# DPET: A Data and Parameter Efficient Training Framework for Green AI

Francesco Scala[1,*], Luigi Pontieri[1] and Sergio Flesca[2]

[1]Institute of High Performance Computing and Networking (ICAR-CNR), Via P. Bucci, 87036 Rende (CS), Italy

[2]Dept. Computer Engineering, Modeling, Electronics, and Systems Engineering (DIMES), University of Calabria, 87036 Rende (CS), Italy

**Abstract**

The worsening climate crisis calls for immediate action to reduce the environmental impact of energy-intensive technologies, including Artificial Intelligence (AI). Reducing AI's environmental footprint involves adopting energy-efficient strategies for training Deep Neural Networks (DNNs). One such strategy is Data Pruning (DP), which decreases the number of training instances, thereby lowering total energy consumption. Several DP methods, such as GraNd and Craig, have been introduced to accelerate model training. On the other hand, Active Learning (AL) techniques, originally designed to iteratively select relevant unlabeled data instances for being labeled by human experts, can also be leveraged to train models on smaller, but informative, subsets. However, despite reducing the volume of training data, many DP and AL-based methods involve expensive computations that may significantly limit their potential for energy savings. In this work"-in-progress", we propose a framework, named *DPET*, that efficiently integrates data selection techniques within an AL-like incremental training. Empirical analyses on a benchmark dataset show that the proposed approach offers a better balance between accuracy and energy efficiency in the training of DNN models.

**Keywords**

Data Pruning, Green-AI, Active Learning, Energy Efficiency, Sustainability

## 1. Introduction

Recent advancements in Artificial Intelligence (AI) have significantly transformed industries like healthcare, finance, and manufacturing, impacting personal and professional life. However, this rapid expansion has raised concerns regarding increased energy consumption and carbon emissions [1]. Deep Learning models, which require vast data and computation for training Deep Neural Networks (DNNs), are major contributors to this surge in energy use [2]. The electricity needed for AI model training, largely generated from non-renewable sources like coal and natural gas, contributes to climate change [3]. In response, Green-AI research aims to reduce AI systems' environmental impact by minimizing energy consumption, utilizing renewable energy, and developing energy-efficient AI hardware. This work particularly addresses the challenge of combining data selection with deep learning methods to lower energy usage while maintaining model accuracy.

**Existing solutions** Several approaches have been proposed to tackle the issue of reducing energy consumption in AI. One key method is Data Pruning (DP), which involves extracting a compact subset, or *coreset*, from a large dataset while preserving its most relevant information [4]. This smaller sample can be used as a more cost-effective substitute for the original dataset in machine learning tasks [5]. However, many DP methods involve heavy computations, which can negate the benefits of reducing the training dataset size. Recent studies show that random sampling schemes often perform as well as, or better than, DP methods [6, 4, 7]. The *Repeated Random Sampling* (*RS2*) method [6] builds on this idea by randomly selecting a data subset for each training epoch, aiming to cut training costs. Despite the potential of DP and random sampling methods, a key limitation is the need to determine

*1st Workshop on Green-Aware Artificial Intelligence, 23rd International Conference of the Italian Association for Artificial Intelligence (AIxIA 2024), November 25–28, 2024, Bolzano, Italy*

*Corresponding author.

francesco.scala@icar.cnr.it (F. Scala); luigi.pontieri@icar.cnr.it (L. Pontieri); sergio.flesca@unical.it (S. Flesca)

0009-0007-5224-0910 (F. Scala); 0000-0003-4513-0362 (L. Pontieri); 0000-0002-4164-940X (S. Flesca)

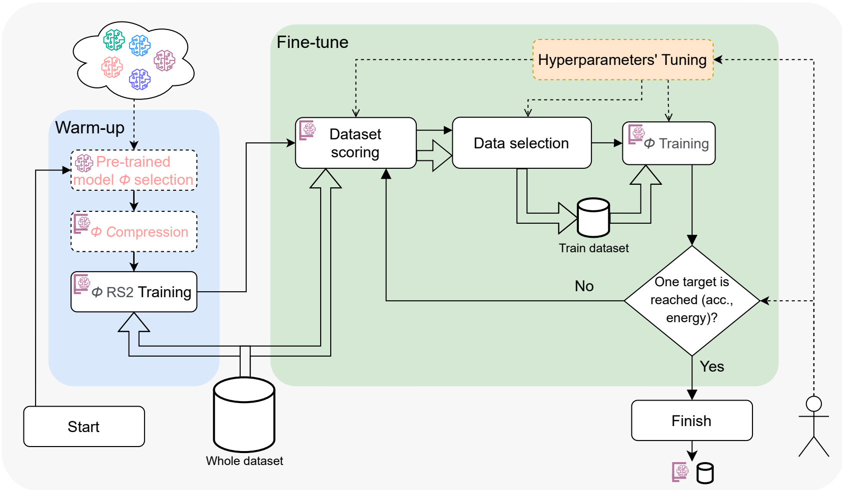



**Figure 1:** Overview of the proposed *DPET* framework.

the optimal data amount beforehand, misjudgments can lead to wasted time and energy. To address this, Active Learning (AL) approaches [8, 9, 10], originally designed to minimize labeling costs, can be leveraged to reduce training costs by focusing on informative data subsets [11, 12, 13]. However, the repeated retraining required by standard AL schemes can be too energy-intensive, as shown in empirical experiments [11].

**Contribution** Given the limitations of current data pruning and sampling methods for efficient Deep Neural Network (DNN) learning, this paper introduces *DPET*, a framework that combines a *RS2*-based DNN warm-up with an iterative AL-like scheme to refine the model with informative data selections.

Experiments on benchmark datasets demonstrate that *DPET* significantly reduces the computational and energetic costs of training large DNN models without sacrificing accuracy. These findings suggest that *DPET* holds promise for promoting more sustainable DNN training, particularly in the context of Green-AI initiatives.

## 2. Proposed approach

Let $\mathscr{D}$ a dataset, to be pruned, that consists of pairs $(x_i, y_i)$, where each $x_i$ is a data instance, and each $y_i$ is a one-hot vector representing a class label, with $C$ classes in total. The goal of data pruning is to extract a representative subset $\mathscr{D}_s$ such that its size is much smaller than $\mathscr{D}$. A DNN model $\phi_\theta$, parameterized by $\theta$, is trained using gradient descent, with the additional benefit that training on $\mathscr{D}_s$ consumes less energy compared to training on the full dataset $\mathscr{D}$.

The proposed *DPET* framework, a "work-in-progress" extension of [13], is currently under development, with the dotted blocks in Figure 1 representing ongoing work. It operates in two distinct phases:

- *Warm-up*: *DPET* first selects the most suitable pre-trained model from an internal pool or some external repositories (like those that have become available recently in many application sectors like, e.g., natural language processing/understanding, computer vision), based on the characteristics of the current dataset. Afterward, it applies optimization techniques such as model pruning, cutout regularization, and low-precision parameter quantization to compress the model and enhance its efficiency. At the end is used the *RS2* algorithm to quickly converge on a preliminary model configuration for $\phi_\theta$. This approach is faster than traditional methods like SGD. However, *RS2*'s performance gain slows down over time, so the algorithm switches to an active learning (AL)-based procedure to continue improving model performance efficiently.

- *Fine-tune*: *DPET* selects additional instances from $\mathcal{D}$ iteratively, using an instance ranking function $f_{rank}$ and a dissimilarity measure $d$ (such as Euclidean distance or KL divergence) to compute an importance score for each instance. The model is updated and trained using both new and "old" (i.e. previously-selected) instances –for the sake of efficiency, the user could require the framework to only select a subset of the old instances leveraging replay-based mechanisms like those used in Continual Learning [14]. This iterative process allows the algorithm to reduce computation and energy costs. This adaptive method is more flexible than traditional data pruning techniques, which require pre-determined reduction levels. Due to the high quantity of parameters involved, a component named *Hyperparameters' Tuner* is involved in the progress to improve *DPET* performances by automatically managing them.

In the end, *DPET* produces a trained model $\phi$ and a *coreset.* This hybrid approach of combining *RS2* with an AL-based fine-tuning procedure helps balance performance and energy efficiency.

**Setting guidelines and implementation choices** AL provides a way to optimize AI model training while reducing energy consumption, in line with Green-AI principles. The idea is to strategically select the most informative data samples from a larger dataset, which can help reduce computational costs required to reach a target accuracy level. However, the extent of energy savings depends on several factors:

- *Data Sampling Effectiveness*: The effectiveness of AL-like data selection strategies is essential to significantly reduce the overall training costs and avoid to undermine model quality. The greater is the per-step data sampling effectiveness (and, hence, the lower is the total number of training instances and AL rounds required), the more substantial is the potential energy saving.
- *Data Sampling Complexity*: AL sampling methods differ in computational cost. Simpler approaches are less resource-intensive, while more complex methods can be costly. If the sampling process is too expensive, it may counteract the overall energy savings.

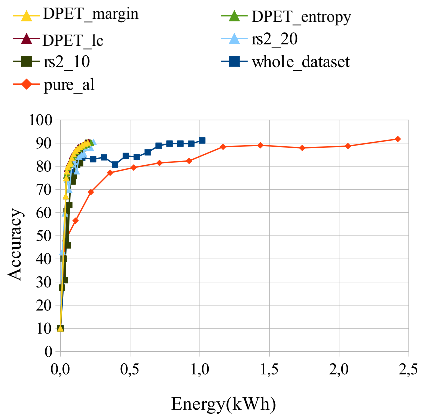
## 3. Experimental Evaluation

**Test setting and terms of comparison** In the experimental evaluation, we used the widely known CIFAR-10 dataset, containing 60,000 images divided into 10 classes. We compared a partial implementation of *DPET* (namely, without using pre-trained models and data replay mechanisms) against several methods: standard full-dataset training (*Standard train*), the *RS2* algorithm [6], the pure AL approach from [11], and state-of-the-art DP methods such as Glister [15], GraphCut [16], CRAIG [17], and GraNd [18], using implementations from DeepCore [4].

Each method was evaluated by measuring both the energy consumption (in Wh) and the accuracy of the trained models. Following the time-to-accuracy approach in [6], we set accuracy targets (from 60% to 90%) and measured the energy requested by each method used to reach these targets, so we measured the energy-to-accuracy, unless it exhausted its budget of energy or epochs beforehand.

**Hyperparameter Configuration** For each test, we trained a ResNet18 model [19] using mini-batch Stochastic Gradient Descent (SGD), alongside a Cross-Entropy loss. We tested *DPET* using three ranking function variants ($f_{rank}$): Least Confidence, Margin Sampling and Entropy scores.

The approach is flexible and can incorporate other AL techniques. For *DPET*, in the warm-up, we ran *RS2* with a 30% data reduction per epoch ($r = 0.3$) over 20 epochs ($bootEpcs = 20$). In the fine-tune rounds, 1,000 instances were selected per round, with 10 optimization epochs. Hyperparameters for *RS2* and the AL method from [11] were set according to their original papers. *RS2* was tested with reduction factors of 20%, 10%, and 5%, with a total budget of 200 epochs.




| Target | 60% | 65% | 70% | 75% | 80% | 85% | 90% |
|---|---|---|---|---|---|---|---|
| Standard train | 20 | 39 | 59 | 78 | 157 | 626 | 1018 |
| AL (margin) [11] | 135 | 218 | 288 | 357 | 711 | 1167 | 2421 |
| GraNd [18] | 838 | 865 | 878 | 902 | 924 | 1042 | 1232 |
| Craig [17] | 175 | 196 | 231 | 248 | 344 | 442 | 627 |
| Glister [15] | 108 | 117 | 133 | 172 | 192 | 380 | 599 |
| GraphCut [16] | 239 | 364 | 396 | 486 | 621 | 762 | 993 |
| RS2 w/o repl 20% | 32 | 39 | 48 | 65 | 82 | 168 | 237 |
| RS2 w/o repl 10% | 44 | 59 | 75 | 97 | 108 | 153 | 197 |
| RS2 w/o repl 5% | 43 | 56 | 62 | 77 | 94 | 121 | - |
| *DPET* (margin) | **19** | **38** | 42 | 47 | **63** | **94** | **196** |
| *DPET* (entropy) | **19** | 39 | **41** | **42** | **63** | 99 | 220 |
| *DPET* (lc) | **19** | **38** | **41** | **42** | **63** | 98 | 204 |

**Figure 1**
Energy-to-accuracy of a ResNet18 for *DPET* compared to *RS2*, AL, standard training and some DP techniques of ResNet18, targeting 90% accuracy on the full CIFAR-10 dataset. Values are reported every 10 training epochs. The plot is shown on the left, while the numerical values are presented in a table on the right.

**Test results** The analysis focuses on three key aspects: computational savings, accuracy, and pruning ratio, comparing the performance of *DPET* with the other techniques. A significant advantage of *DPET* is its iterative approach to data selection, which eliminates the need to pre-determine the amount of data to prune, indeed, it dynamically adds only the necessary data to achieve the target accuracy. This smart data selection allows *DPET* to reach the desired accuracy more quickly, resulting in computational savings, even if the *pruning ratio* is lower than that of other methods.

The results indicate that *DPET*, as shown in the figure 1, outperforms the other techniques in terms of computational savings for the same target accuracy, highlighting the effectiveness of its iterative data selection strategy. This approach not only speeds up the training process but also enables the achievement of all target accuracy levels considered. In contrast, the other analyzed methods (excluding the standard training baseline) fail to meet some of the target accuracy thresholds.

## 4. Conclusion

Based on our analysis, despite its partial implementation, *DPET* stands out as an efficient method for training deep neural network (DNN) models on large datasets, providing significant computational savings compared to standard training and active learning (AL) approaches without compromising model accuracy. It consistently outperforms other data pruning techniques regarding energy consumption across various target accuracy levels. This computational efficiency makes *DPET* particularly suitable for resource-constrained devices and aligns with the goals of Green AI. To enhance the proposed framework, we will finalize the implementation and optimization of the dotted blocks to evaluate the framework's definitive performance.

**Acknowledgment** This work was partially supported by the PNRR research project project FAIR - Future AI Research (PE00000013), Spoke 9 - Green-aware AI, under the NRRP (National Recovery and Resilience Plan) MUR program funded by the NextGenerationEU.

## References

[1] A. de Vries, The growing energy footprint of artificial intelligence, Joule 7 (2023) 2191–2194. doi:https://doi.org/10.1016/j.joule.2023.09.004.

[2] E. Strubell, A. Ganesh, A. McCallum, Energy and policy considerations for deep learning in NLP, in: A. Korhonen, D. R. Traum, L. Màrquez (Eds.), Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2,

2019, Volume 1: Long Papers, Association for Computational Linguistics, 2019, pp. 3645–3650. doi:10.18653/V1/P19-1355.

[3] S. Flesca, F. Scala, E. Vocaturo, F. Zumpano, On forecasting non-renewable energy production with uncertainty quantification: A case study of the italian energy market, Expert Systems with Applications 200 (2022) 116936. doi:https://doi.org/10.1016/j.eswa.2022.116936.

[4] C. Guo, B. Zhao, Y. Bai, Deepcore: A comprehensive library for coreset selection in deep learning, in: Database and Expert Systems Applications: 33rd International Conference, DEXA 2022, Vienna, Austria, August 22–24, 2022, Proceedings, Part I, Springer-Verlag, Berlin, Heidelberg, 2022, pp. 181–195. doi:10.1007/978-3-031-12423-5_14.

[5] N. Sachdeva, J. J. McAuley, Data distillation: A survey, CoRR abs/2301.04272 (2023). doi:10.48550/ARXIV.2301.04272. arXiv:2301.04272.

[6] P. Okanovic, R. Waleffe, V. Mageirakos, K. Nikolakakis, A. Karbasi, D. Kalogerias, N. M. Gürel, T. Rekatsinas, Repeated random sampling for minimizing the time-to-accuracy of learning, in: The Twelfth International Conference on Learning Representations, 2024.

[7] F. Ayed, S. Hayou, Data pruning and neural scaling laws: fundamental limitations of score-based algorithms, 2023. arXiv:2302.06960.

[8] S. Flesca, D. Mandaglio, F. Scala, A. Tagarelli, A meta-active learning approach exploiting instance importance, Expert Systems with Applications 247 (2024) 123320. doi:https://doi.org/10.1016/j.eswa.2024.123320.

[9] S. Flesca, D. Mandaglio, F. Scala, A. Tagarelli, Learning to active learn by gradient variation based on instance importance, in: 2022 26th International Conference on Pattern Recognition (ICPR), 2022, pp. 2224–2230. doi:10.1109/ICPR56361.2022.9956039.

[10] S. Flesca, D. Mandaglio, F. Scala, A. Tagarelli, A meta-active learning approach exploiting instance importance based on learning gradient variation, in: Proceedings of the 31st Symposium of Advanced Database Systems, Galzingano Terme, Italy, July 2nd to 5th, 2023, volume 3478 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2023, pp. 535–544.

[11] D. Park, D. Papailiopoulos, K. Lee, Active learning is a strong baseline for data subset selection, in: Has it Trained Yet? NeurIPS 2022 Workshop, 2022.

[12] F. Scala, S. Flesca, L. Pontieri, Data filtering for a sustainable model training, in: Proceedings of the 32nd Symposium of Advanced Database Systems, Villasimius, Italy, June 23rd to 26th, 2024, volume 3741 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2024, pp. 205–216.

[13] F. Scala, S. Flesca, L. Pontieri, Play it straight: An intelligent data pruning technique for green-ai, in: D. Pedreschi, A. Monreale, R. Guidotti, R. Pellungrini, F. Naretto (Eds.), Discovery Science, Springer Nature Switzerland, Cham, 2025, pp. 69–85.

[14] L. Wang, X. Zhang, H. Su, J. Zhu, A comprehensive survey of continual learning: theory, method and application, IEEE Transactions on Pattern Analysis and Machine Intelligence (2024).

[15] Z. Yang, H. Yang, S. Majumder, J. Cardoso, G. Gallego, Data pruning can do more: A comprehensive data pruning approach for object re-identification, Transactions on Machine Learning Research (2024).

[16] R. Iyer, N. Khargoankar, J. Bilmes, H. Asanani, Submodular combinatorial information measures with applications in machine learning, in: V. Feldman, K. Ligett, S. Sabato (Eds.), Proceedings of the 32nd International Conference on Algorithmic Learning Theory, volume 132 of *Proceedings of Machine Learning Research*, PMLR, 2021, pp. 722–754.

[17] B. Mirzasoleiman, J. Bilmes, J. Leskovec, Coresets for data-efficient training of machine learning models, in: H. D. III, A. Singh (Eds.), Proceedings of the 37th International Conference on Machine Learning, volume 119 of *Proceedings of Machine Learning Research*, PMLR, 2020, pp. 6950–6960.

[18] M. Paul, S. Ganguli, G. K. Dziugaite, Deep learning on a data diet: Finding important examples early in training, in: A. Beygelzimer, Y. Dauphin, P. Liang, J. W. Vaughan (Eds.), Advances in Neural Information Processing Systems, 2021.

[19] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, in: Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778. doi:10.1109/CVPR.2016.90.