Empowering our Critters: Running Energy Efficient Deep Learning Models for On-Edge Bioacoustic Monitoring

Paritosh Borkar
Rochester Institute of Technology
paritoshpborkar@gmail.com

Rishabh Malviya
IIT, Bombay
rishabh.malviya@gmail.com

Jayant Sachdev
Cornell University
js2866@cornell.edu

Abstract

We propose a multi-phase plan to study the feasibility and benefits of deploying energy-efficient deep learning models on edge devices using passive acoustic monitoring (PAM) for climate change mitigation and adaptation. Current approaches involve collecting audio recordings for long periods and then automatically annotating them with bioacoustic models at a central location; we make the case for running these bioacoustic models on-edge to significantly speed up the operational efficiency of PAM. Successful deployments would open up several possibilities, including early warning systems for climate disasters, frequent monitoring of a region's resistance to extreme climate events, and real-time monitoring of the effects of human activities on ecosystems.

1 Introduction

Passive bioacoustic monitoring (PAM) has emerged as a transformative non-invasive technique for ecological research, enabling scientists to systematically record and analyze animal vocalizations using autonomous acoustic sensors to identify species, assess population dynamics, and monitor ecosystem health. In the past decade, PAM has been successfully used to monitor species of owls[12], seabirds [2], bats [5], beaked whales [20], frogs[14] and elephants[25]. However, the current PAM workflow often requires deploying recording units in remote locations for days or weeks, after which these units are physically retrieved to centralized computing facilities for processing. This offline processing paradigm introduces delays of weeks between data collection and actionable insights, limiting the utility of this technique[22]. On-edge bioacoustic models offer a more efficient and sustainable solution to these operational bottlenecks. They enable frequent processing of captured audio directly at the recording site, reducing the need for physical device retrieval and centralized data processing. The processed results and relevant audio segments can be transmitted over a network to researchers. This approach fundamentally restructures the PAM workflow, reducing the weeks long delays to near real-time insights, as illustrated in Figure 1. In addition to reducing these operational bottlenecks, on-edge processing involves light-weight models and optimizations, which would result in reduced energy usage and thus reduced carbon emissions. This is especially important as PAM scales up and becomes more widely deployed [21].

We measure energy consumption during inference using SPEC-approved power measurement devices for direct instrumentation. For CUDA containers, we estimate the energy consumption of these models by first measuring the total floating point operations(FLOPs) calculated during inference at full throughput. Then using the power efficiency of ML models on different GPU architectures[6] we get our final estimations. To evaluate these models, we look at MLPerf[23], a benchmarking framework for deep learning models, to create energy consumption profiles for these models.

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2025.

1.1 Bioacoustic Monitoring and Climate Change

Bioacoustic monitoring isn't just useful for the conservation efforts and tracking of specific animal species. It also has important applications in monitoring climate change and predicting its impact on humans. The biodiversity of an ecosystem serves as a proxy for its resistance to climate perturbations [11], often over multiple decades [16]. Studies show that monitoring bioacoustics can serve as an early warning of climate extreme events such as typhoons[18], tracking invasive species[12] and changes in animal behavior from human logging[25]. Bioacoustics can also play a role in tracking biodiversity recovery in tropical forests[15]. This is important since there have been instances of planting large monocultures of eucalyptus trees for reforestation that have led to increased droughts and forest fires[3] [7] [19]. The real-time monitoring capabilities enabled by on-edge PAM systems represent a crucial tool for tracking reforestation efforts and detecting early warning signals of climate extremes and ecosystem destabilization before these processes reach irreversible tipping points.

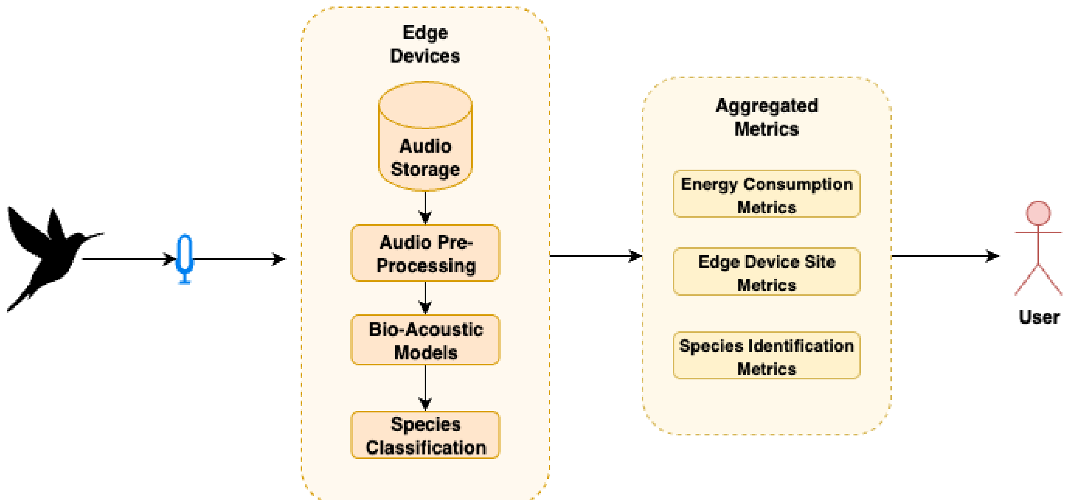

Figure 1: What an inference pipeline for bioacoustic monitoring with on-edge models might look like. Acoustic sensors will feed raw audio into edge devices running optimized models; only lightweight summaries, small audio clips, and alerts will be sent to the cloud for aggregation and climate accounting.

2 Methodology

We will profile state-of-the-art(SOTA) bioacoustic models (CNN + Spectrogram approaches [21] and NatureLM-audio [17]), a foundational audio language model, on edge devices(such as NVIDIA Jetson boards) and CUDA containers. We use the following MLPerf Power [23] metrics for evaluation, as they remain relevant even when systems scale.

- Energy per inference: For cases without direct instrumentation, we can infer the energy consumption by calculating total FLOPs during inference for an edge device or CUDA container and estimate using Giga FLOPs(GFLOPS) per watt of the system at full throughput[6].
- Samples per joule: This is the number of inferences that the system can make at full throughput per joule of energy. It is an indication of the energy efficiency of the edge device.
- Performance-constrained measurement: We compare the performance of a model running on an edge device to its performance running on a cloud server to ensure that there is no severe degradation.

We will measure the performance of these models by training on BirdCLEF 2025 [9], an audio soundscape dataset that spans different taxonomic groups, to quantify both the technical feasibility and the sustainability gains of running these models on edge. Since this dataset includes recordings from a rich ecosystem of threatened and endemic species, it serves as a strong test bed for studying ecosystems vulnerable to climate extremes or industrial impacts.

2.1 Phase 1: Estimating Energy Consumption for a Single Inference

We begin by estimating the energy consumption of a single inference on CUDA containers. We profile a ResNet50 model[8] and the first place solution for BirdCLEF[1](using Noisy Student training[26]), which are trained on mel spectrograms extracted from smaller chunks of the audio. We will measure FLOPs and GFLOPs per watt measurements [6] to estimate the energy consumption for multiple inferences on CUDA containers and average the results. We look at the total energy consumption of running inferences on a batch of audio inputs. Assuming 30 second audio chunks, we would process 2880 audio inputs in a 24 hour period. The results would include identified species and their confidence, timestamps, key audio clips, and location metadata.

2.2 Phase 2: Measuring Energy Consumption for NatureLM-audio

We run the same inferences on a fine-tuned NatureLM-audio foundational model [17] that uses a Llama 3.1-8B LLM under the hood. we run the model on an edge device (NVIDIA Jetson board) and use direct instrumentation to get accurate readings of the energy consumption of the entire system, which also includes peripherals and data transmission costs. For the fine-tuning process, we take an audio-text pair as input, and encode the audio input with a BEATs encoder [4]. The embeddings are then connected to the LLM using a Q-Former, and the LLM is finetuned with LoRa[10]. For 50% of the audioprompt pairs, the correct species label is intentionally omitted. Under 4-bit quantization, the Llama 3.1-8B model requires a minimum of 6 GB VRAM, whic is sufficient to deploy on NVIDIA Jetson.

2.3 Limitations and Potential Topics for Further Research

The energy costs of training and deploying bioacoustic models are beyond the scope of this proposal. We also ignore the effects of temperature and humidity on energy efficiency. Prior work has also explored solar-powered recorders as energy sources for powering edge devices[13], but such considerations remain outside our scope.

3 Future Outlook and Impact

To estimate the impact of successfully deploying deep learning models and running inferences on edge, we take a look at some existing projects that would benefit from this initiative. A 2023 study on the ecological response to typhoons[18] utilized recording audio soundscapes for 30 days before and after a typhoon impact to measure the difference in bird calls. Part of the study involved using machine learning methods to identify bird species after the recordings had been collected. An application of this proposal would be to run these bird classification models at recording sites and prepare early for a potential typhoon impact. In 2020, the U.S. Department of Agriculture's Forest Service published a study on the early detection of invasive Barred Owl species[12][24] using bioacoustics. The study involved the deployment of recording devices across $\sim$400 sites and the collection of audio recordings for 5 to 7 nights at a time over a 5-month period, redeploying the devices after each survey. Even running rudimentary inferences at these sites would reduce the wait time for researchers to analyze the results and helped them make better informed decisions about placing devices at new sites and tracking the behavior of invasive species. If successful, these on-edge monitoring networks can operate continuously and autonomously, providing the scientific foundation necessary for effective climate adaptation and ecosystem preservation strategies in an era of rapid climate change.

References

[1] Nikita Babych. 1st place solution: Multi-iterative noisy student is all you need, 2025.

[2] Abraham L. Borker, Michael W. McKown, Joshua T. Ackerman, John Y. Takekawa, and et al. Spatial and temporal patterns in the acoustic activity of a nocturnal seabird colony. *Urban Naturalist*, 1:1–14, 2014.

[3] Brown, Sarah. Eucalyptus expansion worsens droughts and fires in Brazils Cerrado, conservationists say. *Mongabay*, 2024.

[4] Sanyuan Chen, Yu Wu, Chengyi Wang, Shujie Liu, Daniel Tompkins, Zhuo Chen, Wanxiang Che, Xiangzhan Yu, and Furu Wei. BEATs: Audio pre-training with acoustic tokenizers. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 5178–5193. PMLR, 23–29 Jul 2023.

[5] Laci S. Coleman, W. Mark Ford, Chris A. Dobony, and Eric R. Britzke. A comparison of passive and active acoustic sampling for a bat community impacted by white-nose syndrome. *Journal of Fish and Wildlife Management*, 5(2):217–226, 2014.

[6] Radosvet Desislavov, Fernando Martínez-Plumed, and José Hernández-Orallo. Trends in ai inference energy consumption: Beyond the performance-vs-parameter laws of deep learning. *Sustainable Computing: Informatics and Systems*, 38:100857, April 2023.

[7] Sandra Goded, Johan Ekroos, Jesús Domínguez, Joaquín G. Azcárate, José A. Guitián, and Henrik G. Smith. Effects of eucalyptus plantations on avian and herb species richness and composition in north-west spain. *Global Ecology and Conservation*, 19:e00690, 2019.

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

[9] Maggie Demkin Sohier Dane Stefan Kahl Holger Klinck, Juan Sebastián Cañas and Tom Denton. Birdclef+ 2025, 2025.

[10] Edward J Hu, yelong shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022.

[11] Isbell, Forest and Craven, Dylan and Connolly, John and Loreau, Michel and Schmid, Bernhard and Beierkuhnlein, Carl and Bezemer, T. Martijn and Bonin, Catherine and Bruelheide, Helge and de Luca, Enrica and Ebeling, Anne and Griffin, John N. and Guo, Qinfeng and Hautier, Yann and Hector, Andy and Jentsch, Anke and Kreyling, Jürgen and Lanta, Vojtch and Manning, Pete and Meyer, Sebastian T. and Mori, Akira S. and Naeem, Shahid and Niklaus, Pascal A. and Polley, H. Wayne and Reich, Peter B. and Roscher, Christiane and Seabloom, Eric W. and Smith, Melinda D. and Thakur, Madhav P. and Tilman, David and Tracy, Benjamin F. and van der Putten, Wim H. and van Ruijven, Jasper and Weigelt, Alexandra and Weisser, Wolfgang W. and Wilsey, Brian and Eisenhauer, Nico. Biodiversity increases the resistance of ecosystem productivity to climate extremes. *Nature*, 10 2015.

[12] Damon B. Lesmeister and Julianna M. A. Jenkins. Integrating new technologies to broaden the scope of northern spotted owl monitoring and linkage with usda forest inventory data. *Frontiers in Forests and Global Change*, Volume 5 - 2022, 2022.

[13] Vincent Lostanlen, Antoine Bernabeu, Jean-Luc Béchennec, Mikaël Briday, Sébastien Faucou, and Mathieu Lagrange. Energy efficiency is not enough: Towards a batteryless internet of sounds. In *Proceedings of the International Workshop on the Internet of Sounds (IWIS), Audio Mostly*, Trento, Italy, September 2021.

[14] John Measey, Ben C. Stevenson, Tamara Scott, and etc. Counting chirps: acoustic monitoring of cryptic frogs. *Journal of Applied Ecology*, 54(3):894–902, 2017.

[15] Müller, Jörg and Mitesser, Oliver and Schaefer, H. Martin and Seibold, Sebastian and Busse, Annika and Kriegel, Peter and Rabl, Dominik and Gelis, Rudy and Arteaga, Alejandro and Freile, Juan and Leite, Gabriel Augusto and de Melo, Tomaz Nascimento and LeBien, Jack and Campos-Cerqueira, Marconi and Blüthgen, Nico and Tremlett, Constance J. and Böttger, Dennis and Feldhaar, Heike and Grella, Nina and Falconí-López, Ana and Donoso, David A. and Moriniere, Jerome and Buivalová, Zuzana. Soundscapes and deep learning enable tracking biodiversity recovery in tropical forests. *Nature Communications*, 14, 10 2023.

[16] Dong X. Oliveira BF, Moore FC. Biodiversity mediates ecosystem sensitivity to climate variability. *Commun Biol.*, 06 2022.

[17] David Robinson, Marius Miron, Masato Hagiwara, Benno Weck, Sara Keen, Milad Alizadeh, Gagan Narula, Matthieu Geist, and Olivier Pietquin. Naturelm-audio: an audio-language foundation model for bioacoustics, 2025.

[18] Samuel R. P.-J. Ross, Nicholas R. Friedman, Kenneth L. Dudley, Takuma Yoshida, Masashi Yoshimura, Evan P. Economo, David W. Armitage, and Ian Donohue. Divergent ecological responses to typhoon disturbance revealed via landscape-scale acoustic monitoring. *Global Change Biology*, 30(1):e17067, 2024. e17067 GCB-23-1585.R1.

[19] Unai Sertutxa, Unai Ortega-Barrueta, Ibone Ametzaga-Arregi, and Lorena Peña. The ecological cost of eucalypt plantations for bird communities. *Forest Ecology and Management*, 594:122961, 2025.

[20] Joy E. Stanistreet, Douglas P. Nowacek, Simone Baumann-Pickering, Joel T. Bell, Danielle M. Cholewiak, John A. Hildebrand, Lynne E.W. Hodge, Hilary B. Moors-Murphy, Sofie M. Van Parijs, and Andrew J. Read. Using passive acoustic monitoring to document the distribution of beaked whale species in the western north atlantic ocean. *Canadian Journal of Fisheries and Aquatic Sciences*, 74(12):2098–2109, 2017.

[21] Dan Stowell. Computational bioacoustics with deep learning: a review and roadmap, 2022. Published 2022 Mar 21.

[22] Larissa S. M. Sugai, Thiago S. F. Silva, João W. Ribeiro, and Diego Llusia. Terrestrial passive acoustic monitoring: Review and perspectives. *BioScience*, 69(1):15–25, 2019.

[23] Arya Tschand, Arun Tejusve Raghunath Rajan, Sachin Idgunji, Anirban Ghosh, Jeremy Holleman, Csaba Kiraly, Pawan Ambalkar, Ritika Borkar, Ramesh Chukka, Trevor Cockrell, Oliver Curtis, Grigori Fursin, Miro Hodak, Hiwot Kassa, Anton Lokhmotov, Dejan Miskovic, Yuechao Pan, Manu Prasad Manmathan, Liz Raymond, Tom St. John, Arjun Suresh, Rowan Taubitz, Sean Zhan, Scott Wasson, David Kanter, and Vijay Janapa Reddi. Mlperf power: Benchmarking the energy efficiency of machine learning systems from microwatts to megawatts for sustainable ai, 2025.

[24] Connor M Wood, R J Gutiérrez, John J Keane, and M Zachariah Peery. Early detection of rapid barred owl population growth within the range of the california spotted owl advises the precautionary principle. *The Condor*, 122(1), jan 2020.

[25] Peter H. Wrege, Franck B. Bambi, P. J. F. Malonga, Olivier J. Samba, and Terry Brncic. Early detection of human impacts using acoustic monitoring: An example with forest elephants. *PLOS ONE*, 19(7):e0306932, 2024.

[26] Qizhe Xie, Minh-Thang Luong, Eduard Hovy, and Quoc V. Le. Self-training with noisy student improves imagenet classification, 2020.