Black Sea Journal of Engineering and Science
doi: 10.34248/bsengineering.1858749

Open Access Journal
e-ISSN: 2619 – 8991

Research Article
Volume 9 - Issue 2: 616-623 / March 2026

COMPARATIVE ANALYSIS OF OPEN-SOURCE DEEP LEARNING MODELS IN TERMS OF ENERGY CONSUMPTION, COMPUTATIONAL LOAD, AND PERFORMANCE

Yasin SANCAR[1]*

[1]*Ataturk University, Faculty of Open and Distance Education, Department of Information Technology, 25240, Erzurum, Türkiye*

Abstract: The energy consumption of deep learning models during training and inference processes has become an important performance indicator, especially for applications running on resource-constrained devices. Although there are significant differences in computational costs between different architectures, studies that comprehensively compare the energy efficiency of models are limited. In this study, six widely used models MobileNetV2, EfficientNet-B0, ResNet50, DenseNet121, Xception, and VGG19 were trained and evaluated under the same dataset and experimental settings. Real-time power measurements were performed on an RTX 2070 GPU to calculate each model's total energy consumption during training and inference, average power value, frames per second (FPS), and energy cost per image (J/image). The findings show that lightweight architectures are significantly more efficient: MobileNetV2 achieved the lowest energy consumption at 0.2289 J/image during inference, while EfficientNet-B0 offered balanced performance in terms of accuracy and energy usage. In contrast, VGG19 stood out as the least efficient model due to its high power requirements. The results reveal that model architecture has a direct impact on energy consumption and that model selection plays a critical role in the design of sustainable artificial intelligence systems.

Keywords: Energy consumption, Deep learning, Power profiling, Green AI, Model efficiency

***Corresponding author:** Ataturk University, Faculty of Open and Distance Education, Department of Information Technology, 25240, Erzurum, Türkiye
E mail: yasinsancar@atauni.edu.tr (Y. SANCAR)
Yasin SANCAR https://orcid.org/0000-0002-4200-1293

Received: January 07, 2026
Accepted: February 05, 2026
Published: March 15, 2026

Cite as: Sancar, Y. (2026). Comparative analysis of open-source deep learning models in terms of energy consumption, computational load, and performance. *Black Sea Journal of Engineering and Science*, *9*(2), 616-623.

1. Introduction

Deep learning models have revolutionized various disciplines with their significant achievements in image recognition, natural language processing, and many other tasks in the field of artificial intelligence (Xu et al., 2023). The increasing complexity and prevalence of these models have brought to the fore a serious problem: high energy consumption and its environmental impact (Tuğaç, 2023). In particular, the training and deployment of large-scale deep learning models require significant amounts of energy and increase the carbon footprint. to the extent that training large models such as GPT-3 can be equivalent to the carbon dioxide emissions emitted by a car over its lifetime (Bozkurt, 2024). This situation has led to growing concerns about the sustainability and environmental responsibility of artificial intelligence technologies (Söyler, 2025).

Traditionally, the success of deep learning models has generally been evaluated using performance metrics such as accuracy and inference speed. However, with increasing energy consumption and environmental concerns, the energy efficiency and ecological footprint of models have become evaluation criteria that are at least as important as these metrics (Xu et al., 2023). In this context, understanding the energy consumption characteristics of different deep learning architectures is critical for developing more sustainable and efficient artificial intelligence solutions (Xu et al., 2023).

In this study, the energy consumption performance of six commonly used convolutional neural network architectures MobileNetV2, EfficientNet-B0, ResNet50, DenseNet121, Xception, and VGG19 were comparatively analyzed. These models have a wide range of applications in tasks such as image classification and differ in terms of architectural complexity and performance (Xu et al., 2023). In particular, lighter architectures (e.g., MobileNetV2) are noted to offer low energy consumption potential (Qi et al., 2022). The study aims to analyze the energy consumption patterns exhibited by these models when run on an NVIDIA RTX 2070 graphics processing unit. Although GPUs are widely used to accelerate deep learning workloads, their high energy consumption has often been overlooked (Tang et al., 2019). Therefore, understanding the impact of the hardware platform on energy performance is also of great importance (Wang et al., 2021; Wang et al., 2024).

This comparative analysis aims to assist researchers and developers in making more informed decisions by emphasizing that energy efficiency is an important factor in the selection of deep learning models, alongside accuracy and speed.

1.1. Literature Review

Deep learning models have made a significant impact with their achievements in image recognition, natural language processing, and many other artificial intelligence applications. However, the increasing complexity and prevalence of these models have also raised serious concerns, such as high energy consumption and its environmental impacts (Aquino-Brítez et al., 2025; Li et al., 2022; Mehlin et al., 2023). The training and deployment of large-scale deep learning models require significant amounts of energy, sparking academic debates about the sustainability of artificial intelligence technologies (Ji & Jiang, 2026).

Traditionally, the success of deep learning models has been measured by performance metrics such as accuracy, inference speed, and computational efficiency (Getzner et al., 2023; Gowda et al., 2024). However, due to AI's growing environmental footprint, the energy efficiency and ecological impact of models have become evaluation criteria as important as model performance (Tripp et al., 2024; Xu et al., 2023). In this context, understanding the energy consumption characteristics of different deep learning architectures is critical for developing more sustainable and efficient artificial intelligence solutions (del Rey et al., 2023).

Various studies have compared the energy consumption performance of different convolutional neural network architectures. For example, it has been shown that lighter architectures such as MobileNetV2 have significantly lower energy consumption compared to heavier models such as VGG16 (Dey et al., 2020). Similarly, some studies have noted that models such as MobileNetV2 and ResNet50 still offer competitive accuracy levels with lower energy consumption. These findings highlight the direct impact of model architecture on energy efficiency and emphasize the importance of developing lighter, more efficient architectures (Gowda et al., 2024).

Graphics Processing Units (GPUs) are widely used in deep learning workloads, but their energy consumption is also a significant factor (Tang et al., 2019). It has been determined that dynamic management techniques for GPUs, such as Dynamic Voltage and Frequency Scaling, can significantly reduce energy consumption in deep learning training and inference processes (Tang et al., 2019). Research analyzes the accuracy-efficiency trade-offs by examining the power consumption of deep learning models on different GPUs. Modern GPUs, such as the NVIDIA RTX series, offer high performance while energy efficiency remains a focus for researchers. For example, one study measured the instantaneous power draw of deep learning models on an 8-GPU NVIDIA H100 HGX node and found that the maximum power draw was 18% lower than the value specified by the manufacturer (Latif et al., 2024).

Various methodologies have been developed to accurately measure and estimate the energy consumption of deep learning models. These methodologies include hardware-based measurement tools (e.g., wattmeters) and software-based estimation tools (Bouza et al., 2023; Rodriguez et al., 2024; Tripp et al., 2024). Some studies have presented approaches that aim to determine the model's energy consumption by estimating it layer by layer, without the need for direct energy measurement (Getzner et al., 2023). Furthermore, strategies and tools for reducing energy consumption at different stages of the deep learning lifecycle are also being explored in the literature (Karamchandani et al., 2024). These studies guide developers and researchers in better understanding and managing the environmental impacts of artificial intelligence systems (Tu et al., 2023).

2. Materials and Methods

2.1. Computing Environment

All experiments conducted in this study were performed in a carefully controlled computing environment to ensure the reliability and reproducibility of energy consumption measurements. The experiments were carried out on a system equipped with an NVIDIA GeForce RTX 2070 graphics processing unit, an Intel Core i7 processor, and 16 GB of RAM. Windows 11 Professional was used as the operating system, and compatible versions of Python 3.10, TensorFlow/Keras, CUDA, and cuDNN were preferred for model runs. This eliminated performance fluctuations that could arise from the software infrastructure. The GPU's instantaneous power consumption was sampled at the millisecond level using an NVIDIA Management Library (NVML)-based profiling infrastructure. During measurements, GPU frequency scaling mechanisms were disabled to ensure the hardware operated in a fixed performance mode. The system memory was cleared before each run, the GPU temperature was balanced, and unnecessary processes were disabled to prevent background load from affecting the measurements. These preparatory steps allowed all models to be evaluated under the same thermal and hardware conditions. Power measurements were collected using the NVIDIA Management Library (NVML) with a fixed sampling interval of 200 ms, capturing GPU-only power draw. All experiments were conducted on an NVIDIA GeForce RTX 2070 Super GPU under Windows (WDDM) using CUDA Toolkit 11.2, cuDNN 8.1.0, and NVIDIA driver version 591.59 to ensure measurement reproducibility.

2.2. Deep Learning Architectures Evaluated

Six common deep learning models with different computational intensities and architectural design principles were selected for energy efficiency evaluation: MobileNetV2, EfficientNet-B0, ResNet50, DenseNet121, Xception, and VGG19. These models offer a wide variety in terms of building blocks, number of parameters, and FLOP values. For example, MobileNetV2 has a very lightweight and energy-efficient structure with inverted residual blocks and depth-wise convolutions, while deeper and more traditional convolutional networks such as VGG19 increase energy consumption with high computational costs. Therefore, the selected architectures provide an ideal basis for comparison to examine how energy efficiency depends not only on the number of parameters but also on architectural innovations, layer placement, computational intensity, and tensor flow.

2.3. Dataset and Preprocessing Process

The dataset used in this study is the 320-pixel resolution version of the Imagenette dataset provided by FastAI, which is a curated subset of ImageNet designed for rapid experimentation (Howard & Gugger, 2020). Imagenette is a subset of ImageNet containing only 10 of the 1,000 classes, selected from objects that are relatively easy to identify visually. The Imagenette-320 version used in the study contains a total of 13,394 images, which are divided into training, validation, and test subdirectories. This scale offers a more compact structure compared to the full ImageNet, while still providing sufficient diversity and sample size to evaluate the energy consumption characteristics of deep learning models.

The images were obtained from the official GitHub repository of the dataset (fastai/imagenette) and used directly without any subsampling, filtering, or relabeling. Each model was trained on images rescaled to fit its input size and subjected to the same preprocessing steps during inference. Normalization was performed using the mean and variance values used in the pre-trained versions of each architecture, ensuring a consistent data distribution during both training and inference.

Data augmentation techniques were kept limited to avoid unnecessarily affecting energy measurements, and transformations that could increase computational load were specifically excluded from the study. This choice is a requirement of the experimental design, which aims to ensure that differences in energy consumption stem solely from the computational intensity and layer structure of the architecture. Imagenette's simple class structure allowed models to be trained faster, enabling energy consumption profiles to be reliably extracted through short-term but consistent observations.

2.4. Training and Inference Processes

All models were run under the same training protocol. During the training phase, each model used the same number of epochs, the same batch size, the same optimization algorithm, and the same learning rate settings. Throughout all stages of the training process, the GPU's power consumption was continuously monitored, and the total energy profile was extracted. After training, inference evaluation was performed, and each model was run sequentially on the test set to measure energy consumption, average power level, and output speed. Prior to the inference process, the GPU temperature was stabilized, the memory was cleared, and all potential load sources in the system were disabled. This preparation prevented the inference performance from being affected by thermal conditions or system load and increased the consistency of the measurements.

Model training was conducted using the same training protocol for all architectures, ensuring that energy consumption comparisons between different models were fair and reproducible. During the training process, data was split into a fixed ratio of 80% training and 20% validation, and all models were subjected to the same data augmentation techniques (random horizontal flip, random rotation, normalization). The Adam optimizer was used for all networks during the optimization phase, with an initial learning rate of 0.001 and beta values set to (0.9, 0.999). The Cosine Annealing LR Scheduler was used to gradually decrease the learning rate as training progressed. The batch size was kept constant at 32 for all models, ensuring that memory usage and GPU load remained consistent throughout the experiments. Cross-Entropy Loss was chosen as the loss function, and training was conducted for a total of 10 epochs for each model. To ensure the GPU operated at full load and remained stable, mixed-precision was not used during the training process, thereby minimizing potential fluctuations in energy consumption. This common configuration ensured that the energy behaviors of different architectures could be observed solely in terms of effects arising from architectural differences, strengthening the methodological validity of the energy performance comparison between models.

2.5. Energy Measurement Method

In this study, power consumption refers to the instantaneous electrical power draw measured in watts (W), while energy consumption represents the accumulated energy over time measured in joules (J). The term energy cost is used to express the energy consumed per processed image (J/image).

The total energy consumption of the GPU is defined as the integral of the instantaneous power over time, as shown in equation (1).

$$E = \int_0^T P(t)\, dt \quad (1)$$

Since power is sampled at discrete time intervals in our measurements, the total energy used during an experiment is approximated by summing the sampled power values, as shown in equation (2).

$$E \approx \sum_i P_i \cdot \Delta t \quad (2)$$

The average power consumption over the entire training or inference period is then obtained by dividing the total energy by the elapsed time, as shown in equation (3).

$$P\_avg = E / T \quad (3)$$

To quantify the energy required to process a single input sample, the energy per image is computed by normalizing the total energy by the number of processed images, as shown in equation (4).

$$E\_image = E / N \quad (4)$$

Finally, the inference throughput is expressed in frames per second by dividing the number of processed images by the total inference time, as shown in equation (5).

$$FPS = N / T \quad (5)$$

3. Results

3.1. Training Energy Consumption Analysis

This section comparatively examines the energy consumption dynamics that emerge during the training process of the six evaluated deep learning architectures. The findings reveal that architectural design principles directly determine energy efficiency, particularly highlighting the significant advantages modern and lightweight networks offer compared to traditional architectures. Table 1 shows the energy consumption metrics measured during the training process of the evaluated models. It is observed that modern, lightweight architectures consume significantly less energy, while traditional and deep architectures consume considerably more energy despite similar accuracy levels. Figure 1 compares the total training energy usage of the six deep learning architectures, showing that lightweight models such as MobileNetV2 and EfficientNet-B0 consume significantly less energy than deeper architectures like VGG19.

Table 1. Training energy consumption metrics of evaluated models

Model	Duration (second)	Average Power (Watt)	Total Energy(J)	Model	Duration (second)
MobileNetV2	1162.06	64.11	74,496.88	0.7867	0.9702
EfficientNet-B0	1368.82	73.83	101,059.64	1.0673	0.9845
ResNet50	1212.64	135.22	163,979.70	1.7318	0.9931
DenseNet121	1951.50	99.74	194,652.35	2.0557	0.9839
Xception	1292.86	156.43	202,247.09	2.1359	0.9949
VGG19	21867.35	90.47	1,978,405.28	20.8935	0.9845

The results obtained show that MobileNetV2 is by far the most energy-efficient architecture. The model consumed only 74,496.88 J of energy and operated at an average power of 64.11 W. Furthermore, achieving 97.02% validation accuracy demonstrates that energy efficiency does not result in a significant loss in accuracy. EfficientNet-B0 also offers a strong balance with similarly low energy consumption (101,059.64 J) and high accuracy (98.45%).

In contrast, energy consumption increases significantly in deeper and more complex networks. ResNet50 and DenseNet121 consumed 163,979.70 J and 194,652.35 J of energy, respectively; although both models achieved over 98% accuracy, they lagged behind MobilenetV2 and EfficientNet-B0 in terms of energy-performance balance. This demonstrates that while residual and dense connections increase representational power, they also raise energy costs during training.

Xception consumed a total of 202,247.09 J of energy with a high average power consumption (156.43 W) despite its discretized convolution structure. The model's achievement of 99.49% accuracy is noteworthy; however, its energy cost is approximately three times that of MobilenetV2.

As expected, the highest energy consumption occurred in the VGG19 model. Its total energy consumption of 1,978,405.28 J is excessively high compared to the other evaluated architectures and is approximately 26 times higher than MobileNetV2. This result clearly demonstrates that the deep and dense structure of the VGG architecture, which lacks modern efficiency mechanisms, is not sustainable in terms of energy.

Overall, the findings reveal that low-complexity modern architectures provide the most balanced performance in terms of both accuracy and energy efficiency, while traditional or excessively deep networks are not suitable for sustainable AI applications due to their high computational cost.

3.2. Energy Consumption Analysis of the Inference Process

The energy consumption results of the six evaluated architectures during the inference phase are summarized in Table 2. The results reveal significantly lower total energy requirements compared to the training process, while also demonstrating that the efficiency differences between the models are clearly maintained during the inference phase. Lightweight architectures, in particular, are extremely advantageous in terms of energy consumption, with both lower average power consumption and higher frames per second (FPS) values.

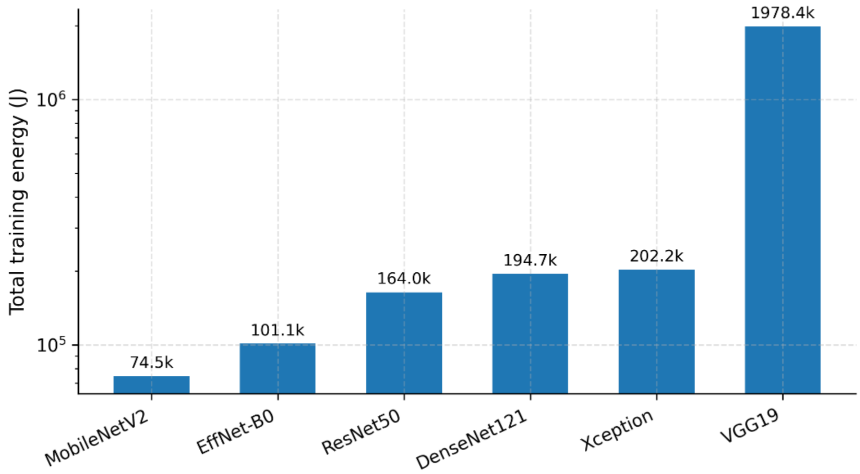

Figure 1. Total training energy consumption of evaluated models.

Table 2. Inference energy consumption metrics of evaluated models

Model	Duration (second)	FPS	Average Power (W)	Total Energy (J)	Energy / Image (J)
MobileNetV2	29.865	131.42	30.09	898.76	0.22898
EfficientNet-B0	30.548	128.49	35.83	1094.61	0.27888
ResNet50	28.801	136.28	71.80	2067.78	0.52682
DenseNet121	39.151	100.25	67.05	2625.26	0.66886
Xception	32.651	120.21	72.18	2356.82	0.60046
VGG19	40.056	97.99	107.99	4325.52	1.10204

Table 2 shows the average power, total energy, energy cost per image, and inference speed measured during the inference phase. It is observed that modern and lightweight architectures consume significantly less energy, while deep and traditional architectures have higher energy costs. Figure 2 illustrates the per-image energy consumption during inference, highlighting the efficiency advantage of MobileNetV2 and EfficientNet-B0 compared to heavier architectures.

The highest energy efficiency during inference was observed in MobileNetV2. The model consumed only 898.76 J of total energy and had the lowest cost among all evaluated architectures at 0.2289 J per image. It also achieved a speed of 131.42 FPS, demonstrating an ideal profile in terms of both energy and performance. EfficientNet-B0 is similarly efficient with a value of 0.2788 J/image, confirming the advantages of modern, compact architectures in the inference phase.

Energy consumption increases in more comprehensive architectures. ResNet50 consumed a total of 2067.78 J of energy and 0.5268 J per image. DenseNet121, on the other hand, incurred a higher cost at 0.6688 J/image. Both models require higher average power during inference due to their structural complexity, yet their FPS values lag behind MobileNetV2 and EfficientNet-B0.

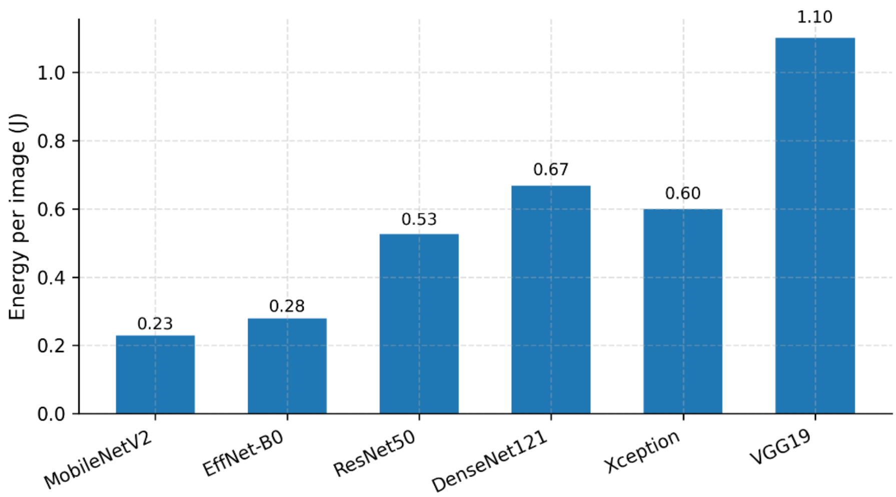

Figure 2. Inference energy cost per image across models.

Despite its separated convolution structure, Xception exhibited a highly intensive computational character with an average power of 72.18 W and consumed 0.6004 J/image. Although its FPS value (120.21) is high, its energy cost is significantly higher than that of lightweight architectures. VGG19 had by far the highest energy consumption in inference, as it did in training. The model's total energy consumption was 4325.52 J, which is approximately 2–5 times higher than all other architectures at 1.1020 J/image. Additionally, the low FPS value (97.99) indicates that the model exhibits a slow and costly inference profile.

The results show that lightweight, modern, and optimized architectures are by far more energy-efficient during inference. Deep and older architectures dramatically increase inference energy costs and are at a disadvantage in sustainable applications, despite achieving similar accuracy levels. This study comprehensively analyzes the energy consumption of six different deep learning architectures during training and inference processes. The findings clearly demonstrate that architectural design principles are decisive for both total energy consumption and energy cost per image. Lightweight architectures such as MobileNetV2 and EfficientNet-B0 stood out with the lowest energy consumption in both training and inference phases; in contrast, deeper or computationally intensive models such as VGG19 and Xception resulted in significant energy consumption. These results reveal that architectural approaches aimed at reducing computational complexity play a critical role in environmental sustainability. The results indicate that higher accuracy does not necessarily imply higher energy efficiency, as lightweight architectures achieve comparable accuracy with substantially lower energy cost.

Comparisons during the training phase show that network depth and parameter count directly affect energy consumption. Notably, the total energy consumption measured during the training of VGG19 was orders of magnitude higher than all other models, confirming that classical architectures can be energy inefficient on modern hardware. In contrast, MobileNetV2 and EfficientNet-B0 provided significant advantages in terms of both time and energy consumption thanks to their lower number of parameters and optimized computation blocks. This aligns with findings frequently emphasized in current research that “lightweight models are critical components for sustainable AI.”

When the inference stage was examined, a similar trend was observed in terms of energy efficiency. MobilenetV2 and EfficientNet-B0 achieved the highest FPS values, emerging as the most suitable architectures not only in terms of energy but also for applications requiring real-time processing. Interestingly, ResNet50, which had moderate energy consumption during training, was found to offer quite efficient FPS values during the inference stage; this may be a result of its good parallelizability despite the depth of the architecture. On the other hand, VGG19 lagged far behind modern architectures in terms of energy efficiency in both training and inference processes, demonstrating that it is not a sustainable option for practical use. Figure 3 visualizes the trade-off between inference speed (FPS) and average power consumption, showing that lightweight architectures achieve high throughput with low power usage.

When all findings are evaluated together, it is seen that not only model size but also the types of convolutions used, block configurations, parameter efficiency, and the load distribution created by the computation flow on the GPU are effective on energy consumption. Mobile-focused design principles, such as depthwise separable convolution and inverted residual blocks, dramatically reduce energy costs. These results emphasize that when developing AI systems operating in energy-constrained environments in the real world, the energy-performance balance must be considered as much as accuracy in model selection.

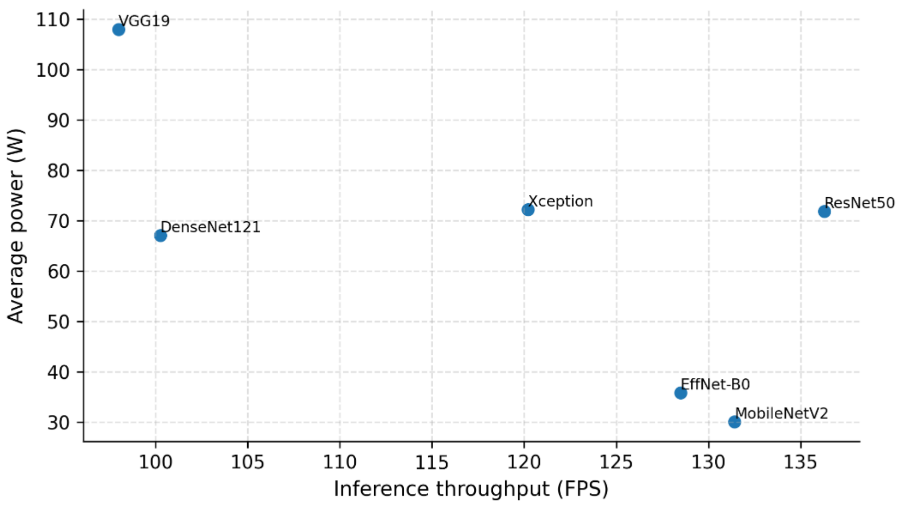

Figure 3. Throughput–Power trade-off (fps vs average power) for all models.

Furthermore, this study has revealed that the training and inference stages can exhibit completely different profiles in terms of energy consumption. Some models consume high energy during training but can be relatively efficient during inference. Therefore, evaluating energy costs based solely on one stage is insufficient; the total energy consumed by the model throughout its life cycle must be taken into account.

In conclusion, this study makes a meaningful contribution to the literature by systematically measuring energy consumption and providing a comprehensive comparison between modern and classical architectures. The findings show that architectural choices are critical for developing sustainable artificial intelligence applications and that energy efficiency must now occupy a central place among performance metrics. In this context, considering the energy-performance balance, the MobileNetV2 and EfficientNet-B0 architectures stand out as the most suitable options, especially for resource-constrained devices and real-time applications.

4. Discussion and Conclusion

This study presents significant findings in the field of energy efficiency by systematically and comprehensively examining the energy consumption that arises during the training and inference processes of six different deep learning architectures. The results reveal that the model architecture has a decisive impact on energy performance; specifically, lightweight and optimized architectures such as MobileNetV2 and EfficientNet-B0 have been shown to achieve the lowest energy consumption in both training and inference phases, making them the most suitable options for sustainable artificial intelligence applications. In contrast, deeper and computationally intensive architectures like VGG19 and Xception, while offering competitive performance in terms of accuracy, fall short in meeting modern requirements for energy efficiency.

The findings show that energy cost is closely related not only to model size but also to the computational complexity of the architecture, layer design, and the capacity for task parallelization on the GPU. Therefore, accuracy alone is not sufficient when evaluating deep learning models; the energy-performance balance must now become a central design and selection criterion. Energy efficiency has become a critical factor in model architecture selection, especially with the proliferation of mobile devices, embedded systems, and real-time AI applications.

This study fills an important gap in the literature by performing energy measurements during both training and inference phases and providing a detailed comparison across different architectures. Future work could explore energy behaviors across different GPU architectures, the energy gains of quantization and model compression techniques, and new architectural design strategies that optimize energy efficiency in the context of the accuracy-performance trade-off. Furthermore, comparisons on large-scale datasets and multiple hardware configurations will provide more comprehensive contributions to sustainable artificial intelligence research.

Although Imagenette provides a controlled benchmark, further experiments on large-scale datasets such as ImageNet are required to assess the generalizability of the observed energy consumption trends.

Overall, this study strongly emphasizes that energy efficiency is an integral metric in the evaluation of artificial intelligence models and establishes an important foundation for the development of sustainable, low-environmental-impact deep learning systems.

Author Contributions

The percentages of the author' contributions are presented below. The author reviewed and approved the final version of the manuscript.

	Y.S.
C	100
D	100
S	100
DCP	100
DAI	100
L	100
W	100
CR	100
SR	100
PM	100
FA	100

C= concept, D= design, S= supervision, DCP= data collection and/or processing, DAI= data analysis and/or interpretation, L= literature search, W= writing, CR= critical review, SR= submission and revision, PM= project management, FA= funding acquisition.

Conflict of Interest

The author declared that there is no conflict of interest.

Ethical Consideration

Ethics committee approval was not required for this study because it did not involve human participants or animal subjects.

Acknowledgements

The author would like to thank the open-source contributors of the Imagenette dataset and the developers of Tensorflow and associated deep learning libraries used in this study. No additional administrative, technical, or material support was received.

References

Aquino-Brítez, S., García-Sánchez, P., Ortiz, A., & Aquino-Brítez, D. (2025). Towards an energy consumption index for deep learning models: A comparative analysis of architectures, GPUs, and measurement tools. *Sensors*, *25*(3), 846.

Bouza, L., Bugeau, A., & Lannelongue, L. (2023). How to estimate carbon footprint when training deep learning models? A guide and review. *Environmental Research Communications*, *5*(11), 115014.

Bozkurt, A. (2024). Stanford HAI yapay zekâ raporu incelemesi. *Bilgi Yönetimi*, *7*(2), 445–457.

del Rey, S., Martínez-Fernández, S., Cruz, L., & Franch, X. (2023). Do DL models and training environments have an impact on energy consumption? *2023 49th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, 150–158.

Dey, S., Singh, A. K., Prasad, D. K., & McDonald-Maier, K. (2020). Temporal motionless analysis of video using CNN in MPSoC. *2020 IEEE 31st International Conference on Application-Specific Systems, Architectures and Processors (ASAP)*, 73–76.

Getzner, J., Charpentier, B., & Günnemann, S. (2023). Accuracy is not the only metric that matters: Estimating the energy consumption of deep learning models. *arXiv*. https://doi.org/10.48550/arXiv.2304.00897

Gowda, S. N., Hao, X., Li, G., Gowda, S. N., Jin, X., & Sevilla-Lara, L. (2024). Watt for what: Rethinking deep learning's energy-performance relationship. *European Conference on Computer Vision*, 388–405.

Howard, J., & Gugger, S. (2020). Fastai: A layered API for deep learning. *Information*, *11*(2), 108.

Ji, Z., & Jiang, M. (2026). A systematic review of electricity demand for large language models: Evaluations, challenges, and solutions. *Renewable and Sustainable Energy Reviews*, *225*, 116159.

Karamchandani, A., Mozo, A., Gómez-Canaval, S., & Pastor, A. (2024). A methodological framework for optimizing the energy consumption of deep neural networks: A case study of a cyber threat detector. *Neural Computing and Applications*, *36*(17), 10297–10338.

Latif, I., Newkirk, A. C., Carbone, M. R., Munir, A., Lin, Y., Koomey, J., Yu, X., & Dong, Z. (2024). Empirical measurements of AI training power demand on a GPU-accelerated node. *arXiv*. https://doi.org/10.48550/arXiv.2412.08602

Li, C., Tsourdos, A., & Guo, W. (2022). A transistor operations model for deep learning energy consumption scaling law. *IEEE Transactions on Artificial Intelligence*, *5*(1), 192–204.

Mehlin, V., Schacht, S., & Lanquillon, C. (2023). Towards energy-efficient deep learning: An overview of energy-efficient approaches along the deep learning lifecycle. *arXiv*. https://doi.org/10.48550/arXiv.2303.01980

Qi, Y., Zhang, S., & Taha, T. M. (2022). TRIM: A design space exploration model for deep neural networks inference and training accelerators. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, *42*(5), 1648–1661.

Rodriguez, C., Degioanni, L., Kameni, L., Vidal, R., & Neglia, G. (2024). Evaluating the energy consumption of machine learning: Systematic literature review and experiments. *arXiv*. https://doi.org/10.48550/arXiv.2408.15128

Söyler, H. (2025). Gerze Meslek Yüksekokulu'nun karbon ayak izi hesaplaması, Monte Carlo analizi ve sürdürülebilirlik stratejileri. *Journal of Anatolian Environmental and Animal Sciences*, *10*(4), 328–337.

Tang, Z., Wang, Y., Wang, Q., & Chu, X. (2019). The impact of GPU DVFS on the energy and performance of deep learning: An empirical study. *Proceedings of the Tenth ACM International Conference on Future Energy Systems*, 315–325.

Tripp, C. E., Perr-Sauer, J., Gafur, J., Nag, A., Purkayastha, A., Zisman, S., & Bensen, E. A. (2024). Measuring the energy consumption and efficiency of deep neural networks: An empirical analysis and design recommendations. *arXiv*. https://doi.org/10.48550/arXiv.2403.08151

Tu, X., Mallik, A., Chen, D., Han, K., Altintas, O., Wang, H., & Xie, J. (2023). Unveiling energy efficiency in deep learning: Measurement, prediction, and scoring across edge devices. *Proceedings of the Eighth ACM/IEEE Symposium on Edge Computing*, 80–93.

Tuğaç, Ç. (2023). Birleşmiş Milletler sürdürülebilir kalkınma amaçlarının gerçekleştirilmesinde yapay zekâ uygulamalarının rolü. *Sayıştay Dergisi*, *34*(128), 1–25.

Wang, F., Zhang, W., Lai, S., Hao, M., & Wang, Z. (2021). Dynamic GPU energy optimization for machine learning training workloads. *IEEE Transactions on Parallel and Distributed Systems*, *33*(11), 2943–2954.

Wang, Y., Hao, M., He, H., Zhang, W., Tang, Q., Sun, X., & Wang, Z. (2024). Drlcap: Runtime GPU frequency capping with deep reinforcement learning. *IEEE Transactions on Sustainable Computing*, *9*(5), 712–726.

Xu, Y., Martínez-Fernández, S., Martinez, M., & Franch, X. (2023). Energy efficiency of training neural network architectures: An empirical study. *arXiv*. https://doi.org/10.48550/arXiv.2302.00967