



Noise-aware Client Selection for carbon-efficient Federated Learning via Gradient Norm Thresholding

PATRICK WILHELM, BIFOLD and Technische Universität Berlin

INESE YILMAZ, Technische Universität Berlin

ODEJ KAO, Technische Universität Berlin

Training large-scale Neural Networks requires substantial computational power and energy. Federated Learning enables distributed model training across geospatially distributed data centers, leveraging renewable energy sources to reduce the carbon footprint of AI training. Various client selection strategies have been developed to align the volatility of renewable energy with stable and fair model training in a federated system. However, due to the privacy-preserving nature of Federated Learning, the quality of data on client devices remains unknown, posing challenges for effective model training. In this paper, we introduce a modular approach on top to state-of-the-art client selection strategies for carbon-efficient Federated Learning. Our method enhances robustness by incorporating a noisy client data filtering, improving both model performance and sustainability in scenarios with unknown data quality. Additionally, we explore the impact of carbon budgets on model convergence, balancing efficiency and sustainability. Through extensive evaluations, we demonstrate that modern client selection strategies based on local client loss tend to select clients with noisy data, ultimately degrading model performance. To address this, we propose a gradient norm thresholding mechanism using probing rounds for more effective client selection and noise detection, contributing to the practical deployment of carbon-efficient Federated Learning.

CCS Concepts: • **Computing methodologies** → *Artificial intelligence*; • **Hardware** → *Impact on the environment*.

Additional Key Words and Phrases: Sustainable AI, Federated Learning, Carbon-aware Training

1 Introduction

The rapid advancement of deep learning has led to models with hundreds of billions of parameters, consistently outperforming previous AI benchmarks. However, this progress comes at a significant cost: larger models demand more computational power, resulting in hours of GPU training time, increased energy consumption, and carbon emissions of up to 24.7 tonnes of CO₂ for training a model with approximately 170 billion parameters [17]. Recent studies project that the energy demand of global data centers will reach 1,000 TWh by 2026, driven in large part by the surge in AI inference and training workloads. Furthermore, carbon emissions from these activities are estimated to contribute up to 8% of global emissions within the next decade [9].

One of the most promising strategies to mitigate the carbon footprint of neural network training is to align computational workloads with the availability of renewable energy sources [20, 21]. Distributed training approaches, particularly those leveraging geospatially distributed data centers, offer a unique opportunity to utilize local renewable energy surpluses. Federated Learning (FL), in particular, enables decentralized training across data centers or edge

devices, making it a suitable candidate for carbon-aware, geospatial model training [2, 25].

To exploit renewable availability, client selection strategies in FL have been developed to dynamically allocate training workloads based on the local computing capacity and renewable energy availability of clients [3]. However, due to the privacy-preserving nature of FL, the quality of local data remains unknown. This presents a challenge: current selection strategies often rely on client-side training loss to determine utility, yet high loss may stem from either valuable, hard examples—or noisy and corrupted data [15]. Consequently, selecting clients solely based on high loss can inadvertently introduce harmful noise into the collaborative training process.

Achieving a balance between sustainability and model performance requires more nuanced approaches that can infer the impact of a client's data without breaching privacy. We propose a noise-aware client selection mechanism that leverages gradient norm statistics as a proxy for data quality. Inspired by the Critical Learning Periods [1, 28], we show that estimating the Fisher Information Matrix via gradient norm during probing rounds allows for identifying clients whose data positively contribute to model convergence—while filtering those with potentially corrupted or noisy samples.

In our experiments, we demonstrate the following key findings:

- Enhancing traditional loss-based client selection with a single gradient norm thresholding round improves resilience against noisy or corrupted client data.
- Integrating carbon budgets into the client selection process enables a more balanced trade-off between model performance and sustainability in carbon-aware FL.

Our findings are supported by a series of experiments, presented throughout the remainder of this paper. In Section 2, we discuss related work on carbon-aware FL, noise-robust training and data valuation. Section 3 outlines our system design, including carbon data integration and our proposed client selection approach. In Section 4, we evaluate the effectiveness of gradient norm thresholding for filtering noisy clients. Finally, in Section 5 and 6, we summarize our contributions and provide an outlook on incorporating data valuation techniques into privacy-preserving FL.

2 Related Work and Background

Previous studies align client selection strategies with the high volatility of renewable energy within the power grid [2, 25]. Recent studies have investigated various strategies for client selection, including choosing clients according to the size of their local datasets [18], or favoring clients that exhibit higher training losses [4, 15]. None of those take data quality into account.

Authors' Contact Information: Patrick Wilhelm, patrick.wilhelm@tu-berlin.de, BIFOLD and Technische Universität Berlin; Inese Yilmaz, inese.yilmaz@campus.tu-berlin.de, Technische Universität Berlin; Odej Kao, odej.kao@tu-berlin.de, Technische Universität Berlin.

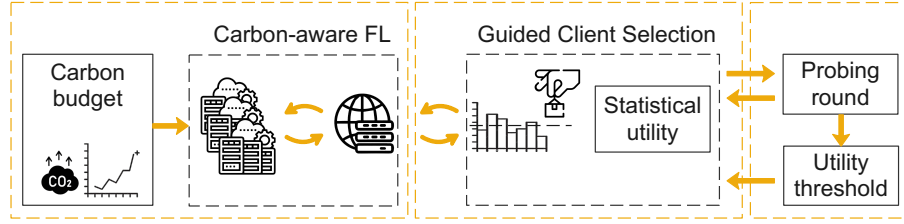


Fig. 1. System Overview: Clients represent datacenters in 30 regions within the US. While the data quality is unknown we introduce probing rounds in the beginning of federated training to identify clients with noisy data. Integrating carbon budgets we can balance clients with high data utility and their current carbon intensity.

Noisy data can significantly disrupt the training process by producing high loss and erratic gradients, leading to unstable parameter updates and reduced learning efficiency [6]. While examples with slightly higher losses are valuable for guiding meaningful learning, excessively large training losses often indicate noisy or corrupted data that should be avoided [6, 8, 16, 22].

Critical learning periods in neural networks are early phases of training during which the model is most plastic and responsive, forming lasting internal representations that significantly influence its final performance [1, 7]. During this phase, distinguishing between informative, high-impact examples and harmful noisy data, enables more effective and stable training [12]. Several works approximate the Hessian as an indicator for critical learning periods using Fisher Information Matrix [11, 27], which by itself can be approximated via gradient norm for more efficient computation [28].

3 System Design and Approach

In Federated Learning (FL), data is typically non-IID across clients, leading to significant variations in data quality and utility. Some clients may hold high-quality, diverse, and representative datasets that support convergence and generalization, while others may hold noisy, biased, or redundant data that hinders training and degrades overall performance.

In our system, we extend client selection strategies through an explicit probing round at the beginning and consecutive filtering via a coefficient threshold to maintain model quality and avoid selecting harmful data.

While aligning computation with renewable energy availability promotes sustainability, it restricts the client pool and introduces variability in participation. This often leads to reliance on suboptimal clients, negatively affecting training efficiency and model accuracy. While the use of a limited carbon budget increases client availability, it is essential to prioritize clients not only based on energy availability but also on their expected contribution to model performance.

We enable the selection of higher-emission clients through the use of a fixed carbon budget and improve budget efficiency via informed client selection. To achieve this, we model the trade-off between clients' statistical utility and carbon intensity by adapting Oort's reward calculation mechanism [15]. Figure 1 visualizes our noise-aware client selection extension for carbon-aware FL.

3.1 Client Selection through Gradient Norm Probing

To enable robust client selection under real-world data noise, we begin by evaluating all clients during an initial probing round. Clients are assessed using the statistical utility formulation proposed in Oort [15], but instead of using the loss approximation, we compute utility based on the gradient norm once at the beginning: $U(i) = |B_i| \cdot \sqrt{\frac{1}{|B_i|} \sum_{k \in B_i} \|\nabla f(k)\|^2}$, where $\nabla f(k)$ is the gradient of the loss function with respect to sample k , and $\|\nabla f(k)\|$ denotes its L2 norm. U_i is the utility and B_i the local data of client i .

This utility, further referred to as probing utility, reflects the curvature of the local loss landscape, capturing data informativeness and noise sensitivity more effectively than training loss. While more computationally intensive, it offers greater robustness in federated settings.

The server aggregates client probing utilities and applies a utility-variance-based threshold. Specifically, a client is retained if its probing utility satisfies $utility \geq c \cdot \max(utility)$, where $c \in [0, 1]$ is a configurable coefficient that controls the exclusion level, and its effect is evaluated in 4.1. This ensures that only the highest-utility clients are retained, while those falling below the threshold are excluded from subsequent training rounds.

This approach is not limited to the gradient norm-based utility examined in our study; other metrics can be used in its place or combined to evaluate clients during the probing round. Different thresholding strategies can then be applied to enable more robust client selection. Moreover, the overall method is compatible with other client selection approaches.

3.2 Utility Aware Carbon Budget Allocation

To address the limitations of energy-aligned training—such as reduced client availability and reliance on suboptimal participants—we explore a client selection method that accounts for a fixed carbon budget and allocates resources based on client utility. Our method builds on Oort's client selection strategy [15], enhancing it with carbon-awareness to improve training efficiency without exceeding emissions constraints.

We replace Oort's exploration mechanism with the probing round introduced earlier, during which we can also compute statistical utility as proposed in Oort[15] for all clients. Rather than selecting only the top-scoring clients based on utility alone, we formulate a budget-aware optimization problem. In each round, the server selects a subset of clients that maximizes the total utility score while

ensuring that the round's total emissions remain within the carbon budget:

$$\max_{S \subseteq C} \sum_{i \in S} r_i \quad \text{s.t.} \quad \sum_{i \in S} c_i \leq B_t, \quad |S| \leq K$$

Here, C is the set of all clients, r_i is the utility score of client i (obtained from the probing round), c_i is the carbon intensity of client i in round t , B_t is the carbon budget allocated for round t , and K is the number of clients selected per round.

By explicitly modeling the trade-off between utility and emissions, this strategy improves budget efficiency and mitigates the performance degradation often caused by over-reliance on low-emission clients.

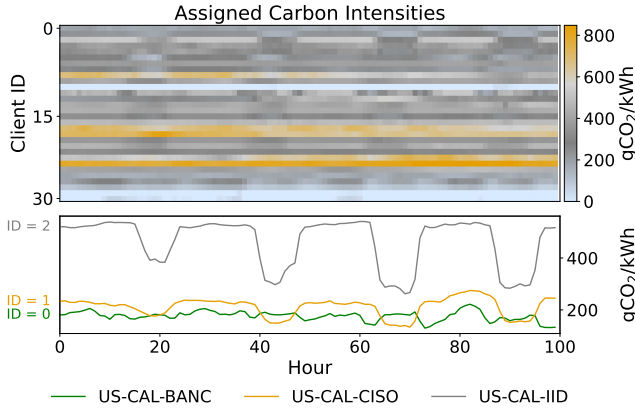


Fig. 2. Assigned carbon intensity values for clients over 100 training rounds. Each round corresponds to one hour, spanning from 2023-01-15 00:00:00 to 2023-01-19 04:00:00 (UTC). The bottom plot highlights daytime intensity patterns for three clients from different regions.

4 Experiments

We evaluate our proposed strategies under two practical considerations: data quality variation and the temporal-spatial variability of carbon intensity across energy regions.

We perform experiments using CIFAR-10 [13] with data distributed across clients non-IID under Dirichlet ($\alpha = 10$) on 30 clients. A simple CNN (2 convolutional and 3 fully connected layers) is trained with 10 clients per round, 2 local epochs, batch size 32, using Adam optimizer (learning rate = 0.001).

To assess robustness to data noise, specifically inaccuracies or inconsistencies in input features that disrupt the relationship between features and their corresponding labels, we simulate feature noise by adding zero-mean Gaussian noise with standard deviation $\sigma = 1$ to each image, followed by restricting pixel values to the valid range $[0, 1]$. This perturbs image details while slightly preserving the overall structure. We replace the data of 6 clients with corrupted, noisy variants.

To model and assess how carbon intensity data can be utilized in combination with statistical utility, each of the 30 clients is assigned to a U.S. energy region, using hourly carbon intensity historical data from Electricity Maps [5], as illustrated in Fig. 2. For simplicity, we

assume that all clients have equal power demand, with each client consuming 1kWh during training, and each training round lasting one hour.

For carbon budgeting, energy curtailment data [5] is combined with carbon intensity data: if curtailment is available, the client is assumed to operate on curtailed (zero-emission) energy and assigned an intensity of 0; otherwise, the previously recorded intensity value is used.

4.1 Handling Noisy Clients via Probing-Based Client Evaluation

We evaluate gradient norm thresholding, described in Section 3.1, on a data distribution that includes six corrupted clients, by applying it to two standard baselines: Random and Oort [15]. In this configuration, we use only Oort's statistical utility for its reward mechanism, excluding system utility to focus specifically on data quality. The resulting thresholded variants are referred to as RandomWT and OortWT, respectively.

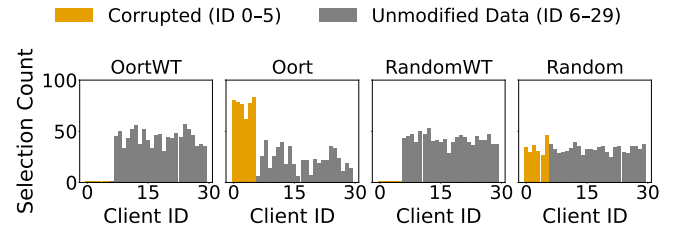


Fig. 3. Client selection counts, 6 out of 30 clients contain noisy data.

Without thresholding, corrupted clients are frequently selected for training. As shown in Figure 3, Oort consistently prioritizes these clients due to its utility metric, which is based on local training loss—typically increased for noisy data. Although Oort includes an optional blacklisting mechanism that allows the exclusion of clients after participating in a specified number of rounds, it continues to favor corrupted clients prior to that point. In contrast, gradient norm thresholding enables the identification of noisy clients and their exclusion of subsequent training rounds.

Selecting an appropriate threshold is key to filtering out noisy or harmful clients while retaining useful ones. In our setup, a value

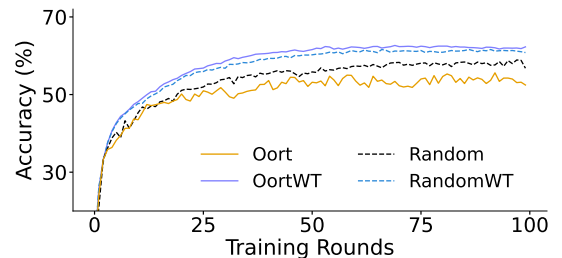


Fig. 4. Convergence curves under noisy data scenario. Baseline methods are compared with their thresholded variants. Thresholding reduces the impact of corrupted clients, resulting in faster and more stable convergence.

of $c = 0.5$ provides a good balance, in practice, this would require evaluation on a small, representative subset of the data.

The results in Figure 4 demonstrate that the overselection of noisy clients significantly degrades Oort’s performance, leading to reduced accuracy. While Random selection does not systematically favor noisy clients, it still includes them with high probability, which disrupts convergence. In contrast, the thresholded variants, OortWT and RandomWT, result in faster and more stable convergence and achieve higher final accuracy. The improvement is particularly beneficial for methods like Oort, which tend to prioritize high-loss clients—enhancing robustness in the presence of noisy data.

While the average carbon intensity of corrupted clients is lower than that of others (see client IDs 0–5 in Fig.2), their selection leads to increased carbon emissions while achieving lower accuracy. Although the probing round introduces a one-time computational cost, it allows reaching maximum accuracy earlier, allowing for a reduction in training rounds, as demonstrated in Figure 5, which shows emissions measured at the point of maximum accuracy.

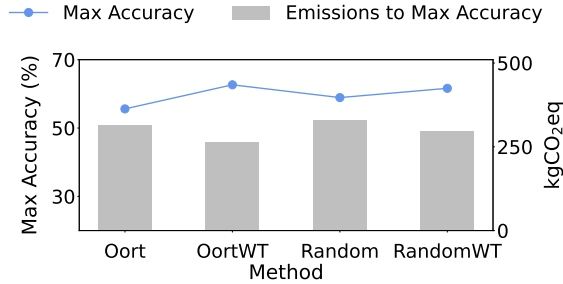


Fig. 5. Accuracy–emissions comparison across methods. Bars show the carbon emissions spent during training up to the point of maximum accuracy, while the line shows the corresponding maximum accuracy.

4.2 Evaluating Budget-Constrained Utility-Aware Client Selection

We evaluate the utility-aware carbon budget allocation strategy introduced in Section 3.2, here referred to as OortCA. We compare it to the original Oort strategy, which serves as the baseline with unconstrained emissions usage—meaning *all clients are available* for selection regardless of their carbon intensity. The total emissions from Oort serve as a reference point for setting carbon budgets and define the emission baseline.

Carbon budget is set for the entire training and divided evenly across rounds, with unused carbon carried over. If the budget is spent before reaching the desired number of clients in a round, as the remaining are selected curtailment clients with the highest utility scores.

We test OortCA across different levels of carbon budget relative to this emission baseline. Evaluation proceeds as follows: we begin with a zero-carbon budget, which restricts selection to clients using only curtailed (i.e., zero-emission) energy, significantly limiting the available client pool. We then increase the carbon budget in 10% increments, gradually expanding the set of eligible clients for

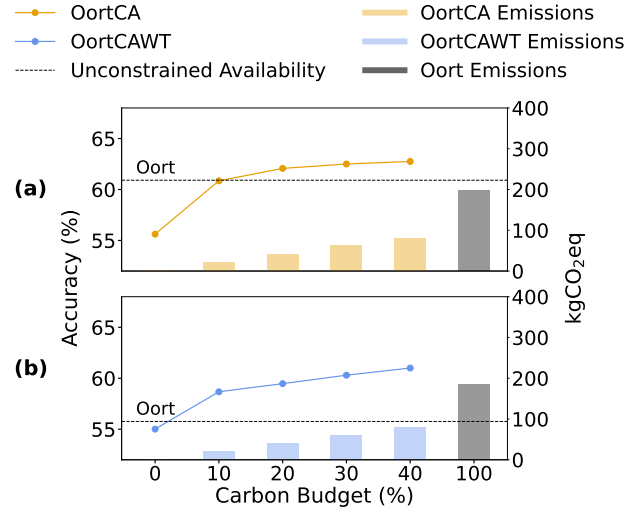


Fig. 6. Impact of carbon budgeting on CO₂-emissions and model performance, compared to standard Oort client selection with full client availability. Results are shown for two scenarios: (a) clients with clean data, and (b) clients with corrupted data.

selection and allowing higher-emission clients to be included, as shown in Figure 7.

As shown in Figure 6, by effectively balancing client utility despite reduced availability due to budget constraints, OortCA achieves final accuracy comparable to the unconstrained Oort baseline — while using only 40% of its emissions.

We further repeat the experiment under the noisy data setup, combining the utility-aware budget allocation (Section 3.2) with the gradient norm thresholding approach (Section 3.1). The resulting method, referred to as OortCAWT, enables more robust client selection in the presence of corrupted data while maximizing the use of limited carbon budgets and spending the budget on clean, high-utility clients.

Figure 8 presents extended results using the DenseNet-121 [10] model evaluated on the CIFAR-100 dataset and experiments with the EfficientNet-B1 [23] evaluated on the Tiny ImageNet. The Tiny ImageNet contains 100,000 64×64 color images of 200 classes. CIFAR-100 [14] contains 60,000 32×32 color images across 100 classes.

The results suggest that in a clean data scenario, the client selection strategy manages to achieve similar model performance even with a reduced/constrained client availability. While in under noisy data conditions model performance varies drastically. In this case, filtering out noisy data via gradient norm improves robustness. Moreover, when combined with carbon-aware training (e.g., budgeted carbon emissions), this approach enables accuracy improvements while accounting for reduced carbon emissions. This opens the possibility of strategically allocating carbon budgets to achieve accuracy gains in noisy data scenarios.

5 Future Work

In this section, we outline potential future directions informed by the experiments and insights gained from this work.

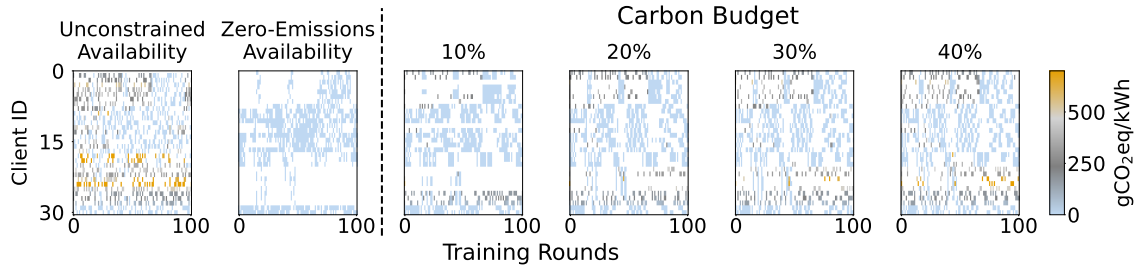


Fig. 7. Client selection during training: Comparison between two approaches—one with unconstrained emissions allowing full client participation in each round, and another enforcing a hard constraint where only zero-emission clients are selected. A third, more flexible strategy, balancing carbon budgeting with broader client availability during training.

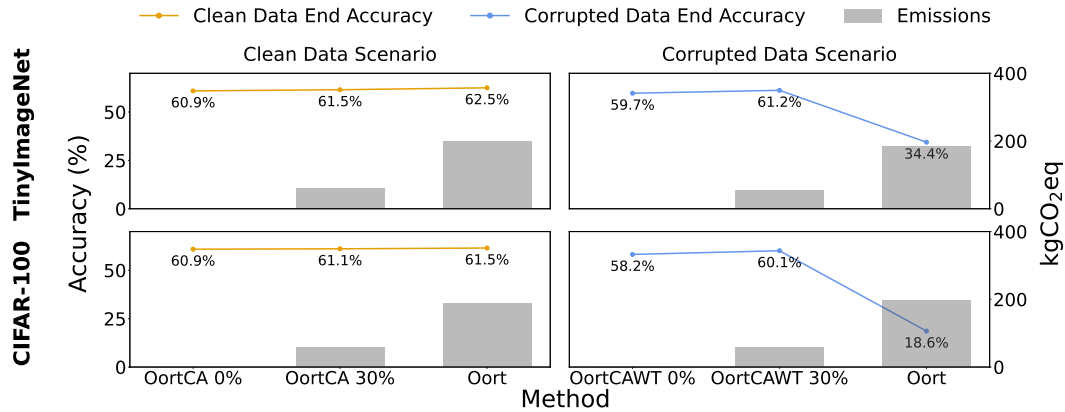


Fig. 8. Impact of carbon budgeting on CO₂-emissions and model performance using the DenseNet-121 model evaluated on the CIFAR-100 dataset and the EfficientNet-B1 model evaluated on the Tiny ImageNet dataset. Results are shown for two scenarios: clients with clean data and corrupted data, where noise was added to 6 out of 30 clients.

Data Valuation in Carbon-Aware Federated Learning FL enables privacy-preserving distributed training, offering a platform to explore data valuation methods for sustainable AI. Techniques such as Federated Shapley Values and Gradient Alignment should be evaluated in carbon-aware settings to balance fairness, bias mitigation, and carbon efficiency [24, 26]. Asynchronous Federated Learning (AFL) is a promising alternative, removing reliance on potentially inaccurate and costly carbon intensity forecasts. Coupling AFL with data valuation and carbon budgeting may yield practical, carbon-efficient training strategies. Another direction involves leveraging critical learning periods—stages where models are most sensitive to high-quality data. Aligning these phases with moderately carbon-intensive, high-value data could improve both convergence and sustainability [1]. To reduce computation during probing rounds, data coresets can be estimated via a single round of local inference [19]. This lowers the cost of gradient norm calculations and can aid in efficiently onboarding new clients.

6 Conclusion

In this paper, we investigated the impact of noisy data on client selection strategies and their implications for carbon-aware computing. We also explored the influence of carbon budgeting for larger

client availability and its effect on model performance. Our experiments demonstrate that gradient norm thresholding is effective for filtering noisy clients, and that strategic use of carbon budgets can mitigate the effects of sparse client availability during periods of renewable energy volatility. We hope this work not only advances understanding in this area but also encourages further research into practical and efficient approaches to carbon-aware machine learning.

References

- [1] Alessandro Achille, Matteo Rovere, and Stefano Soatto. 2017. Critical learning periods in deep neural networks. *arXiv preprint arXiv:1711.08856* (2017).
- [2] Jieming Bian, Lei Wang, Shaolei Ren, and Jie Xu. 2024. CAFE: Carbon-Aware Federated Learning in Geographically Distributed Data Centers. In *Proceedings of the 15th ACM International Conference on Future and Sustainable Energy Systems*. 347–360.
- [3] Roozbeh Bostandoost, Walid A. Hanafy, Adam Lechowicz, Noman Bashir, Prashant Shenoy, and Mohammad Hajiesmaili. 2025. Data-driven Algorithm Selection for Carbon-Aware Scheduling. *SIGENERGY Energy Inform. Rev.* 4, 5 (April 2025), 148–153. doi:10.1145/3727200.3727222
- [4] Yae Jee Cho, Jianyu Wang, and Gauri Joshi. 2020. Client selection in federated learning: Convergence analysis and power-of-choice selection strategies. *arXiv preprint arXiv:2010.01243* (2020).
- [5] Electricity Maps. 2024. Electricity Maps Data Portal – Carbon Intensity Data. <https://www.electricitymaps.com/data-portal> Accessed: 2024-10-17.
- [6] Minghong Fang, Xiaoyu Cao, Jinyuan Jia, and Neil Gong. 2020. Local Model Poisoning Attacks to Byzantine-Robust Federated Learning. In *29th USENIX*

Security Symposium (USENIX Security 20). USENIX Association, 1605–1622. <https://www.usenix.org/conference/usenixsecurity20/presentation/fang>

- [7] Aditya Sharad Golatkar, Alessandro Achille, and Stefano Soatto. 2019. Time matters in regularizing deep networks: Weight decay and data augmentation affect early learning dynamics, matter little near convergence. *Advances in Neural Information Processing Systems* 32 (2019).
- [8] Shivani Gupta and Atul Gupta. 2019. Dealing with Noise Problem in Machine Learning Data-sets: A Systematic Review. *Procedia Computer Science* 161 (2019), 466–474. doi:10.1016/j.procs.2019.11.146 The Fifth Information Systems International Conference, 23–24 July 2019, Surabaya, Indonesia.
- [9] Udit Gupta, Mariam Elgamal, Gage Hills, Gu-Yeon Wei, Hsien-Hsin S. Lee, David Brooks, and Carole-Jean Wu. 2022. ACT: designing sustainable computer systems with an architectural carbon modeling tool. In *Proceedings of the 49th Annual International Symposium on Computer Architecture (New York, New York) (ISCA '22)*. Association for Computing Machinery, New York, NY, USA, 784–799. doi:10.1145/3470496.3527408
- [10] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4700–4708.
- [11] Stanislaw Jastrzebski, Zachary Kenton, Nicolas Ballas, Asja Fischer, Yoshua Bengio, and Amos Storkey. 2018. On the relation between the sharpest directions of DNN loss and the SGD step length. *arXiv preprint arXiv:1807.05031* (2018).
- [12] Michael Kleinman, Alessandro Achille, and Stefano Soatto. 2023. Critical learning periods emerge even in deep linear networks. *arXiv preprint arXiv:2308.12221* (2023).
- [13] Alex Krizhevsky. 2009. *Learning multiple layers of features from tiny images*. Technical Report. University of Toronto. <https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf>
- [14] Alex Krizhevsky, Geoffrey Hinton, et al. 2009. Learning multiple layers of features from tiny images. (2009).
- [15] Fan Lai, Xiangfeng Zhu, Harsha V. Madhyastha, and Mosharaf Chowdhury. 2021. Oort: Efficient Federated Learning via Guided Participant Selection. In *15th USENIX Symposium on Operating Systems Design and Implementation (OSDI 21)*. USENIX Association, 19–35. <https://www.usenix.org/conference/osdi21/presentation/lai>
- [16] Suyi Li, Yong Cheng, Wei Wang, Yang Liu, and Tianjian Chen. 2020. Learning to Detect Malicious Clients for Robust Federated Learning. *arXiv:2002.00211 [cs.LG]* <https://arxiv.org/abs/2002.00211>
- [17] Alexandra Sasha Luccioni, Sylvain Viguier, and Anne-Laure Ligozat. 2023. Estimating the carbon footprint of bloom, a 176b parameter language model. *Journal of Machine Learning Research* 24, 253 (2023), 1–15.
- [18] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*. PMLR, 1273–1282.
- [19] Baharan Mirzasoleiman, Jeff Bilmes, and Jure Leskovec. 2020. Coresets for data-efficient training of machine learning models. In *International Conference on Machine Learning*. PMLR, 6950–6960.
- [20] David Patterson, Joseph Gonzalez, Urs Hölzle, Quoc Le, Chen Liang, Lluís-Miquel Munguia, Daniel Rothchild, David R So, Maud Texier, and Jeff Dean. 2022. The carbon footprint of machine learning training will plateau, then shrink. *Computer* 55, 7 (2022), 18–28.
- [21] Thanathorn Sukprasert, Abel Souza, Noman Bashir, David Irwin, and Prashant Shenoy. 2024. On the limitations of carbon-aware temporal and spatial workload shifting in the cloud. In *Proceedings of the Nineteenth European Conference on Computer Systems*. 924–941.
- [22] Kahou Tam, Li Li, Bo Han, Chengzhong Xu, and Huazhu Fu. 2023. Federated Noisy Client Learning. *IEEE Transactions on Neural Networks and Learning Systems* (2023), 1–14. doi:10.1109/TNNLS.2023.3336050
- [23] Mingxing Tan and Quoc Le. 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*. PMLR, 6105–6114.
- [24] Tianhao Wang, Johannes Rausch, Ce Zhang, Ruoxi Jia, and Dawn Song. 2020. A principled approach to data valuation for federated learning. *Federated Learning: Privacy and Incentive* (2020), 153–167.
- [25] Philipp Wiesner, Ramin Khalili, Dennis Grinwald, Pratik Agrawal, Lauritz Thamsen, and Odej Kao. 2024. Fedzero: Leveraging renewable excess energy in federated learning. In *Proceedings of the 15th ACM International Conference on Future and Sustainable Energy Systems*. 373–385.
- [26] Xinyi Xu, Lingjuan Lyu, Xingjun Ma, Chenglin Miao, Chuan Sheng Foo, and Bryan Kian Hsiang Low. 2021. Gradient Driven Rewards to Guarantee Fairness in Collaborative Machine Learning. In *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (Eds.), Vol. 34. Curran Associates, Inc., 16104–16117. https://proceedings.neurips.cc/paper_files/paper/2021/file/8682cc30db9c025ecd3fee433f8ab54c-Paper.pdf
- [27] Gang Yan, Hao Wang, and Jian Li. 2021. Critical learning periods in federated learning. *arXiv preprint arXiv:2109.05613* (2021).
- [28] Gang Yan, Hao Wang, Xu Yuan, and Jian Li. 2023. CriticalFL: A Critical Learning Periods Augmented Client Selection Framework for Efficient Federated Learning. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (Long Beach, CA, USA) (KDD '23)*. Association for Computing Machinery, New York, NY, USA, 2898–2907. doi:10.1145/3580305.3599293

Received 19 Mai 2025