

Green Machine Learning

Verónica Bolón-Canedo, Laura Morán-Fernández, Brais Cancela and
Amparo Alonso-Betanzos *

CITIC, Universidade da Coruña, A Coruña, Spain

Abstract. Green machine learning refers to research that is more environmentally friendly and inclusive, not only by producing novel results without increasing the computational cost, but also by ensuring that any researcher with a laptop has the opportunity to perform high-quality research without the need to use expensive cloud servers. Efficient machine learning approaches (especially deep learning) are starting to receive some attention in the research community. This tutorial is concerned with the development of machine learning algorithms that optimize efficiency rather than only accuracy. We provide an overview of this recent field, together with a review of the novel contributions to the ESANN 2023 special session on Green Machine Learning.

1 Introduction

Over the past few years, Artificial Intelligence (AI) and Machine Learning (ML) have brought about a revolution in numerous industries. They have greatly enhanced efficiency and accuracy in sectors like healthcare, finance, transportation, education, and entertainment. To achieve higher performance, ML models have become increasingly complex, resulting in a larger number of parameters to estimate. However, these advancements come at a cost: the energy consumption required for training and running these models has risen significantly.

In the next coming years, this energy consumption is projected to multiply, potentially reaching over 30% of the world's total energy consumption by 2030. Large Language Models (LLMs), such as the recently launched ChatGPT with GPT-4 as backbone model, contribute to this trend with their substantial energy requirements. For instance, ChatGPT with GPT-3.5 allegedly consumed 1,287 megawatts and generated 552 metric tons of carbon dioxide emissions during its training, as reported by various sources. Not to mention that the more recent GPT-4 is estimated to be 10 times larger than its predecessor.

As the environmental impact of this disruptive technology grows almost exponentially, great concerns arise about its carbon footprint, and thus a new paradigm has emerged: Green machine learning. Green ML focuses on developing and implementing sustainable practices and techniques into the design, training, and deployment of the models so as to reduce AI systems' energy consumption and environmental footprint.

*This work was supported by the Ministry of Science and Innovation of Spain (Grant PID2019-109238GB-C22 / AEI / 10.13039 / 501100011033) and together with "NextGenerationE"/PRTR (TED2021-130599A-I00) and by Xunta de Galicia (Grants ED431G 2019/01 and ED431C 2022/44).

Traditional ML algorithms often require large amounts of computational power and data, resulting in significant energy consumption from data centers and increasing greenhouse gas emissions. Green ML seeks to mitigate these environmental impacts by optimizing algorithms, improving hardware efficiency, and adopting sustainable data management practices. Thus, Green ML is characterized by low carbon footprints, small model sizes, low computational complexity, and logical transparency. It offers energy-effective solutions in cloud centers as well as mobile/edge devices. Green ML also provides a clear and logical decision-making process to gain people's trust.

This special session delves into the emerging field of Green ML, exploring the various approaches, methodologies, and innovations that aim to make AI and ML more environmentally sustainable, such as federated learning, transfer learning, etc. We will explore the key challenges and opportunities in reducing energy consumption, minimizing carbon emissions, and promoting ethical and responsible AI practices. The adoption of these approaches not only has those environmental advantages, essential for the future and mandatory in the UE [1], but also allows for cost savings and increased efficiency without compromising performance or accuracy.

2 Green Machine Learning

The binomial artificial intelligence and environment has a dual perspective. On the one hand, we have artificial intelligence as a new technology with the potential to move towards a green economy (green by AI), and on the other, AI as a polluting agent in need of the design of energy-efficient algorithms (green in AI).

According to the existing literature, green algorithms are usually defined as those algorithms capable of maximizing energy efficiency and reducing the environmental impact of AI models, while supporting the use of this technology to respond to different environmental challenges. In this way, two different types of algorithms are being referred to. On the one hand, those which use and training are energy efficient (systems that are themselves green, green by design). On the other hand, algorithms specifically created or used to help meet the environmental challenges set out, among others, in the Paris Agreement on climate change, in the United Nations Sustainable Development Goals¹, or in the most recent European Green Deal².

The debate revolves around the difference between red artificial intelligence, and green artificial intelligence [2]. In 2018 a study³ revealed that the computational needs to train large ML models were doubling every 3.4 months since 2012 (deviating quite a bit from Moore's Law which states that this should happen every 18 months). In 2020, another paper [3] introduces the concept of red AI,

¹<https://sdgs.un.org/goals>

²https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en

³<https://openai.com/research/ai-and-compute>

which “buy” better results at the cost of using massive computational resources. Their authors analyzed more than 60 papers at the most prestigious conferences and concluded that the vast majority (between 75% and 90% depending on the conference) prioritized accuracy over efficiency.

However, the energy consumption of machine learning algorithms should not be seen as something impossible to reduce and does not have to be assumed as the cost of progress in this field. Green machine learning can be defined as machine learning research that produces novel results without increasing computational cost and, ideally, reducing it. In the specialized literature, we can find several strategies that can be carried out to reduce this consumption and will be following described.

- **Algorithmic optimization.** Making algorithms more efficient can have many benefits, in addition to reducing their environmental footprint. This is one of the most productive strategies and the focus of green algorithm development. One area that is quite active is making the inference process efficient, using quantization or energy-aware pruning [4].
- **Hardware optimization.** Choosing more computationally efficient hardware can also contribute to energy savings, as some GPUs have substantially higher efficiency in terms of floating point operations per second (FLOPS) per watt of power usage compared to others. Another important issue is in parallelization. One obvious way to reduce the training time of algorithms is to distribute the computations among several processing cores. Anthony et al. [5] showed that, for a given task, increasing the number of cores to 15 improved both execution time and greenhouse emissions. However, when the reduction in execution time is smaller than the relative increase in the number of cores, distributing the computation can worsen carbon emissions. They performed some experiments demonstrating how marginal improvements in runtime lead to disproportionately large emissions. Finally, edge computing is also a relevant strategy in this context, as the idea is to carry out the computation at the locations where the data is collected or used, thus avoiding the need of sending the data to a datacenter or to the cloud, at the same time that it takes advantage of the limited computational and energy resources of the IoT devices.
- **Choice of data center.** The carbon footprint is directly proportional to the efficiency of the data center and the carbon intensity of its location. The latter is perhaps the most important factor for the total carbon footprint due to the huge variation between countries, from less than 20 gCO_{2e} kWh⁻¹ in Norway and Switzerland to over 800 gCO_{2e} kWh⁻¹ in Australia, South Africa, and some US states.
- **Reducing the pragmatic scaling factor.** Limiting the number of times an algorithm is run, especially those that are computationally expensive, is surely the easiest way to reduce energy consumption. Another possible

strategy is to limit the time spent on hyperparameter tuning, for example with less exhaustive searches [6, 7, 8].

Machine learning risks contributing significantly to climate change if it follows the energy consumption trend of large models. Therefore, if researchers and developers are aware of the energy and CO₂ footprint they are causing, they are more likely to take measures to reduce it. This is why several tools for calculating and predicting energy consumption from artificial intelligence algorithms are emerging. Carbontracker⁴ is one such tool, in particular for estimating the consumption of deep learning models (famously computationally and energy intensive). Another openly available tool is Green Algorithms⁵, which can be easily integrated with computational processes as it requires a minimum amount of information and does not interfere with existing code.

3 Recent contributions

As an area of active research and development in recent years, there are a variety of recent contributions in Green ML that are focused on a variety of strategies:

- **Environmental Impact Assessment:** Beyond algorithmic advancements, researchers have also been working on frameworks and tools to assess the environmental impact of machine learning systems. Over the recent years, Yigitcanlar et al. [9] explored how cryptocurrency mining has led to increased energy consumption worldwide, concluding that Bitcoin miners are expected to consume approximately 130 terawatt hours of energy (TWh), which represents approximately 0.6% of global electricity consumption. This puts the bitcoin economy on par with the CO₂ emissions of small developing countries such as Sri Lanka or Jordan. In the field of Natural Language Processing (NLP), Strubell et al. [10] raised the alarm when they discovered that training a widely used computational model emitted as much CO₂ as five cars in their lifetime. More recently, George et al. [11] reported that language models like GPT-3 [12] or ChatGPT reportedly consumed over 700,000 liters of water during its training phase, which is equivalent to the amount of water used by an average American household in about 20 years.
- **Energy-Efficient Models:** Researchers have focused on designing optimization techniques that reduce the computational resources required, thus minimizing energy consumption. These approaches include sparse training methods [13, 14, 15], quantization techniques [16, 17, 18], and low-precision arithmetic operations [19, 20], which decrease both the memory footprint and the computational complexity of training models.
- **Hardware Acceleration:** Another area of focus in Green ML is the development of specialized hardware accelerators (like GPUs or TPUs [21])

⁴<https://carbontracker.org>

⁵<http://www.green-algorithms.org>

tailored specifically for machine learning tasks, and the ability to customize machine learning models to be used in that specific hardware [22, 23]. For instance, works in edge computing [24, 25] suggest that we can achieve not only less-energy demanding algorithms but also privacy-preserving architectures that can be trained without sharing all of the data.

- **Data Center Optimization:** Researchers have developed algorithms and frameworks that dynamically manage server loads [26, 27], adjust cooling systems [28, 29], and optimize resource allocation to reduce energy consumption in data centers [30].
- **Energy-Efficient Structures:** Green ML contributes to making structures more energy-efficient. ML models can analyze sensor data from smart buildings [31, 32] and cities [33, 34] to optimize heating, ventilation, and air conditioning (HVAC) systems, lighting, and energy usage patterns, resulting in significant energy savings and reduced carbon emissions [35, 36].

Other important fields of study like climate change [37], sustainable agriculture [38, 39], renewable energy forecasting [40, 41] or waste management [42] are gaining special interest over the last years, increasing the scope of the use of machine learning in different green scenarios.

4 Contributions to the special session

The Green Machine Learning special session has garnered contributions from multiple research groups, showing methodologies that tackle both the theoretical and applied dimensions of the core subject matter. The subsequent section provides concise introductions to each accepted paper, highlighting their significance and relevance.

Two of the accepted papers propose more efficient methods in the field of feature selection. Feature selection (FS) is a fundamental task in machine learning, as it can help in reducing dimension and thus contributes to more understandable models. García-Castillo et al. [43], in particular, develop a novel and efficient method for feature selection in domain adaptation, a type of transfer learning where the source and target domains share the feature space and task but differ in their distributions. The researchers present an alternative method, Mutual Information Maximization (MIM), as opposed to the commonly used evolutionary algorithms in the field. MIM offers advantages such as eliminating the need for an iterative search process and being computationally less demanding. Through experiments conducted on two datasets, they demonstrate that their proposed method outperforms two previously suggested alternatives that rely on evolutionary algorithms: Sticky Binary Particle Swarm Optimization using classifiers or data complexity metrics. Specifically, their method shows superior efficiency, speed, and the capability to select a smaller subset of features, while still achieving competitive accuracy results in classification tasks.

In the work of Suárez-Marcote et al. [44], an efficient and green approach for feature selection is proposed, using principles from information theory. A key

novelty of their method lies in employing logarithmic division and fixed-point precision techniques. The authors conducted experiments involving two popular information-theoretic filter methods, namely Mutual Information Maximization and Joint Mutual Information Maximization. The results of their study suggest that adopting low-precision representations has the potential to decrease energy consumption without compromising information content. Additionally, the use of logarithmic division provides an extra level of energy efficiency, which proves beneficial in resource-constrained environments. In terms of classification performance, the modified algorithms employing fixed-point representation and logarithmic division demonstrated comparable results to the baseline approaches, even when there were slight discrepancies in the obtained feature rankings.

Lourenço et al. [45] present a scalable and sustainable methodology for condition-based maintenance of wheel out-of-roundness. The methodology involves developing an anomaly detection model using locality-sensitive hashing to analyze extensive time series datasets generated by railway network sensors. The implementation leverages the Apache Spark framework for parallel processing, facilitating efficient data analysis across multiple clusters. The performance study conducted demonstrates the robustness and efficiency of the fault detection method, even when confronted with various wheel profiles, track irregularities, train speeds, sensor positions, and non-linear environmental factors that may introduce artificial noise and signal distortion.

Shumska et al. [46] introduce a novel and efficient approach, called block-diagonal dissimilarity extension, within the Generalized Matrix Learning Vector Quantization (GMLVQ) framework, and applied it to agricultural multispectral images. Their findings suggest that, for the Statlog dataset classification in Euclidean space, using the quadratic form yields the highest accuracy, while the full and block-diagonal configurations of GMLVQ matrices yield comparable results. Their models demonstrate similar accuracy levels as previously reported methods such as neural networks, support vector machines, and k-nearest neighbors, while offers the advantages of interpretability, reduced complexity, and the ability to visualize multispectral data.

References

- [1] European Union. Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts, 2021.
- [2] Payal Dhar. The carbon impact of artificial intelligence. *Nat. Mach. Intell.*, 2(8):423–425, 2020.
- [3] Roy Schwartz, Jesse Dodge, Noah A Smith, and Oren Etzioni. Green ai. *Communications of the ACM*, 63(12):54–63, 2020.
- [4] Tien-Ju Yang, Yu-Hsin Chen, and Vivienne Sze. Designing energy-efficient convolutional neural networks using energy-aware pruning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5687–5695, 2017.
- [5] Lasse F Wolff Anthony, Benjamin Kanding, and Raghavendra Selvan. Carbontracker: Tracking and predicting the carbon footprint of training deep learning models. *arXiv preprint arXiv:2007.03051*, 2020.

- [6] James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *Journal of machine learning research*, 13(2), 2012.
- [7] Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, 25, 2012.
- [8] Lisha Li, Kevin Jamieson, Giulia DeSalvo, Afshin Rostamizadeh, and Ameet Talwalkar. Hyperband: A novel bandit-based approach to hyperparameter optimization. *The Journal of Machine Learning Research*, 18(1):6765–6816, 2017.
- [9] Tan Yigitcanlar, Rashid Mehmood, and Juan M Corchado. Green artificial intelligence: Towards an efficient, sustainable and equitable technology for smart cities and futures. *Sustainability*, 13(16):8952, 2021.
- [10] Emma Strubell, Ananya Ganesh, and Andrew McCallum. Energy and policy considerations for deep learning in nlp. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3645–3650, 2019.
- [11] A Shaji George, AS Hovan George, and AS Gabrio Martin. The environmental impact of ai: A case study of water consumption by chat gpt. *Partners Universal International Innovation Journal*, 1(2):97–104, 2023.
- [12] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.
- [13] Cédric Renggli, Saleh Ashkboos, Mehdi Aghagolzadeh, Dan Alistarh, and Torsten Hoefer. Sparcml: High-performance sparse communication for machine learning. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 1–15, 2019.
- [14] Yi Tay, Dara Bahri, Liu Yang, Donald Metzler, and Da-Cheng Juan. Sparse sinkhorn attention. In *International Conference on Machine Learning*, pages 9438–9447. PMLR, 2020.
- [15] Sameer Bibikar, Haris Vikalo, Zhangyang Wang, and Xiaohan Chen. Federated dynamic sparse training: Computing less, communicating less, yet learning better. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 6080–6088, 2022.
- [16] Anis Elgabli, Jihong Park, Amrit Singh Bedi, Chaouki Ben Issaid, Mehdi Bennis, and Vaneet Aggarwal. Q-gadmm: Quantized group admm for communication efficient decentralized machine learning. *IEEE Transactions on Communications*, 69(1):164–181, 2020.
- [17] Sarit Khirirat, Sindri Magnússon, Arda Aytakin, and Mikael Johansson. A flexible framework for communication-efficient machine learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 8101–8109, 2021.
- [18] Guangxuan Xiao, Ji Lin, Mickael Seznec, Hao Wu, Julien Demouth, and Song Han. Smoothquant: Accurate and efficient post-training quantization for large language models. In *International Conference on Machine Learning*, pages 38087–38099. PMLR, 2023.
- [19] Ofir Zafrir, Guy Boudoukh, Peter Izsak, and Moshe Wasserblat. Q8bert: Quantized 8bit bert. In *2019 Fifth Workshop on Energy Efficient Machine Learning and Cognitive Computing-NeurIPS Edition (EMC2-NIPS)*, pages 36–39. IEEE, 2019.
- [20] Zhenhua Liu, Yunhe Wang, Kai Han, Wei Zhang, Siwei Ma, and Wen Gao. Post-training quantization for vision transformer. *Advances in Neural Information Processing Systems*, 34:28092–28103, 2021.
- [21] Norm Jouppi, George Kurian, Sheng Li, Peter Ma, Rahul Nagarajan, Lifeng Nai, Nishant Patil, Suvinay Subramanian, Andy Swing, Brian Towles, et al. Tpu v4: An optically reconfigurable supercomputer for machine learning with hardware support for embeddings. In *Proceedings of the 50th Annual International Symposium on Computer Architecture*, pages 1–14, 2023.

- [22] Mario Osta, Mohamad Alameh, Hamoud Younes, Ali Ibrahim, and Maurizio Valle. Energy efficient implementation of machine learning algorithms on hardware platforms. In *2019 26th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, pages 21–24. IEEE, 2019.
- [23] Dimitrios Stamoulis, Ruizhou Ding, Di Wang, Dimitrios Lymberopoulos, Bodhi Priyanta, Jie Liu, and Diana Marculescu. Single-path nas: Designing hardware-efficient convnets in less than 4 hours. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 481–497. Springer, 2019.
- [24] Shivani Singh, Razia Sulthana, Tanvi Shewale, Vinay Chamola, Abderrahim Benslimane, and Biplab Sikdar. Machine-learning-assisted security and privacy provisioning for edge computing: A survey. *IEEE Internet of Things Journal*, 9(1):236–260, 2021.
- [25] Yichen Wan, Youyang Qu, Longxiang Gao, and Yong Xiang. Privacy-preserving blockchain-enabled federated learning for b5g-driven edge computing. *Computer Networks*, 204:108671, 2022.
- [26] Arezoo Ghasemi and Abolfazl Toroghi Haghighat. A multi-objective load balancing algorithm for virtual machine placement in cloud data centers based on machine learning. *Computing*, 102:2049–2072, 2020.
- [27] Haiwei Dong, Ali Munir, Hanine Tout, and Yashar Ganjali. Next-generation data center network enabled by machine learning: Review, challenges, and opportunities. *IEEE Access*, 9:136459–136475, 2021.
- [28] Yuanlong Li, Yonggang Wen, Dacheng Tao, and Kyle Guan. Transforming cooling optimization for green data center via deep reinforcement learning. *IEEE transactions on cybernetics*, 50(5):2002–2013, 2019.
- [29] Zhiwei Cao, Xin Zhou, Han Hu, Zhi Wang, and Yonggang Wen. Toward a systematic survey for carbon neutral data centers. *IEEE Communications Surveys & Tutorials*, 24(2):895–936, 2022.
- [30] Rafael Moreno-Vozmediano, Rubén S Montero, Eduardo Huedo, and Ignacio M Llorente. Efficient resource provisioning for elastic cloud services based on machine learning techniques. *Journal of Cloud Computing*, 8(1):1–18, 2019.
- [31] Basheer Qolomany, Ala Al-Fuqaha, Ajay Gupta, Driss Benhaddou, Safaa Alwajidi, Junaid Qadir, and Alvis C Fong. Leveraging machine learning and big data for smart buildings: A comprehensive survey. *IEEE Access*, 7:90316–90356, 2019.
- [32] Anh-Duc Pham, Ngoc-Tri Ngo, Thi Thu Ha Truong, Nhat-To Huynh, and Ngoc-Son Truong. Predicting energy consumption in multiple buildings using machine learning for improving energy efficiency and sustainability. *Journal of Cleaner Production*, 260:121082, 2020.
- [33] Marijana Zekić-Sušac, Saša Mitrović, and Adela Has. Machine learning based system for managing energy efficiency of public sector as an approach towards smart cities. *International journal of information management*, 58:102074, 2021.
- [34] Taher M Ghazal, Mohammad Kamrul Hasan, Munir Ahmad, Haitham M Alzoubi, and Muhammad Alshurideh. Machine learning approaches for sustainable cities using internet of things. In *The Effect of Information Technology on Business and Marketing Intelligence Systems*, pages 1969–1986. Springer, 2023.
- [35] Nikola Milojevic-Dupont and Felix Creutzig. Machine learning for geographically differentiated climate change mitigation in urban areas. *Sustainable Cities and Society*, 64:102526, 2021.
- [36] Maomao Zhang, Abdulla-Al Kafy, Pengnan Xiao, Siyu Han, Shangjun Zou, Milan Saha, Cheng Zhang, and Shukui Tan. Impact of urban expansion on land surface temperature and carbon emissions using machine learning algorithms in wuhan, china. *Urban Climate*, 47:101347, 2023.

- [37] David Rolnick, Priya L Donti, Lynn H Kaack, Kelly Kochanski, Alexandre Lacoste, Kris Sankaran, Andrew Slavin Ross, Nikola Milojevic-Dupont, Natasha Jaques, Anna Waldman-Brown, et al. Tackling climate change with machine learning. *ACM Computing Surveys (CSUR)*, 55(2):1–96, 2022.
- [38] Ruhollah Taghizadeh-Mehrjardi, Kamal Nabiollahi, Leila Rasoli, Ruth Kerry, and Thomas Scholten. Land suitability assessment and agricultural production sustainability using machine learning models. *Agronomy*, 10(4):573, 2020.
- [39] Balsher Singh Sidhu, Zia Mehrabi, Navin Ramankutty, and Milind Kandlikar. How can machine learning help in understanding the impact of climate change on crop yields? *Environmental Research Letters*, 18(2):024008, 2023.
- [40] Sheraz Aslam, Herodotos Herodotou, Syed Muhammad Mohsin, Nadeem Javaid, Nouman Ashraf, and Shahzad Aslam. A survey on deep learning methods for power load and renewable energy forecasting in smart microgrids. *Renewable and Sustainable Energy Reviews*, 144:110992, 2021.
- [41] Monika Sandelic, Saeed Peyghami, Ariya Sangwongwanich, and Frede Blaabjerg. Reliability aspects in microgrid design and planning: Status and power electronics-induced challenges. *Renewable and Sustainable Energy Reviews*, 159:112127, 2022.
- [42] Wanjun Xia, Yanping Jiang, Xiaohong Chen, and Rui Zhao. Application of machine learning algorithms in municipal solid waste management: A mini review. *Waste Management & Research*, 40(6):609–624, 2022.
- [43] Guillermo Castillo-García, Laura Morán-Fernández, and Verónica Bolón-Canedo. Efficient feature selection for domain adaptation using mutual information maximization. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 2023.
- [44] Samuel Suárez-Marcote, Laura Morán-Fernández, and Verónica Bolón-Canedo. Logarithmic division for green feature selection: an information-theoretic approach. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 2023.
- [45] Afonso Lourenço, Carolina Ferraz, Jorge Meira, Goreti Marreiros, Verónica Bolón-Canedo, and Amparo Alonso-Betanzos. Automated green machine learning for condition-based maintenance. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 2023.
- [46] Mariya Shumska and Kerstin Bunte. Multispectral texture classification in agriculture. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 2023.