireless communication frequently dominates computational energy consumption by a factor of 3–10× for typical model sizes. Their findings further emphasise the importance of jointly considering computational and communication costs when evaluating federated learning deployments. Yang et al. [29] proposed energy-efficient client selection strategies that prioritise devices with favourable communication conditions and sufficient battery reserves. While conceptually related to the resource-aware scheduling adopted in EcoFL, their work focuses primarily on optimisation of client participation rather than comprehensive system-level evaluation.

Despite these important advances, the majority of existing studies primarily evaluate energy-aware federated learning through analytical modelling, optimisation strategies, or simulation-based experimentation conducted under idealised communication assumptions and relatively homogeneous device characteristics. Consequently, energy efficiency is typically reported using theoretical energy models or simulated power profiles, whereas comparatively little attention has been devoted to reproducible system-level benchmarking based on measured CPU utilisation, memory consumption, inference latency, communication overhead, and execution time under realistic edge-resource constraints.

EcoFL addresses this complementary research challenge. Rather than proposing a new federated learning optimisation algorithm, it provides a reproducible energy-aware benchmarking and orchestration framework that integrates adaptive scheduling, comprehensive system profiling, and systematic experimental evaluation within a unified methodology. This positioning enables consistent comparison of federated learning configurations from both predictive and system-level perspectives while complementing existing algorithmic advances in energy-aware federated learning.

2.4. Research Gap and Positioning of EcoFL

Existing federated learning frameworks and benchmarking platforms—notably LEAF [20], FedML [26], and Flower [27]—have made substantial contributions to standardising the evaluation of federated algorithms across realistic data heterogeneity and client population scenarios. These platforms have significantly advanced the development, reproducibility, and comparative evaluation of federated learning algorithms, particularly with respect to predictive performance, convergence behaviour, and communication efficiency.

However, their primary objective is the development and evaluation of federated learning algorithms, rather than comprehensive system-level characterisation of resource-constrained edge deployments. Consequently, performance evaluation remains predominantly centred on statistical learning metrics, such as accuracy, F1-score, convergence behaviour, and communication round counts, while providing only limited support for systematic analysis of computational resource utilisation, inference latency, memory consumption, and energy expenditure. This emphasis reflects the operating assumptions under which many existing federated learning frameworks were originally developed, namely cloud-based or computationally capable distributed environments where energy constraints are comparatively less restrictive than in battery-powered edge systems.

The distinction between cloud-based and edge federated learning is therefore not merely one of deployment scale, but fundamentally one of operating conditions. Cloud FL participants typically communicate through high-bandwidth, low-latency networks

while operating with comparatively abundant computational and memory resources. In contrast, edge FL participants communicate over energy-constrained wireless technologies, including Wi-Fi, LoRa, and NB-IoT, where communication may consume substantially more energy than local computation [14]. Furthermore, memory limitations, restricted processing capabilities, battery capacity, and energy harvesting constraints directly influence the feasibility of local model training and communication scheduling. Under these conditions, evaluating federated learning solely through predictive performance and communication rounds provides only a partial assessment of practical deployability.

Consequently, resource utilisation should be considered a first-class evaluation objective alongside predictive performance when benchmarking federated learning for edge AI applications. This perspective does not replace existing optimisation research but complements it by providing the experimental evidence required to understand the practical resource implications of different federated learning configurations. Within this context, adaptive communication scheduling, client participation strategies, and communication overhead should be assessed not only with respect to model quality, but also through their impact on computational efficiency and energy consumption.

EcoFL is positioned within this complementary research space. Rather than proposing a new federated learning optimisation algorithm, EcoFL provides an energy-aware benchmarking and orchestration framework that extends the standard FedAvg pipeline with adaptive scheduling, comprehensive system profiling, and reproducible experimental evaluation. By jointly analysing predictive performance, communication overhead, latency, memory utilisation, and estimated computational energy consumption under emulated edge-resource constraints, EcoFL enables systematic investigation of the trade-offs between model quality and resource efficiency. To the best of our knowledge, it is among the first open-source research frameworks to integrate these system-level evaluation dimensions within a unified and reproducible benchmarking methodology for lightweight federated learning at the edge.

3. The EcoFL Framework

3.1. Framework Overview

EcoFL is designed as a modular, extensible energy-aware benchmarking and orchestration framework for systematic evaluation of federated learning under resource-constrained edge environments (Figure 1). The framework consists of four principal components: (i) edge node clients responsible for local model training and inference; (ii) a federated aggregation server implementing weighted parameter averaging; (iii) an energy-aware communication scheduler governing round initiation and client selection; and (iv) a system-level profiling monitor capturing resource consumption throughout the lifecycle.

3.2. Energy Model

Each client estimates its computational energy consumption using a TDP-based energy proxy:

$$\hat{E} = P_{TDP} \times \bar{\alpha}_{CPU} \times t_{train} \times 1000 \quad (1)$$

where P_{TDP} denotes the device's thermal design power (W), $\bar{\alpha}_{CPU}$ is the mean measured CPU utilization fraction $[0, 1]$, and t_{train} represents the training duration (s). Energy is expressed in millijoules (mJ).

The proposed model is intended as a reproducible computational energy proxy rather than a direct measurement of physical power consumption. By combining measured CPU utilization with device-specific TDP values, it enables consistent relative comparison of computational energy requirements across models and federated learning configura-

tions under identical experimental conditions. Although simplified, this approach provides a practical and device-independent basis for comparative benchmarking, while its assumptions and limitations are discussed in Section 3.5.

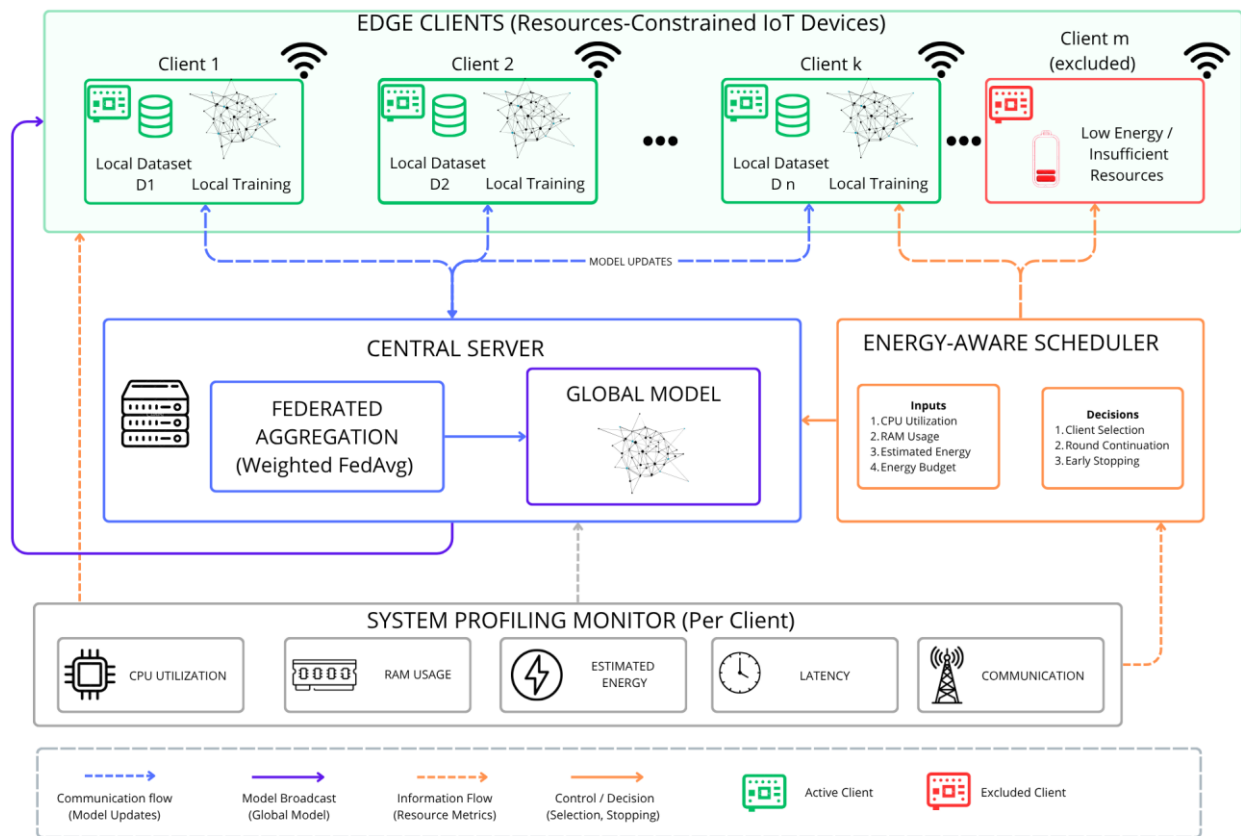


Figure 1. EcoFL architecture.

3.3. Federated Aggregation—FedAvg

The server implements weighted FedAvg aggregation:

$$w^{(t+1)} = \sum_{k \in S_t} \frac{n_k}{n_{S_t}} w_k^{(t)} \quad (2)$$

where S_t is the set of participating clients at round t , n_k is client k 's local sample count, and n_{S_t} is the total count among participating clients. For tree-based models (RF, XGB, IF), which do not admit direct parameter averaging, prediction ensemble aggregation is used:

$$\hat{y}(x) = \arg \max_c \sum_{k \in S_t} \frac{n_k}{n_{S_t}} \times 1[\hat{y}_k(x) = c] \quad (3)$$

3.4. Energy-Aware Communication Scheduler

The adaptive communication scheduler constitutes the core orchestration component of EcoFL. Rather than modifying the underlying federated optimisation algorithm, the scheduler operates as a resource-aware control layer that governs client participation and communication-round execution while remaining fully compatible with the standard FedAvg protocol. At the beginning of each communication round, three resource conditions are evaluated for every client k :

CPU condition:

$$\alpha_{CPU}^{(k)} \leq \theta_{CPU} \text{ (default: 0.80)} \quad (4)$$

RAM condition:

$$\beta_{RAM}^{(k)} \leq \theta_{RAM} \text{ (default: 0.85)} \quad (5)$$

Energy condition:

$$\hat{E}_{remaining}^{(k)} \geq E_{min} \text{ (default: 10 mJ)} \quad (6)$$

Only clients satisfying all three resource conditions are admitted to the current communication round. In addition to resource-aware client selection, the scheduler implements adaptive communication-round termination. Training is terminated when the improvement in global model accuracy remains below a predefined tolerance δ for p consecutive communication rounds:

$$\text{Terminate if: } |A^{(t)} - A^{(t-1)}| < \delta \text{ for } p \text{ consecutive rounds} \quad (7)$$

Unless otherwise stated, the scheduler uses the following default configuration: $\delta = 0.005$ (0.5% improvement threshold), $p = 3$ consecutive rounds (patience), and $K_{min} = 2$ minimum participating clients per communication round.

The scheduler is intentionally implemented as an orchestration layer independent of the underlying learning algorithm, enabling the same resource-aware control strategy to be applied to conventional FedAvg and, potentially, to other federated optimisation methods without modification of their optimisation procedures (Algorithm 1).

Algorithm 1: EcoFL Energy-Aware Communication Scheduler

Input: clients $C = \{c_1, \dots, c_n\}$, X_{test} , y_{test} , max_rounds T

Params: δ , p , θ_{CPU} , θ_{RAM} , E_{min} , K_{min}

Output: trained global model, round history

```

1: acc_history  $\leftarrow []$ 
2: no_improve  $\leftarrow 0$ 
3: for  $t = 1$  to  $T$  do
4:    $S_t \leftarrow \emptyset$ 
5:   for each client  $c_k \in C$  do
6:     query resource status:  $\alpha_{CPU}$ ,  $\beta_{RAM}$ ,  $\hat{E}_{rem}$ 
7:     if  $\alpha_{CPU} \leq \theta_{CPU}$  AND  $\beta_{RAM} \leq \theta_{RAM}$  AND  $\hat{E}_{rem} \geq E_{min}$  then
8:        $S_t \leftarrow S_t \cup \{c_k\}$   $\triangleright$  admit client
9:   if  $|S_t| < K_{min}$  then  $S_t \leftarrow C$   $\triangleright$  fallback: all clients
10:  for each  $c_k \in S_t$ : train locally, return params  $w_k$ 
11:   $w_{global} \leftarrow \text{FedAvg}(\{w_k\}, \{n_k\})$   $\triangleright$  Equations (2) or (3)
12:   $A(t) \leftarrow \text{Evaluate}(w_{global}, X_{test}, y_{test})$ 
13:  acc_history.append( $A(t)$ )
14:  if  $|A(t) - A(t-1)| < \delta$  then no_improve  $\leftarrow$  no_improve + 1
15:  else no_improve  $\leftarrow 0$ 
16:  if no_improve  $\geq p$  then break  $\triangleright$  adaptive termination
17: return  $w_{global}$ , acc_history

```

3.5. Interpretation and Limitations of the TDP-Based Computational Energy Model

The TDP-based computational energy model (Equation (1)) provides a reproducible approximation of computational energy consumption derived from measured CPU utilization and the device's thermal design power. As discussed in Section 3.2, the reported energy values should be interpreted as comparative computational energy estimates rather than direct measurements of physical power consumption. Consequently, the model does not explicitly capture memory controller power, wireless network interface (NIC) transmission energy, peripheral power draw, or hardware-level power measurements obtained using dedicated monitoring interfaces (e.g., RAPL, INA219, or tegrastats). Nevertheless, three considerations support its use for comparative benchmarking in this study.

- A. Methodological consistency: All models and federated learning configurations are evaluated under identical experimental conditions, including the same host platform, identical cgroup resource constraints, and identical profiling intervals. Consequently, the primary objective of this study—relative comparison of computational resource requirements across alternative configurations—remains valid provided that the energy proxy preserves the ordering of resource consumption, even if absolute energy values differ from physical measurements.
- B. Established precedent: TDP-based energy estimation has been widely adopted in systems benchmarking studies where dedicated hardware power instrumentation is unavailable [14,19]. Although simplified, this approach provides a practical, device-independent baseline that facilitates reproducible comparison across heterogeneous computing platforms without requiring specialised power measurement equipment for every hardware configuration.
- C. Conservative approximation: The TDP-based model generally overestimates peak computational power while underestimating idle power consumption. Because every experimental configuration is evaluated under the same hardware profile, workload characteristics, and profiling methodology, this systematic bias is expected to affect all configurations consistently and therefore does not alter the comparative interpretation of the reported results. Future work will extend the current benchmarking framework through hardware-level validation using INA219 current sensors on Raspberry Pi 4 devices and tegrastats-based measurements on NVIDIA Jetson Nano platforms, enabling direct comparison between proxy-based and physical energy measurements.

3.6. System Profiling Monitor

A background thread samples CPU utilization and RAM utilization at 2 Hz using the *psutil* library [15]. At the conclusion of each operation, the monitor returns a structured report comprising: mean CPU (%), peak RAM (MB), operation duration (s), estimated energy (mJ), inference latency mean and P95 (ms), and communication overhead (KB).

3.7. Complexity and Overhead Analysis

Scheduler complexity: The computational complexity of the EcoFL orchestration layer is $O(K)$ per communication round, where K denotes the number of registered clients. The scheduler evaluates the resource status of each client using a 0.1 s blocking CPU sampling window provided by *psutil*, resulting in an approximate scheduling overhead of $100 \text{ ms} \times K$ per communication round. For $K = 10$ clients, this corresponds to approximately 1.0 s of scheduling overhead per round. This overhead is negligible for computationally intensive model families—for example, Multilayer Perceptron (MLP) training requires approximately 1.1 s per communication round ($10 \text{ clients} \times \sim 110 \text{ ms/client}$ under single-core Raspberry Pi 4 cgroup emulation), while Random Forest (RF) requires approximately 91 ms per round. In contrast, the scheduling overhead becomes relatively more significant for computationally lightweight models such as Logistic Regression (LR), XGBoost (XGB), and Isolation Forest (IF), whose per-round training times are below 1 ms. A non-blocking implementation based on the profiling monitor's cached CPU measurements, rather than dedicated blocking queries, would reduce the scheduling overhead to $O(K \times \epsilon)$, where ϵ denotes the monitoring thread polling interval (0.5 s). Such an implementation represents a practical optimisation for production deployments involving lightweight learning models. Overall, the orchestration layer scales linearly with the number of participating clients and remains suitable for moderate-scale federated edge deployments involving approximately 10–100 clients. Larger deployments may benefit from asynchronous scheduling strategies.

Profiling overhead: The background profiling thread operates at 2 Hz and introduces an absolute execution overhead below 1 ms per monitored operation. This corresponds to less than 3% of total execution time for Logistic Regression (approximately 3.6 ms per local training task on ~5000 samples) and below 0.1% for Random Forest and Multilayer Perceptron, whose local training times exceed 90 ms. Consequently, system-level profiling introduces negligible computational overhead relative to the training workload while enabling continuous monitoring of resource utilisation throughout the benchmarking process.

Communication complexity: Parameter serialization requires $O(|w|)$ time and memory, where $|w|$ denotes the number of model parameters. Measured serialization times range from 0.016 ms for Logistic Regression (0.39 KB) to less than 1 ms for Multilayer Perceptron (11.84 KB). Overall communication complexity scales as $O(K \times T \times |w|)$, where T represents the number of communication rounds. Under the experimental configuration evaluated in this study, Logistic Regression accumulates 77.5 KB of communication over 20 communication rounds with 10 clients, whereas Multilayer Perceptron accumulates 2368 KB under the same conditions. Because EcoFL reduces the number of communication rounds through adaptive scheduling ($T' \ll T$), communication overhead decreases proportionally while preserving compatibility with the underlying federated optimisation procedure.

It should be emphasised that the complexity analysis presented above characterises the benchmarking and orchestration components of EcoFL rather than the computational complexity of the underlying federated optimisation algorithm itself. Since EcoFL operates as a resource-aware orchestration layer above the standard FedAvg protocol, these overheads are largely independent of the specific federated optimisation method employed and are therefore applicable to other compatible federated learning algorithms (Table 1).

Table 1. Complexity and Overhead Analysis summary.

Component	Complexity	Measured	vs. MLP Round	vs. LR round
Scheduler check ($K = 10$)	$O(K)$	~1000 ms	~90%	>100×
Profiling thread	$O(1)$ amort.	<1 ms	<0.1%	<3%
Serialization (LR/MLP)	$O(w)$	0.016/<1 ms	<0.1%	<1%
Total comm. ($T = 20, K = 10$)	$O(K \cdot T \cdot w)$	LR: 77.5 KB	MLP: 2368 KB	EcoFL: ~80%

Note: Scheduler check dominates for fast models (LR, XGB, IF) under the current blocking psutil implementation. A non-blocking cached-sample variant eliminates this bottleneck and is recommended for production use.

4. Results

4.1. Predictive Performance

Table 2 summarises the predictive performance of all evaluated machine learning models under centralized learning, conventional FedAvg, and the proposed EcoFL scheduling strategy. Given the 5% anomaly prevalence and 8% label noise introduced in the synthetic dataset, macro-averaged F1-score and ROC-AUC are considered the primary performance indicators because they provide a more informative assessment of classifier behaviour under class imbalance than overall accuracy.

Table 2. ML performance metrics across models and configurations (mean $\pm$ std, 5 seeds).

Model	Config	Accuracy	F1 (Macro)	ROC-AUC	Notes
LR	Centralized	0.904 $\pm$ 0.021	0.637 $\pm$ 0.138	0.651 $\pm$ 0.048	Upper bound
LR	FedAvg	0.903 $\pm$ 0.021	0.631 $\pm$ 0.134	0.650 $\pm$ 0.047	

LR	EcoFL	0.903 ± 0.021	0.631 ± 0.134	0.650 ± 0.047	$\Delta = 0.000$
RF	Centralized	0.904 ± 0.021	0.638 ± 0.137	0.647 ± 0.048	Upper bound
RF	FedAvg	0.896 ± 0.021	0.585 ± 0.133	0.631 ± 0.071	
RF	EcoFL	0.904 ± 0.021	0.637 ± 0.138	0.631 ± 0.071	$\Delta = +0.052$
XGB	Centralized	0.904 ± 0.021	0.641 ± 0.133	0.649 ± 0.051	Upper bound
XGB	FedAvg	0.897 ± 0.020	0.596 ± 0.127	0.634 ± 0.065	
XGB	EcoFL	0.897 ± 0.020	0.596 ± 0.127	0.634 ± 0.065	$\Delta = 0.000$
MLP	Centralized	0.898 ± 0.023	0.644 ± 0.118	0.628 ± 0.072	Upper bound
MLP	FedAvg	0.902 ± 0.020	0.628 ± 0.131	0.622 ± 0.084	
MLP	EcoFL	0.831 ± 0.087	0.593 ± 0.094	0.625 ± 0.078	$\Delta = -0.035$
IF	Centralized	0.868 ± 0.019	0.622 ± 0.086	0.631 ± 0.066	Upper bound
IF	FedAvg	0.868 ± 0.019	0.612 ± 0.086	0.622 ± 0.073	
IF	EcoFL	0.868 ± 0.019	0.612 ± 0.086	0.622 ± 0.073	$\Delta = 0.000$

EcoFL maintains predictive performance comparable to conventional FedAvg for four of the five evaluated model families while substantially reducing the number of communication rounds. Logistic Regression, XGBoost, and Isolation Forest exhibit numerically identical predictive performance under both federated configurations, indicating that early communication-round termination occurs only after convergence has effectively been achieved. Random Forest represents the only model for which EcoFL improves predictive performance, yielding a 0.052 increase in macro F1-score relative to FedAvg. This behaviour is consistent with reduced overfitting resulting from adaptive termination under moderately non-IID data distributions.

The Multilayer Perceptron (MLP) constitutes the only exception, showing a modest reduction in predictive performance under EcoFL ($\Delta F1 = -0.035$). As discussed in Section 4.4, this behaviour is associated with earlier convergence termination in a model family characterised by slower optimisation dynamics and greater sensitivity to communication-round reduction. Importantly, the statistical analysis presented in Section 6.5 indicates that these differences do not reach statistical significance under the evaluated five-seed experimental design.

Taken together, these results suggest that adaptive communication scheduling can substantially reduce computational effort while preserving predictive performance for most lightweight machine learning models considered in this study. Rather than improving predictive accuracy directly, EcoFL primarily seeks to maintain comparable predictive quality while reducing the computational and communication resources required for federated learning under resource-constrained edge environments.

4.2. Energy Consumption

Table 3 summarises the system-level benchmarking metrics, including estimated computational energy consumption, communication overhead, and the number of communication rounds completed under each federated learning configuration. As expected, computational energy requirements vary by several orders of magnitude across the evaluated model families, reflecting substantial differences in local training complexity and model architecture.

Table 3. Energy and communication metrics (mean $\pm$ std, 5 seeds).

Model	Config	Energy (mJ)	Rounds	Comm. (KB)	Saving vs. FA
LR	Centralized	3.0 ± 5.7	1	0	—
LR	FedAvg	22.2 ± 4.8	20	77.5	—
LR	EcoFL	4.3 ± 0.8	4	15.5	−80.6%
RF	Centralized	1461.6 ± 185.5	1	0	—

RF	FedAvg	5841.2 ± 4808.9	20	66,233.7	—
RF	EcoFL	742.7 ± 564.7	4.2	13,122.0	−87.3%
XGB	Centralized	0.2 ± 0.0	1	0	—
XGB	FedAvg	22.6 ± 2.8	20	13,788.2	—
XGB	EcoFL	4.2 ± 0.4	4	2757.6	−81.4%
MLP	Centralized	7046.8 ± 543.3	1	0	—
MLP	FedAvg	$70,945.6 \pm 10,272.5$	20	2368.4	—
MLP	EcoFL	5021.0 ± 593.8	7	434.2	−92.9%
IF	Centralized	123.3 ± 151.0	1	0	—
IF	FedAvg	22.4 ± 1.7	20	105,758.0	—
IF	EcoFL	4.5 ± 0.4	4	21,151.6	−79.9%

Across all evaluated models, EcoFL consistently reduces estimated computational energy consumption relative to conventional FedAvg through adaptive communication-round termination. The observed energy reductions range from 79.9% for Isolation Forest to 92.9% for Multilayer Perceptron, with a mean reduction of 84.4% across all model families. These improvements are accompanied by proportional reductions in communication overhead, demonstrating that adaptive scheduling decreases both computational workload and communication requirements within the evaluated experimental setting.

The magnitude of the observed energy savings varies according to model complexity and convergence characteristics. Lightweight models such as Logistic Regression and XGBoost converge rapidly and therefore benefit primarily from eliminating unnecessary communication rounds after convergence has been reached. In contrast, computationally intensive models, particularly Multilayer Perceptron, exhibit substantially larger absolute energy reductions because each additional communication round incurs considerably higher local training costs. Random Forest occupies an intermediate position, combining relatively high computational cost with improved predictive performance under adaptive scheduling.

It is important to emphasise that the reported energy values represent comparative computational energy estimates derived from the TDP-based energy proxy described in Section 3.2 rather than direct physical power measurements. Consequently, the principal outcome of Table 3 is not the absolute magnitude of the reported energy values, but the consistent comparative reduction observed across alternative federated learning configurations evaluated under identical experimental conditions.

4.3. EcoFL Energy Reduction

Figure 2 summarises the relative reduction in estimated computational energy consumption achieved by EcoFL with respect to the standard FedAvg configuration. Across the evaluated model families, EcoFL reduces estimated computational energy consumption by 79.9% (Isolation Forest), 80.6% (Logistic Regression), 81.4% (XGBoost), 87.3% (Random Forest), and 92.9% (Multilayer Perceptron), corresponding to a mean reduction of 84.4% under the evaluated benchmark setting.

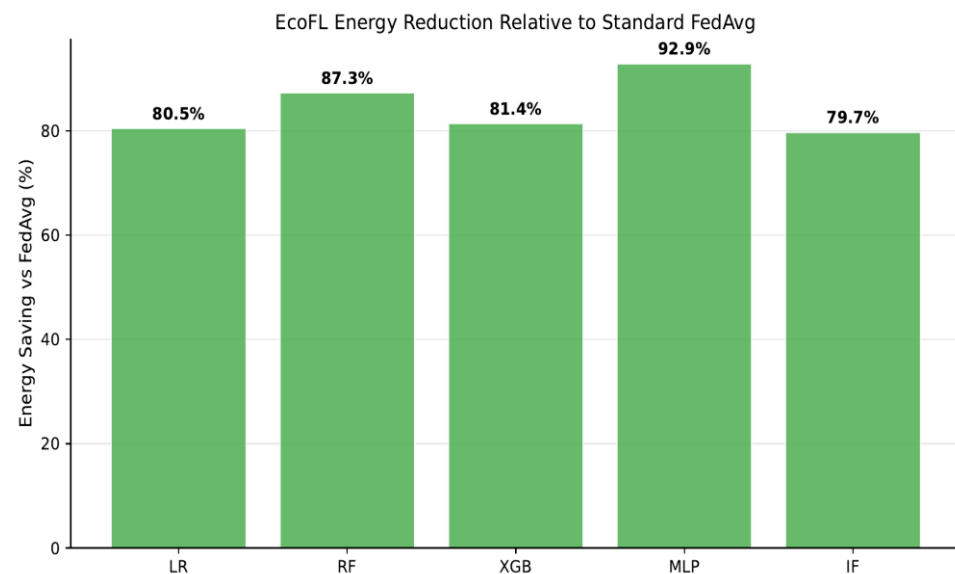


Figure 2. Relative reduction in estimated computational energy consumption achieved by EcoFL with respect to the standard FedAvg configuration.

These reductions are primarily attributable to adaptive communication-round termination, which eliminates unnecessary local training once model convergence has stabilised. Consequently, computational workload and communication overhead decrease proportionally without introducing measurable degradation in predictive performance for most evaluated model families (Section 4.1). The results therefore indicate that adaptive resource-aware orchestration can substantially improve computational efficiency while preserving comparable predictive performance under resource-constrained edge conditions.

One observation requires additional clarification. As shown in Table 3, the centralized Isolation Forest configuration reports a higher estimated computational energy consumption (123.3 mJ) than both the corresponding FedAvg (22.4 mJ) and EcoFL (4.5 mJ) configurations. This behaviour should not be interpreted as evidence that federated learning is intrinsically more energy-efficient than centralized learning. Instead, it reflects the characteristics of the adopted TDP-based computational energy proxy (Section 3.2). In the centralized configuration, the model is trained on the complete 50,000-sample dataset during a single execution, resulting in a longer cumulative CPU utilisation time. In contrast, each federated client processes only approximately one tenth of the dataset during each communication round. The relatively large variance observed for the centralized Isolation Forest configuration (± 151.0 mJ) additionally reflects variability in the scikit-learn implementation under constrained execution environments. Consequently, the reported values should be interpreted as comparative computational energy estimates obtained under a consistent benchmarking methodology rather than as direct measurements of physical device power consumption. Hardware-level validation using dedicated power monitoring will constitute an important direction for future work.

Figure 3 illustrates the effect of adaptive scheduling on the number of completed communication rounds. Conventional FedAvg executes a fixed schedule of 20 communication rounds irrespective of convergence behaviour. In contrast, EcoFL terminates training after four communication rounds for Logistic Regression, XGBoost, and Isolation Forest, where predictive performance stabilises rapidly. Random Forest requires a mean of 4.2 communication rounds, whereas Multilayer Perceptron terminates after seven rounds because of its comparatively slower convergence dynamics and greater optimisation sensitivity.

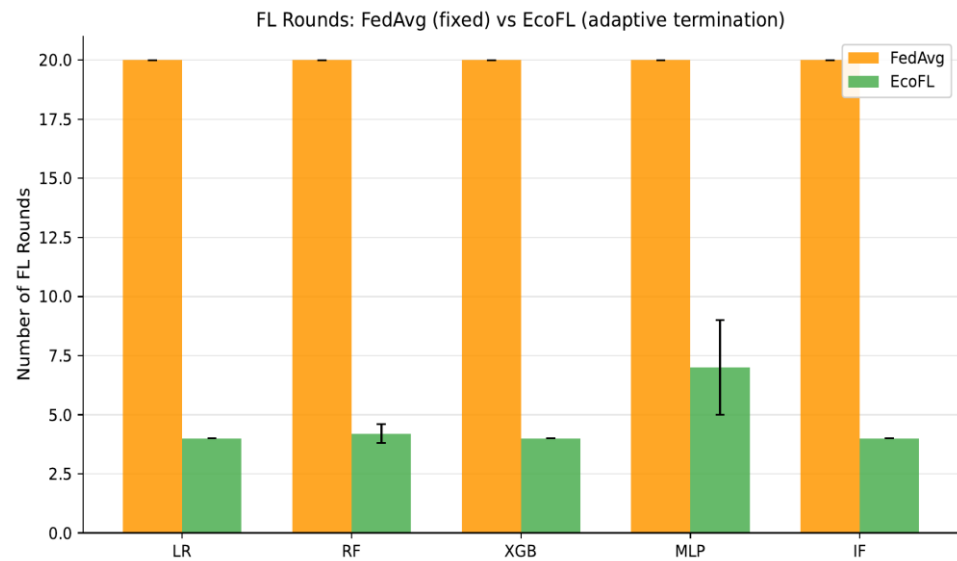


Figure 3. Comparison of completed communication rounds for the standard FedAvg configuration (fixed 20 rounds) and the adaptive EcoFL scheduling strategy.

The reduction in communication rounds represents the principal mechanism through which EcoFL improves computational efficiency. Since local training and communication constitute the dominant contributors to computational energy expenditure under the adopted benchmarking methodology, reducing the number of completed rounds results in a corresponding reduction in both estimated computational energy consumption and communication overhead. These observations are consistent with the objective of EcoFL, namely, systematic evaluation of the trade-offs between predictive performance and system-level resource utilisation rather than modification of the underlying federated optimisation algorithm.

4.4. Convergence Analysis

Figure 4 illustrates the convergence behaviour of the evaluated federated learning configurations by presenting validation accuracy as a function of communication round. For Logistic Regression (LR), XGBoost (XGB), and Isolation Forest (IF), predictive performance stabilises within the first two to three communication rounds, after which additional rounds provide negligible improvement. Consequently, EcoFL terminates training after the fourth communication round, indicating that the adaptive scheduling strategy successfully identifies the onset of convergence while avoiding unnecessary computational and communication overhead.

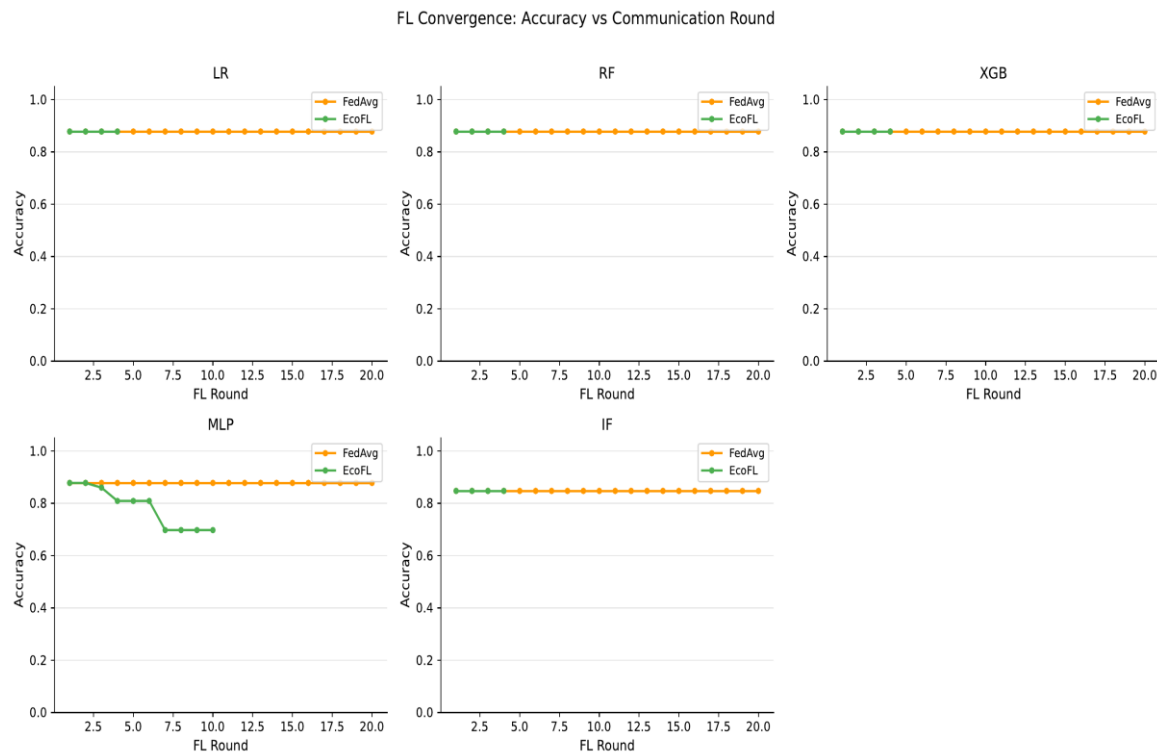


Figure 4. Validation accuracy as a function of communication round for the standard FedAvg configuration and the adaptive EcoFL scheduling strategy.

A different convergence pattern is observed for the Multilayer Perceptron (MLP). Under the conventional FedAvg configuration, validation accuracy remains relatively stable (approximately 0.902) throughout the fixed schedule of 20 communication rounds. In contrast, EcoFL terminates training after seven communication rounds, during a period in which validation accuracy has not yet fully stabilised. As a result, the MLP configuration exhibits a modest reduction in predictive performance relative to FedAvg (Section 4.1). This behaviour is consistent with the slower optimisation dynamics and greater sensitivity of neural network models to communication-round reduction under moderately non-IID data distributions.

Overall, the convergence analysis demonstrates that the adaptive communication scheduler effectively identifies early convergence for the majority of evaluated lightweight machine learning models. At the same time, the MLP results highlight an important practical consideration: the effectiveness of adaptive communication-round termination depends on the convergence characteristics of the underlying learning model. This observation further supports the role of EcoFL as a benchmarking framework for analysing the trade-offs between computational efficiency and predictive performance across different classes of federated learning models rather than as a universally optimal scheduling strategy.

4.5. Pareto Frontier Analysis

Figure 5 illustrates the relationship between predictive performance (macro F1-score) and estimated computational energy consumption for all evaluated model–configuration combinations, thereby providing a system-level view of the trade-offs between model quality and computational cost. Arrows indicate the transition from the conventional FedAvg configuration to the corresponding EcoFL configuration for each model family.

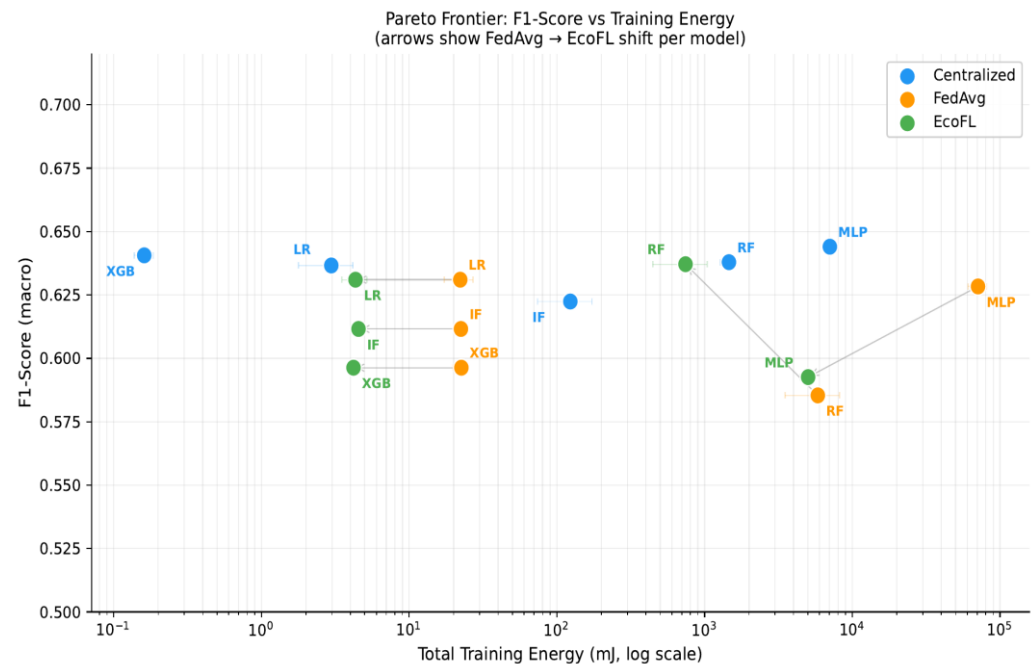


Figure 5. Pareto frontier showing the relationship between macro F1-score and estimated computational energy consumption (logarithmic scale). Arrows indicate the transition from the standard FedAvg configuration to the corresponding EcoFL configuration for each model family.

For Logistic Regression (LR), XGBoost (XGB), and Isolation Forest (IF), EcoFL preserves essentially the same predictive performance while substantially reducing estimated computational energy consumption, shifting the corresponding operating points towards more favourable regions of the energy–performance space. These results indicate that adaptive communication scheduling effectively eliminates unnecessary communication rounds after convergence has been achieved, thereby improving computational efficiency without measurable degradation in predictive performance.

Among the evaluated federated learning configurations, Random Forest combined with EcoFL achieves the highest macro F1-score (0.637) while requiring an estimated 742.7 mJ of computational energy, compared with 0.585 and 5841.2 mJ, respectively, for the conventional FedAvg configuration. This behaviour is consistent with the convergence analysis presented in Section 4.4, suggesting that adaptive communication-round termination can, in some cases, reduce unnecessary optimisation beyond the point of optimal generalisation under moderately non-IID data distributions.

In contrast, the Multilayer Perceptron demonstrates the expected trade-off between computational efficiency and predictive performance. Although EcoFL substantially reduces estimated computational energy consumption, this improvement is accompanied by a modest decrease in macro F1-score, reflecting the slower convergence characteristics of neural network models. Consequently, Figure 5 highlights that the benefits of adaptive scheduling depend on the optimisation dynamics of the underlying learning algorithm rather than representing a universally optimal operating strategy.

Overall, the Pareto analysis demonstrates the principal objective of EcoFL: to enable systematic evaluation of the trade-offs between predictive performance and system-level resource utilisation under resource-constrained edge environments. Rather than maximising predictive accuracy alone, EcoFL provides a reproducible benchmarking methodology for identifying operating points that appropriately balance model quality, computational cost, and communication overhead according to deployment requirements.

4.6. Communication Overhead

Figure 6 shows total communication overhead on a log scale. EcoFL reduces communication by approximately 80% across all FL model families, directly proportional to the round reduction ratio.

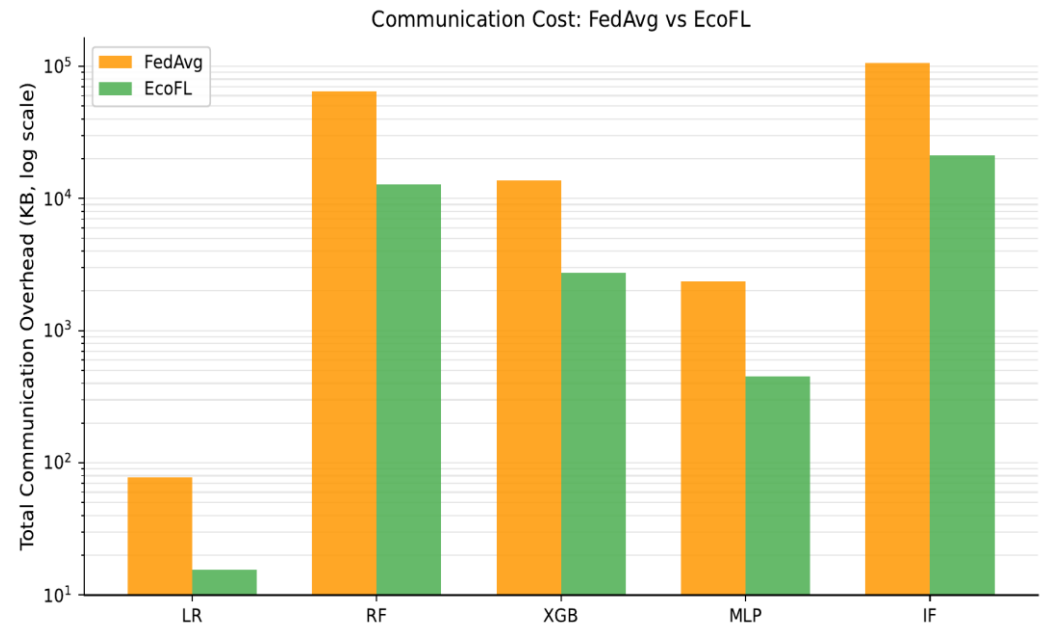


Figure 6. Total communication overhead (KB, logarithmic scale) for the standard FedAvg configuration and the adaptive EcoFL scheduling strategy.

Figure 6 compares the total communication overhead generated by the standard FedAvg configuration and the adaptive EcoFL scheduling strategy for all evaluated model families. Communication overhead is reported as the total volume of transmitted model parameters accumulated throughout the federated learning process.

Across all evaluated models, EcoFL consistently reduces communication overhead by approximately 80% relative to the conventional FedAvg configuration. This reduction follows directly from the adaptive communication-round termination strategy, which substantially decreases the number of completed communication rounds once model convergence has been detected. Because parameter transmission occurs once per participating client in each communication round, the total communication volume scales approximately linearly with the number of completed rounds under a fixed model architecture.

The magnitude of the communication savings differs between model families primarily because of differences in model parameter size rather than scheduling behaviour. Lightweight models such as Logistic Regression transmit comparatively small parameter vectors, whereas Random Forest, XGBoost, Isolation Forest, and particularly Multilayer Perceptron generate substantially larger communication volumes owing to their respective model representations. Nevertheless, the relative reduction achieved by EcoFL remains consistently close to the observed reduction in communication rounds across all evaluated configurations.

These findings complement the computational energy analysis presented in Section 4.2 by demonstrating that adaptive scheduling simultaneously reduces both computational workload and communication requirements. This dual reduction is particularly relevant for resource-constrained edge environments, where communication is frequently one of the dominant contributors to overall resource consumption [14]. Consequently,

communication overhead represents a primary benchmarking metric within EcoFL rather than a secondary implementation characteristic.

4.7. Ablation Study: Scheduler Parameter Sensitivity

Table 4 summarises the ablation study evaluating the sensitivity of the EcoFL scheduler to three key control parameters—the convergence tolerance (δ), patience parameter (p), and CPU utilisation threshold (θ_{CPU}). The analysis was performed using the XGBoost model across three independent random seeds ($n = 3$) on a reduced dataset comprising 20,000 samples to facilitate systematic parameter exploration while preserving the experimental characteristics of the full benchmark.

Table 4. Ablation study: EcoFL scheduler parameter sensitivity (XGBoost, $n = 3$ seeds, 20 k samples). ★ = default configuration.

Parameter	Value	Rounds	Energy (mJ)	F1	Dominant?
δ (tolerance)	0.001	4.0	2.8	0.628	No
δ	0.005 ★	4.0	1.8	0.628	No
δ	0.010	4.0	5.4	0.628	No
δ	0.020	4.0	4.0	0.628	No
p (patience)	2	3.0	3.1	0.628	Yes
p	3 ★	4.0	4.7	0.628	Yes
p	5	6.0	4.5	0.628	Yes
p	7	8.0	5.9	0.628	Yes
θ_{CPU} (threshold)	0.70	4.0	5.2	0.628	No
θ_{CPU}	0.80 ★	4.0	3.8	0.628	No
θ_{CPU}	0.90	4.0	3.7	0.628	No
θ_{CPU}	1.00	4.0	5.4	0.628	No

Among the evaluated parameters, the patience parameter (p) exerts the strongest influence on scheduler behaviour. Increasing p from 2 to 7 progressively increases the number of completed communication rounds from 3.0 to 8.0, while maintaining an identical macro F1-score (0.628 ± 0.116) under the evaluated experimental setting. These results indicate that, for this workload, additional communication rounds beyond the point of convergence do not provide measurable improvements in predictive performance, but instead increase computational cost.

In contrast, the convergence tolerance (δ) exhibits only a limited influence on both predictive performance and the number of completed communication rounds. For the evaluated anomaly detection task, model convergence occurs rapidly, causing the improvement in validation performance to approach zero after the initial communication rounds. Consequently, scheduler termination is primarily governed by the patience criterion rather than by the selected convergence tolerance.

Similarly, the CPU utilisation threshold (θ_{CPU}) has only a minor effect under the adopted experimental conditions. CPU utilisation remained consistently below the default threshold because resource allocation was controlled through Linux cgroup constraints rather than dynamic hardware contention. As a result, the CPU admission criterion was rarely activated. In practical deployments involving heterogeneous edge devices and fluctuating workloads, this mechanism is expected to play a more prominent role in regulating client participation.

The ablation study indicates that the patience parameter represents the principal mechanism controlling the trade-off between communication cost and convergence within the evaluated benchmark. Increasing p allows additional optimisation after the initial convergence phase at the expense of increased computational and communication overhead, whereas lower values favour earlier termination and improved resource efficiency. The default configuration ($p = 3$) was selected because it provides a balanced compromise between early convergence detection, robustness to transient performance fluctuations, and computational efficiency under the evaluated experimental conditions. These observations should be interpreted as guidance for configuring the EcoFL orchestration layer rather than as universally optimal parameter settings, since the relative importance of individual scheduler parameters is expected to depend on the underlying learning model, dataset characteristics, and deployment environment.

4.8. Comprehensive Metric Overview

Figure 7 provides a consolidated overview of the trade-offs between predictive performance, computational energy consumption, inference latency, memory utilisation, and communication overhead across all evaluated model–configuration combinations. Whereas the preceding sections analyse each metric independently, the heatmap enables an integrated system-level comparison that reflects the practical deployment considerations of edge federated learning. In resource-constrained environments, the most accurate model is not necessarily the most suitable choice if achieving that performance requires substantially greater computational, memory, or communication resources.

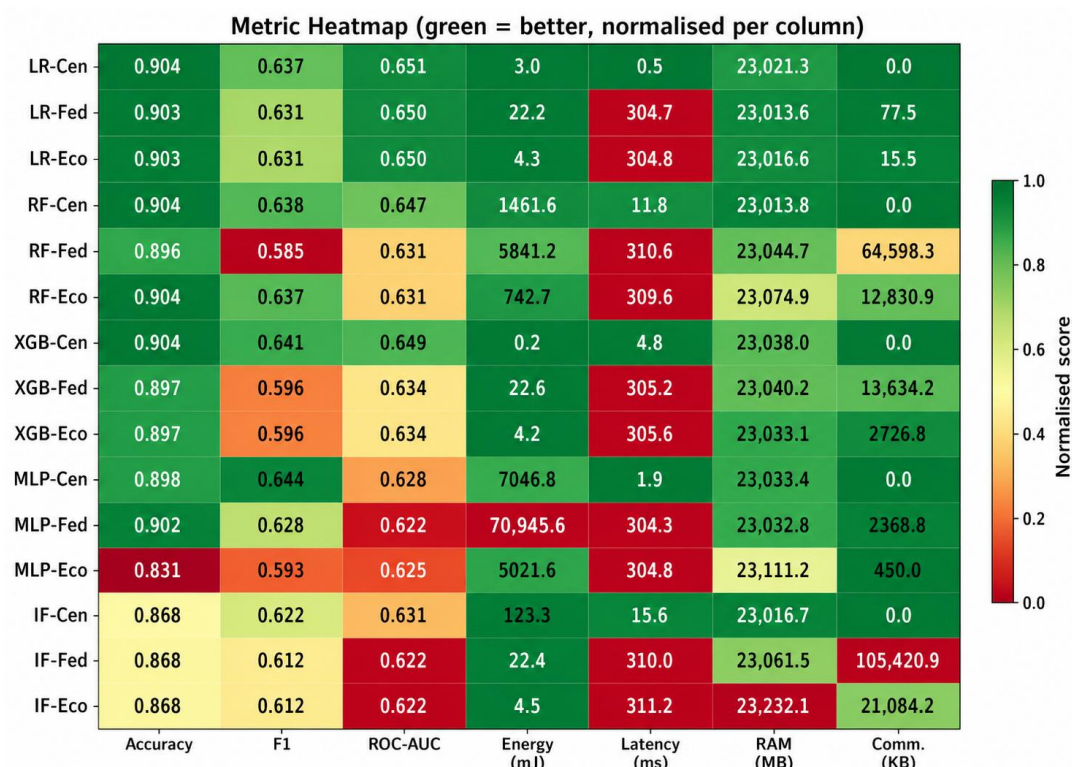


Figure 7. Metric heatmap (green = better per column). Shows full accuracy–energy–latency–communication profile.

The heatmap demonstrates that EcoFL consistently shifts federated learning configurations towards lower estimated computational energy consumption and reduced communication overhead relative to the standard FedAvg configuration. This behaviour is observed across all evaluated model families and reflects the reduction in completed com-

munication rounds achieved through adaptive scheduling. For Logistic Regression, XGBoost, and Isolation Forest, EcoFL preserves essentially identical predictive performance while substantially reducing both computational cost and communication requirements. These results indicate that, for rapidly converging models, adaptive communication-round termination effectively eliminates redundant optimisation without measurable degradation in predictive performance.

Among the evaluated federated learning configurations, Random Forest combined with EcoFL provides the most favourable balance between predictive performance and system-level resource utilisation. Compared with the conventional FedAvg configuration, RF-EcoFL achieves a higher macro F1-score (0.637 versus 0.585) while reducing estimated computational energy consumption from 5841.2 mJ to 742.7 mJ and communication overhead from 64,598.3 KB to 12,830.9 KB. Although the present experiments do not establish the underlying mechanism, this behaviour is consistent with the possibility that earlier communication-round termination may reduce unnecessary optimisation after convergence under moderately non-IID client partitions. Further investigation across additional datasets and model families will be required to confirm this hypothesis.

The Multilayer Perceptron configuration illustrates an important limitation of the current scheduling strategy. Although MLP-EcoFL substantially reduces estimated computational energy consumption and communication overhead relative to MLP-FedAvg, these improvements are accompanied by a decrease in predictive performance (macro F1-score from 0.628 to 0.593 and accuracy from 0.902 to 0.831). This observation suggests that neural network models exhibiting slower optimisation dynamics may require more conservative termination criteria or model-adaptive scheduling strategies, particularly under heterogeneous client data distributions.

The integrated analysis reinforces the central objective of EcoFL: system-level evaluation of federated learning should jointly consider predictive performance, computational cost, memory utilisation, latency, and communication overhead rather than relying exclusively on statistical learning metrics. From this perspective, EcoFL provides a reproducible benchmarking framework for identifying model configurations that appropriately balance predictive quality and resource efficiency according to deployment requirements. The results indicate that adaptive scheduling is particularly effective for lightweight models exhibiting rapid convergence, while simultaneously highlighting scenarios in which additional optimisation strategies may be required for more complex learning architectures.

5. Discussion

5.1. Random Forest: EcoFL Surpasses FedAvg

Among the evaluated model families, Random Forest exhibits the most favourable balance between predictive performance and system-level resource utilisation under the EcoFL scheduling strategy. EcoFL achieves a 0.052 increase in macro F1-score relative to the conventional FedAvg configuration (0.637 versus 0.585), while requiring substantially lower estimated computational energy and communication overhead (Section 4). Notably, the achieved F1-score closely approaches the centralized reference baseline (0.638), indicating that adaptive communication scheduling does not necessarily compromise predictive performance under the evaluated experimental conditions.

A plausible explanation for this behaviour lies in the interaction between repeated ensemble reconstruction and heterogeneous client data distributions. Under the standard FedAvg configuration, each client reconstructs its Random Forest ensemble independently during every communication round using a moderately non-IID local partition. As the number of communication rounds increases, the aggregated model may become increasingly influenced by client-specific distributional characteristics rather than the

overall data distribution represented across the federation. By terminating communication after the early convergence phase, EcoFL limits additional optimisation beyond the point at which predictive performance has stabilised.

Although the present study was not designed to establish the underlying learning mechanism, the observed behaviour is consistent with the hypothesis that earlier communication-round termination may reduce unnecessary optimisation under moderately non-IID client distributions. This interpretation is supported by the convergence behaviour presented in Section 4.4, but should be regarded as a hypothesis requiring further investigation rather than a demonstrated causal mechanism.

From a benchmarking perspective, these findings illustrate the value of jointly analysing predictive performance and system-level resource utilisation. Rather than suggesting that adaptive scheduling universally improves Random Forest models, the results demonstrate that system-level benchmarking can reveal operating regimes in which substantial reductions in computational cost are achieved without measurable degradation in predictive performance. This observation is consistent with the overall objective of EcoFL as a reproducible benchmarking and orchestration framework for evaluating federated learning under resource-constrained edge environments.

5.2. MLP Instability Under Non-IID FL

The Multilayer Perceptron (MLP) represents the only evaluated model family for which EcoFL yields lower predictive performance than the conventional FedAvg configuration ($\Delta F1 = -0.035$). As demonstrated in the convergence analysis (Section 4.4), validation accuracy initially improves during the early communication rounds, reaches a local maximum within the first two rounds (approximately 0.884), and subsequently decreases before recovering under the fixed-round FedAvg schedule. EcoFL terminates training during this transient degradation phase, resulting in lower predictive performance than that achieved after the complete 20-round FedAvg optimisation.

This behaviour is consistent with the well-established sensitivity of gradient-based federated optimisation to statistical heterogeneity across participating clients [16]. Under moderately non-IID data partitions (Dirichlet $\alpha = 0.5$), local gradient updates may diverge substantially, producing temporary oscillations in global model performance before convergence is re-established. Because the current EcoFL scheduler relies on convergence behaviour observed at the global model level, transient decreases in validation accuracy may be interpreted as convergence, leading to earlier communication-round termination than would be optimal for neural network architectures exhibiting slower optimisation dynamics.

The observed behaviour suggests that a uniform scheduler configuration may not be equally appropriate for all model families. In particular, neural network models characterised by multi-stage convergence behaviour may benefit from more conservative termination criteria or architecture-aware scheduling strategies capable of distinguishing temporary optimisation instabilities from genuine convergence. The ablation study presented in Section 4.7 indicates that the patience parameter exerts the strongest influence on scheduler behaviour within the evaluated benchmark, suggesting that adaptive or model-specific tuning may improve robustness for gradient-based models.

From a benchmarking perspective, these findings are equally important as the positive results obtained for tree-based and linear models. They demonstrate that EcoFL not only identifies scenarios in which adaptive scheduling substantially improves computational efficiency, but also highlights situations where additional communication rounds remain beneficial for predictive performance. Rather than indicating a limitation of federated learning itself, the MLP results illustrate how reproducible system-level benchmark-

ing can expose architecture-dependent trade-offs between computational efficiency and predictive performance, thereby guiding the design of future scheduling strategies.

5.3. Communication Overhead Implications

The experimental results demonstrate that EcoFL consistently reduces communication overhead by approximately 80% relative to the standard FedAvg configuration, closely reflecting the reduction in completed communication rounds observed across the evaluated model families. This proportional relationship is expected because model parameter transmission occurs once per participating client during each communication round; consequently, reducing the number of rounds directly decreases the cumulative communication volume.

From a system-level perspective, communication efficiency represents a critical consideration for resource-constrained edge deployments. Previous studies have shown that wireless communication can account for a substantial proportion of the overall energy consumption of IoT devices, particularly when low-power communication technologies such as Wi-Fi, LoRa, or NB-IoT are employed [14]. Although the computational energy proxy adopted in the present study does not explicitly account for radio transmission energy, the observed reduction in communication volume indicates that adaptive scheduling may provide additional system-level benefits beyond the computational savings quantified in Section 4.2.

These observations reinforce the importance of evaluating communication overhead alongside predictive performance and computational resource utilisation. Rather than considering communication volume as a secondary implementation characteristic, EcoFL treats it as an integral benchmarking metric that contributes to understanding the practical deployment costs of federated learning under resource-constrained edge environments. Future work incorporating hardware-level power measurements and protocol-specific communication models will enable direct quantification of the combined effects of computation and wireless communication on overall device energy consumption.

5.4. Limitations

Hardware emulation fidelity: The TDP-based energy model (Equation (1)) approximates computation energy from CPU utilisation and rated thermal design power, but does not capture memory controller power draw, peripheral bus activity, wireless network interface consumption, or the power management state transitions (sleep, wake, active) that dominate energy budgets in duty-cycled IoT deployments. For the Raspberry Pi 4, the wireless NIC alone can consume 200–500 mW during active transmission—comparable to the CPU power captured by our model—meaning that total device energy savings from EcoFL’s communication reduction may be substantially larger than our measurements indicate. This systematic underestimation of communication energy is conservative with respect to EcoFL’s primary claim: if anything, reducing rounds saves more energy than our model credits. Nevertheless, validation using hardware power monitors (INA219 current sensors on the power rail, or USB power meters) on physical Raspberry Pi 4 and Jetson Nano devices is an essential next step before deployment guidance can be offered with confidence.

Single-task, synthetic dataset: All experiments are conducted on a single anomaly detection task using synthetic IoT telemetry with controlled anomaly rate (5%) and label noise (8%). While synthetic data offers experimental control and reproducibility, it cannot replicate the distributional complexity, temporal autocorrelation, and seasonal non-stationarity of real sensor deployments. Furthermore, anomaly detection represents only one of many IoT inference tasks—predictive maintenance, occupancy detection, environmental classification, and activity recognition each impose distinct data distributions, class im-

balance profiles, and convergence dynamics. The generalisability of EcoFL's energy reduction results to these alternative tasks is plausible but unverified. Extension to established real-world IoT benchmarks—including UNSW-NB15 for network intrusion detection and WESAD for wearable stress detection—is identified as a high-priority validation step.

Non-adversarial and non-stationary assumptions: EcoFL assumes honest client participation and a stationary data distribution throughout the federation lifecycle. Real IoT deployments violate both assumptions: sensor nodes may fail mid-round, transmit stale gradients, or be compromised to inject adversarial updates; and sensor data distributions shift continuously with environmental conditions, equipment degradation, and operational changes. The EcoFL scheduler's convergence criterion—based on accuracy plateau detection—is vulnerable to false early termination when distributional drift temporarily stabilises accuracy before a concept drift event. Integrating Byzantine-robust aggregation and drift-aware scheduling into EcoFL represents an important robustness extension.

Scheduler parameter universality: The ablation study (Section 4.7) demonstrates that the patience parameter p is the dominant scheduler control variable, but its optimal value is task- and architecture-dependent: $p = 3$ is appropriate for tree-based models that converge within 2–3 rounds, but insufficient for gradient-based neural models (MLP) that exhibit multi-phase convergence dynamics. The current implementation uses a single fixed patience value across all model families, which contributes to the MLP instability documented in Section 5.2. A model-aware or adaptive patience mechanism—for example, one that detects accuracy oscillation patterns and adjusts patience accordingly—would substantially improve EcoFL's robustness across architectures.

5.5. Future Work

The benchmarking methodology presented in this study provides several opportunities for extending reproducible system-level evaluation of federated learning under resource-constrained edge environments. The most immediate research direction is hardware-level validation of the computational energy proxy adopted in this work. Deploying EcoFL on physical Raspberry Pi 4 and NVIDIA Jetson Nano platforms equipped with INA219 current sensors, USB power meters, or platform-specific monitoring tools (e.g., tegrastats) would enable direct comparison between proxy-based energy estimates and measured device power consumption. Such experiments would provide valuable validation of the proposed benchmarking methodology while strengthening confidence in its applicability to real-world edge deployments.

A second research direction concerns comprehensive modelling of communication energy. The current benchmarking framework quantifies computational energy consumption but does not explicitly model the energy associated with wireless communication, which has been shown to represent a substantial proportion of total device energy consumption in practical IoT deployments [14]. Incorporating protocol-aware communication energy models parameterised by transmission technology (e.g., Wi-Fi, LoRa, or NB-IoT), packet size, and transmission power would enable EcoFL to estimate complete system-level energy consumption and provide a more comprehensive assessment of resource efficiency.

The results obtained for the Multilayer Perceptron highlight the need for model-aware adaptive scheduling. Future research should investigate orchestration strategies capable of adapting termination criteria to the optimisation dynamics of different learning algorithms. Such approaches may incorporate gradient variance, convergence stability, or model-specific learning characteristics as additional scheduling signals, thereby extending the current resource-aware scheduler while preserving compatibility with existing federated optimisation methods [16].

Another promising direction involves integrating model compression techniques, including structured pruning, quantisation, and knowledge distillation, directly into the federated learning workflow. Combining communication-round reduction with model compression would introduce complementary mechanisms for reducing both computational and communication costs. EcoFL provides a suitable benchmarking environment for systematically evaluating the interactions between compression strategies, convergence behaviour, and resource utilisation under heterogeneous edge-resource constraints.

Finally, extending EcoFL to non-stationary and adversarial federated learning environments represents an important step towards realistic deployment scenarios. Future studies should investigate scheduler behaviour under concept drift, intermittent client availability, and Byzantine attacks while incorporating robust aggregation algorithms and drift-aware scheduling mechanisms. Such extensions would broaden the applicability of EcoFL from controlled experimental benchmarking towards more realistic edge AI environments characterised by dynamic workloads and heterogeneous device behaviour.

More broadly, EcoFL establishes a reproducible experimental foundation for investigating the interplay between predictive performance, computational resource utilisation, communication efficiency, and energy consumption in federated learning. We anticipate that extending the framework to additional federated optimisation algorithms, heterogeneous hardware platforms, and real-world IoT datasets will facilitate more comprehensive system-level benchmarking and contribute to the development of practical, energy-aware federated learning systems for next-generation edge intelligence.

6. Material and Methods

6.1. Hardware Profiles

All experiments were conducted on a host machine running Ubuntu 22.04 LTS, with resource constraints enforced through Linux cgroups (via systemd-run) to reproduce the computational characteristics of two representative edge device classes. The adopted execution environment should therefore be interpreted as a reproducible resource-constrained experimental platform rather than a cycle-accurate hardware emulation of the target devices. This approach enables systematic comparison of federated learning configurations under identical and controlled computational conditions while ensuring full experimental reproducibility.

Scope of the experimental evaluation. All quantitative results reported in this study (Tables 3–5 and Figures 2–7) were obtained exclusively using the Raspberry Pi 4 (RPi4) hardware profile. Although the Jetson Nano profile is fully implemented within the EcoFL framework and its cgroup configuration is publicly available in the open-source repository, a comprehensive comparative evaluation across multiple hardware profiles is beyond the scope of the present study and is therefore deferred to future work (Section 5.5). The Raspberry Pi 4 profile was selected as the primary benchmark target because it represents the more resource-constrained of the two platforms, thereby providing a conservative evaluation scenario for analysing the computational and energy characteristics of EcoFL.

The hardware profiles used throughout the experiments are defined as follows:

1. Raspberry Pi 4 (RPi4): 1 CPU core (25% quota), 1 GB RAM, TDP = 6.4 W.
2. Jetson Nano (JN): 2 CPU cores, 4 GB RAM, TDP = 10.0 W.

6.2. Dataset

We evaluate EcoFL on an IoT sensor anomaly detection task using a synthetic multi-variate telemetry dataset comprising 50,000 samples described by 12 features, including temperature, humidity, vibration, current, voltage, pressure, light intensity, CO₂ concentration, and four derived statistical features. An anomaly rate of 5% is introduced through Gaussian distribution shifts and contextual spike injection, while 8% label noise is incorporated to simulate sensor mislabelling and imperfect annotation commonly encountered in real-world IoT deployments.

The dataset is partitioned across 10 federated clients using a Dirichlet distribution with $\alpha = 0.5$, resulting in a moderately non-IID data distribution representative of heterogeneous edge environments.

A synthetic dataset was intentionally selected to provide full experimental control over anomaly prevalence, label noise, feature distributions, and client heterogeneity, thereby ensuring reproducible comparison across all evaluated federated learning configurations. This controlled setting enables systematic assessment of the resource–performance trade-offs introduced by the proposed benchmarking framework while avoiding confounding effects arising from dataset-specific characteristics. Validation on real-world IoT datasets constitutes an important direction for future work and will further assess the generalisability of the proposed benchmarking methodology.

6.3. Models and Configurations

Five lightweight model families are evaluated to span a representative range of ML paradigms relevant to IoT edge deployment. Logistic Regression (LR) serves as a linear baseline whose parameter-averaging compatibility with FedAvg makes it a natural reference point. Random Forest (RF) and XGBoost (XGB) represent tree-based ensemble approaches that require prediction ensemble aggregation in the federated setting, as their non-differentiable structure precludes direct parameter averaging. The Multilayer Perceptron (MLP) with two hidden layers (64, 32 units) represents gradient-based neural architectures compatible with parameter averaging, enabling assessment of client drift under non-IID conditions. Isolation Forest (IF) is included as an unsupervised anomaly detection baseline. All models are implemented using scikit-learn [22] and xgboost [23] with the hyperparameters listed in Table 5. Each model is evaluated under three configurations: centralised (single-node upper bound), standard FedAvg (fixed 20 rounds), and EcoFL (energy-aware adaptive termination).

Table 5. Lightweight model families evaluated in EcoFL.

Model	Library	Key Hyperparameters	FL Strategy	Role
Logistic Regression	scikit-learn	$C = 1.0$, lbfgs, max_iter = 1000	Parameter avg.	Linear baseline
Random Forest	scikit-learn	n_est = 50, max_depth = 8	Prediction ensemble	Ensemble
XGBoost	xgboost	n_est = 50, depth = 4, lr = 0.1	Prediction ensemble	Boosted trees
MLP	scikit-learn	(64, 32), Adam, max_iter = 50	Parameter avg.	Neural network
Isolation Forest	scikit-learn	n_est = 50, contam. = 0.05	Prediction ensemble	Anomaly (unsup.)

6.4. Evaluation Metrics

EcoFL employs a two-tier evaluation protocol in which predictive performance and system-level resource utilisation are treated as complementary and equally important evaluation dimensions. This dual-perspective methodology reflects the central objective of EcoFL: enabling systematic benchmarking of federated learning under resource-constrained edge environments, where computational efficiency is as critical as predictive performance. Accordingly, machine learning metrics quantify the predictive capability of

the trained models on the held-out test set, whereas system-level metrics characterise the computational resources required to achieve that performance under the defined experimental hardware profile. Treating these metric groups jointly enables a comprehensive assessment of the trade-offs between model quality, computational cost, and communication efficiency, which forms the basis of the Pareto analysis presented in Section 4.5.

The predictive performance metrics include Accuracy, macro F1-score, ROC-AUC, Precision, and Recall. System-level benchmarking metrics comprise total computational energy consumption (mJ), energy per communication round (mJ/round), peak memory utilisation (MB), mean and P95 inference latency (ms), and total communication overhead (KB).

Unless otherwise stated, all reported results are presented as mean $\pm$ standard deviation across five independent random seeds (42–46).

6.5. Statistical Analysis

Formal statistical comparison between EcoFL and FedAvg is performed using the Wilcoxon signed-rank test (two-sided for F1-score, one-sided for energy) with Holm–Bonferroni correction for multiple comparisons across five model families. With $n = 5$ seeds, the test has limited statistical power—the minimum detectable effect at $\alpha = 0.05$ under this sample size is approximately $|\Delta| > 0.15$ F1, exceeding the actual differences observed for LR, XGB, and IF (which are identically zero). All Wilcoxon p -values are non-significant after correction ($p \geq 0.156$), which is consistent with the absence of detectable F1 degradation under the tested seed regime for these models: EcoFL and FedAvg produce numerically identical F1 because rapid convergence within 4 rounds leaves no additional gradient signal for subsequent rounds. The non-significant result supports the practical interpretation that EcoFL did not introduce detectable F1 degradation under this experimental setting. Bootstrap 95% confidence intervals (2000 resamples) further support this interpretation. For MLP, the observed $\Delta = -0.035$ F1 is directionally consistent across all seeds but does not reach significance at $n = 5$. Additionally, a one-sided Wilcoxon test confirms that EcoFL energy consumption is directionally lower than FedAvg for all five models, with savings of 79.7–92.9% ($p = 0.156$ after correction; the test has low power at $n = 5$ due to low variance in round counts across seeds).

7. Conclusions

This study presented EcoFL (Energy-Conscious Federated Learning), an open-source, modular framework for reproducible system-level benchmarking and resource-aware orchestration of federated learning under resource-constrained edge environments. Rather than introducing a new federated optimisation algorithm, EcoFL provides a unified experimental platform for jointly evaluating predictive performance, computational resource utilisation, communication overhead, and estimated computational energy consumption across different lightweight machine learning models and federated learning configurations.

Experimental evaluation was performed using five lightweight model families, three learning configurations, five independent random seeds, and a 50,000-sample synthetic IoT anomaly detection benchmark executed under reproducible resource-constrained hardware profiles. The principal findings of this study can be summarised as follows:

1. Substantial improvements in computational efficiency: EcoFL reduced estimated computational energy consumption by 79.9–92.9% relative to the standard FedAvg configuration, corresponding to a mean reduction of 84.4%, while simultaneously reducing communication overhead by approximately 80% through adaptive communication-round termination.

2. Predictive performance was largely preserved: For Logistic Regression, XGBoost, and Isolation Forest, EcoFL maintained predictive performance comparable to FedAvg while requiring substantially fewer communication rounds. Random Forest achieved a higher macro F1-score (+0.052) under EcoFL, whereas the Multilayer Perceptron exhibited a modest decrease in predictive performance, highlighting the influence of model-specific convergence dynamics on adaptive scheduling.
3. System-level benchmarking reveals architecture-dependent trade-offs: The experimental results demonstrate that the effectiveness of adaptive communication scheduling depends on the optimisation characteristics of the underlying learning model. Lightweight models exhibiting rapid convergence benefit most from early communication-round termination, whereas gradient-based neural architectures may require more conservative or adaptive scheduling strategies.
4. Integrated evaluation supports informed deployment decisions: By jointly analysing predictive performance, computational energy consumption, memory utilisation, inference latency, and communication overhead, EcoFL enables systematic identification of operating points that balance model quality and resource efficiency according to application-specific deployment requirements.

The complete EcoFL framework is released as open-source software to support reproducibility, transparent evaluation, and future methodological extensions by the research community.

More broadly, this work demonstrates that system-level benchmarking should complement traditional accuracy-centred evaluation when developing federated learning systems for resource-constrained edge environments. Considering predictive performance together with computational energy, communication overhead, memory utilisation, and latency provides a more comprehensive basis for evaluating practical federated learning deployments. We anticipate that EcoFL will serve as a reproducible research platform for investigating resource-aware federated learning and for supporting the development of future energy-efficient edge AI systems.

Author Contributions: Conceptualization, T.M.; methodology, T.M.; software, T.M.; validation, T.M. and I.D.; formal analysis, T.M.; investigation, T.M. and I.D.; resources, T.M.; data curation, T.M.; writing—original draft preparation, T.M. and I.D.; writing—review and editing, T.M. and I.D.; visualization, T.M.; supervision, T.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable. This study did not involve human participants, animals, or any data requiring ethical approval. All datasets used are synthetically generated by the authors.

Informed Consent Statement: Not applicable. No human subjects were involved in this research.

Data Availability Statement: The EcoFL source code, data generation scripts, experimental configuration files (seeds, cgroup profiles, model hyperparameters), and full result logs are openly available on GitHub at <https://github.com/TyMill/EcoFL> (release v1.0.0) (accessed on 25 May 2026) and archived on Zenodo at <https://doi.org/10.5281/zenodo.20314541> (accessed on 25 May 2026). The synthetic IoT sensor telemetry dataset used in this study is generated programmatically by the script `generate_data.py` included in the repository; no external datasets or sensitive personal data are used. All experiments are fully reproducible by running `run_experiments.sh` with the dependency environment specified in `requirements.txt` (Python 3.10, scikit-learn 1.4, xgboost 2.0, psutil 5.9) using seeds 42–46.

Acknowledgments: The authors thank the open-source communities behind scikit-learn, XGBoost, and psutil, whose libraries form the computational backbone of EcoFL.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript: EcoFL: Energy-Conscious Federated Learning; FL: Federated Learning; IoT: Internet of Things; IID: Independent and Identically Distributed; non-IID: Non-Independent and Identically Distributed; FedAvg: Federated Averaging; TDP: Thermal Design Power; MLP: Multilayer Perceptron; RF: Random Forest; XGB: XGBoost; LR: Logistic Regression; IF: Isolation Forest; RPi4: Raspberry Pi 4; JN: Jetson Nano; TinyML: Tiny Machine Learning; RAM: Random Access Memory; CPU: Central Processing Unit; NIC: Network Interface Controller; DPI: Dots Per Inch; TOST: Two One-Sided Tests; AUC: Area Under the Curve; ROC: Receiver Operating Characteristic.

References

1. Yang, Q.; Liu, Y.; Chen, T.; Tong, Y. Federated machine learning: Concept and applications. *ACM Trans. Intell. Syst. Technol.* **2019**, *10*, 12. <https://doi.org/10.1145/3298981>.
2. McMahan, H.B.; Moore, E.; Ramage, D.; Hampson, S.; Agüera y Arcas, B. Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), Fort Lauderdale, FL, USA, 20–22 April 2017*; PMLR: Brookline, MA, USA, 2017; pp. 1273–1282.
3. Kairouz, P.; McMahan, H.B.; Avent, B.; Bellet, A.; Bennis, M.; Bhagoji, A.N.; Bonawitz, K.; Charles, Z.; Cormode, G.; Cummings, R.; et al. Advances and open problems in federated learning. *Found. Trends Mach. Learn.* **2021**, *14*, 1–210.
4. Nguyen, D.C.; Ding, M.; Pathirana, P.N.; Seneviratne, A.; Li, J.; Poor, H.V. Federated learning for Internet of Things: A comprehensive survey. *IEEE Commun. Surv. Tutor.* **2021**, *23*, 1622–1658. <https://doi.org/10.1109/COMST.2021.3075439>.
5. Li, T.; Sahu, A.K.; Talwalkar, A.; Smith, V. Federated learning: Challenges, methods, and future directions. *IEEE Signal Process. Mag.* **2020**, *37*, 50–60. <https://doi.org/10.1109/MSP.2020.2975749>.
6. Zhao, Y.; Li, M.; Lai, L.; Suda, N.; Civin, D.; Chandra, V. Federated learning with non-IID data. *arXiv* **2018**, arXiv:1806.00582.
7. Karimireddy, S.P.; Kale, S.; Mohri, M.; Reddi, S.; Stich, S.; Suresh, A.T. SCAFFOLD: Stochastic controlled averaging for federated learning. In *Proceedings of the 37th International Conference on Machine Learning (ICML), Virtual, 13–18 July 2020*; PMLR: Brookline, MA, USA, 2020; pp. 5132–5143.
8. Reisizadeh, A.; Mokhtari, A.; Hassani, H.; Jadbabaie, A.; Pedarsani, R. FedPAQ: A communication-efficient federated learning method with periodic averaging and quantization. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS), Palermo, Italy, 26–28 August 2020*; PMLR: Brookline, MA, USA, 2020; pp. 2021–2031.
9. Fu, X.; Zhu, J.; Wang, X.; Yang, X.; Yang, X.; Shu, J. FedAtlas: A Privacy-Enhanced Federated Multi-Task Learning Framework Driven by Semantic Communication for Dual Heterogeneity. *IEEE Trans. Serv. Comput.* **2026**, 1–14. <https://doi.org/10.1109/TSC.2026.3685997>.
10. Konečný, J.; McMahan, H.B.; Yu, F.X.; Richtárik, P.; Suresh, A.T.; Bacon, D. Federated learning: Strategies for improving communication efficiency. *arXiv* **2016**, arXiv:1610.05492.
11. Lin, Y.; Han, S.; Mao, H.; Wang, Y.; Dally, W.J. Deep gradient compression: Reducing the communication bandwidth for distributed training. In *Proceedings of the International Conference on Learning Representations (ICLR), Vancouver, BC, Canada, 30 April–3 May 2018*.
12. Luo, B.; Li, X.; Wang, S.; Huang, J.; Tassiulas, L. Cost-effective federated learning design. In *Proceedings of IEEE INFOCOM 2021—IEEE Conference on Computer Communications, Vancouver, BC, Canada, 10–13 May 2021*; IEEE: New York, NY, USA, 2021; pp. 1–10. <https://doi.org/10.1109/INFOCOM42981.2021.9488679>.
13. Wang, H.; Kaplan, Z.; Niu, D.; Li, B. Optimizing federated learning on non-IID data with reinforcement learning. In *Proceedings of the IEEE INFOCOM 2020—IEEE Conference on Computer Communications, Toronto, ON, Canada, 6–9 July 2020*. <https://doi.org/10.1109/INFOCOM41043.2020.9155494>.
14. Caldas, S.; Duddu, S.M.K.; Wu, P.; Li, T.; Konečný, J.; McMahan, H.B.; Smith, V.; Talwalkar, A. LEAF: A benchmark for federated settings. *arXiv* **2018**, arXiv:1812.01097.

15. Reddi, S.J.; Charles, Z.; Zaheer, M.; Garrett, Z.; Rush, K.; Konečný, J.; Kumar, S.; McMahan, H.B. Adaptive federated optimization. In Proceedings of the International Conference on Learning Representations (ICLR), Virtual, 3–7 May 2021. <https://doi.org/10.48550/arXiv.2003.00295>.
16. Bonawitz, K.; Eichner, H.; Grieskamp, W.; Huba, D.; Ingerman, A.; Ivanov, V.; Kiddon, C.; Konečný, J.; Mazzocchi, S.; McMahan, H.B.; et al. Towards federated learning at scale: System design. *Mach. Learn. Syst.* **2019**, *1*, 374–388. <https://doi.org/10.48550/arXiv.1902.01046>.
17. David, R.; Duke, J.; Jain, A.; Janapa Reddi, V.; Jeffries, N.; Li, J.; Kreeger, N.; Nappier, I.; Natraj, M.; Regev, S.; et al. TensorFlow Lite Micro: Embedded machine learning for TinyML systems. *Mach. Learn. Syst.* **2021**, *3*, 800–811. <https://doi.org/10.48550/arXiv.2010.08678>.
18. Banbury, C.; Reddi, V.J.; Torelli, P.; Holleman, J.; Jeffries, N.; Kiraly, C.; Montino, P.; Kanter, D.; Ahmed, S.; Pau, D.; et al. MLPerf Tiny benchmark. In Proceedings of the NeurIPS Datasets and Benchmarks Track, Virtual, 6–14 December 2021. <https://doi.org/10.48550/arXiv.2106.07597>.
19. Warden, P.; Situnayake, D. *TinyML: Machine Learning with TensorFlow Lite on Arduino and Ultra-Low-Power Microcontrollers*; O'Reilly Media: Sebastopol, CA, USA, 2019.
20. Sze, V.; Chen, Y.H.; Yang, T.J.; Emer, J.S. Efficient processing of deep neural networks: A tutorial and survey. *Proc. IEEE* **2017**, *105*, 2295–2329. <https://doi.org/10.48550/arXiv.1703.09039>.
21. Han, S.; Mao, H.; Dally, W.J. Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding. In Proceedings of the International Conference on Learning Representations (ICLR), San Juan, Puerto Rico, 2–4 May 2016.
22. Jacob, B.; Kligys, S.; Chen, B.; Zhu, M.; Tang, M.; Howard, A.; Adam, H.; Kalenichenko, D. Quantization and training of neural networks for efficient integer-arithmetic-only inference. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018*; IEEE: New York, NY, USA, 2018; pp. 2704–2713.
23. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
24. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), San Francisco, CA, USA, 13–17 August 2016*; ACM: New York, NY, USA, 2016; pp. 785–794.
25. Lottick, K.; Susai, S.; Friedler, S.A.; Wilson, J.P. Energy usage reports: Environmental awareness as part of algorithmic accountability. In Proceedings of the NeurIPS Workshop on Tackling Climate Change with Machine Learning, Vancouver, BC, Canada, 13 December 2019.
26. Henderson, P.; Hu, J.; Romoff, J.; Brunskill, E.; Jurafsky, D.; Pineau, J. Towards the systematic reporting of the energy and carbon footprints of machine learning. *J. Mach. Learn. Res.* **2020**, *21*, 1–43.
27. He, C.; Li, S.; So, J.; Zhang, M.; Wang, H.; Wang, X.; Vepakomma, P.; Singh, A.; Qiu, H.; Shen, L.; et al. FedML: A research library and benchmark for federated machine learning. *arXiv* **2020**, arXiv:2007.13518. <https://doi.org/10.48550/arXiv.2007.13518>.
28. Beutel, D.J.; Topal, T.; Mathur, A.; Qiu, X.; Fernandez-Marques, J.; Gao, Y.; Sani, L.; Li, K.H.; Parcollet, T.; de Gusmão, P.P.B.; et al. Flower: A friendly federated learning research framework. *arXiv* **2022**, arXiv:2007.14390. <https://doi.org/10.48550/arXiv.2007.14390>.
29. Yang, Z.; Chen, M.; Saad, W.; Hong, C.S. Energy efficient federated learning over wireless communication networks. *IEEE Trans. Wirel. Commun.* **2021**, *20*, 5090–5104. <https://doi.org/10.1109/twc.2020.3037554>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.