IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 7, April 2025
Impact Factor: 7.67

Energy-Efficient Deep Learning Models for Edge AI: A Green Computing Perspective

Ms. Nidhi[1], Ms. Maanvika[2], Mr. Rajesh A Rajgor[3]
Assistant Professor (Computer Science and Engineering)1-3
JECRC University, Jaipur, Sitapura, Vidhani, Rajasthan, India
nbarwar1992@gmail.com, choudharymaanvika@gmail.com, raaj.rajgor1808@gmail.com

Abstract: *The exponential growth of artificial intelligence (AI) and deep learning applications has led to significant computational demands and energy consumption, particularly in centralized cloud-based systems. As the global focus shifts toward sustainable and environmentally responsible computing, Edge AI emerges as a compelling solution by enabling on-device intelligence close to data sources. However, deploying deep learning models on resource-constrained edge devices introduces critical challenges related to computational efficiency, power consumption, and real-time performance. This paper presents a comprehensive analysis of energy-efficient deep learning architectures tailored for Edge AI systems, emphasizing green computing principles. We explore lightweight model designs, such as MobileNet, TinyML, SqueezeNet, and quantized neural networks, and propose a novel hybrid model that balances accuracy, latency, and energy usage. Our approach includes pruning, quantization, knowledge distillation, and hardware-aware neural architecture search (NAS) techniques to optimize model deployment on edge devices such as Raspberry Pi, NVIDIA Jetson Nano, and ESP32.*

The research further evaluates the proposed models through empirical analysis across benchmark datasets (e.g., CIFAR-10, ImageNet) using real-time power monitoring tools. The results show a significant reduction in energy consumption (up to 50%) while maintaining competitive inference accuracy. Additionally, we provide insights into the trade-offs between model complexity and sustainability, making a case for responsible AI deployment in smart cities, healthcare wearables, and IoT systems. This paper contributes to the growing body of work in green AI and aims to set a foundation for future energy-aware intelligent systems

Keywords: Edge AI, Energy-Efficient Deep Learning, Green Computing, Neural Architecture Search, Sustainable AI, Low-Power Inference

DOI: 10.48175/IJARSCT-25480