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

Paritosh Borkar, Rishabh Malviya, Jayant Sachdev

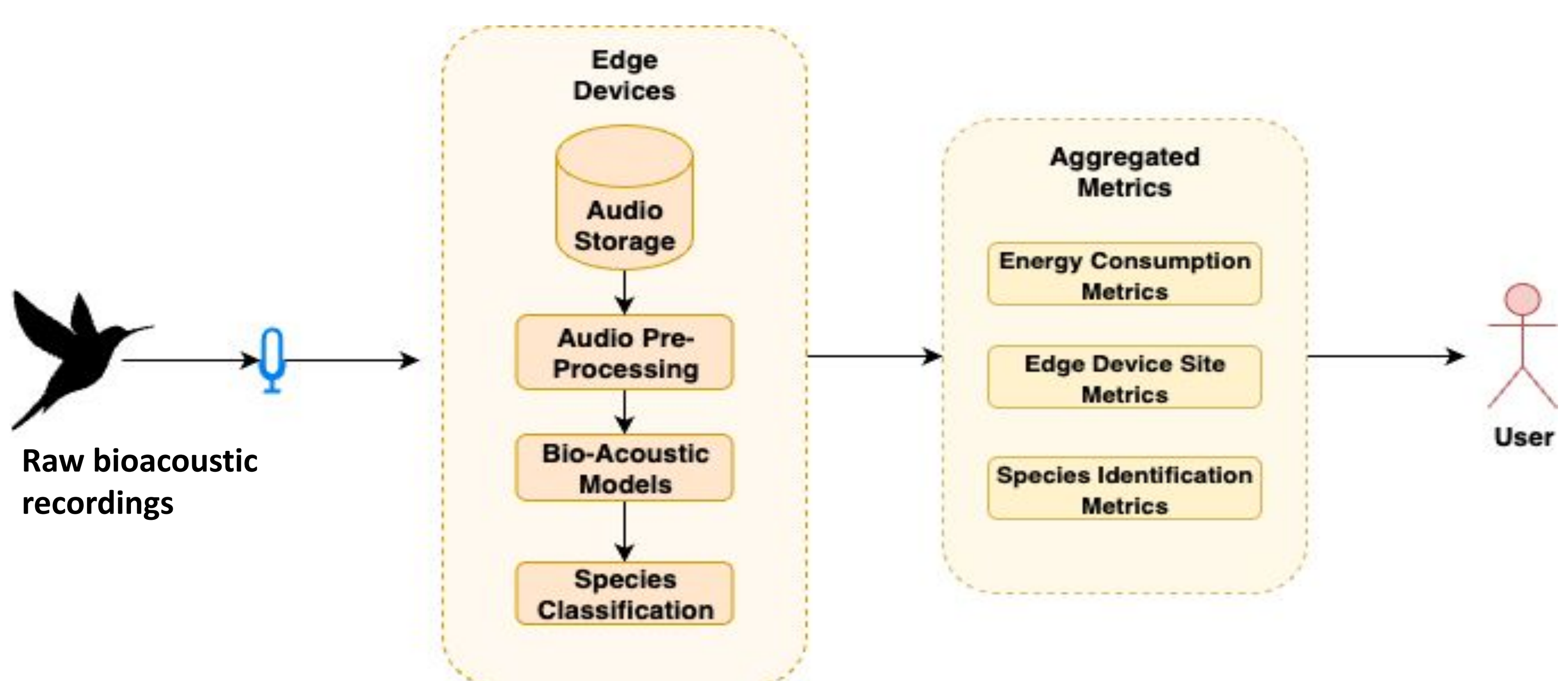
NEURAL INFORMATION PROCESSING SYSTEMS

Introduction

Ecosystems are vulnerable to climate extremes and shifts in historic weather patterns. Bioacoustics can play a role in studying the effects of climate change on these ecosystems. We propose a plan to measure the energy efficiency of deploying deep learning models on edge devices, running parallel to recorders collecting bioacoustic data, to make inferences faster and allow climate scientists and conservationists to respond quickly to climate extremes.

Role of Bioacoustic Monitoring for Climate Change

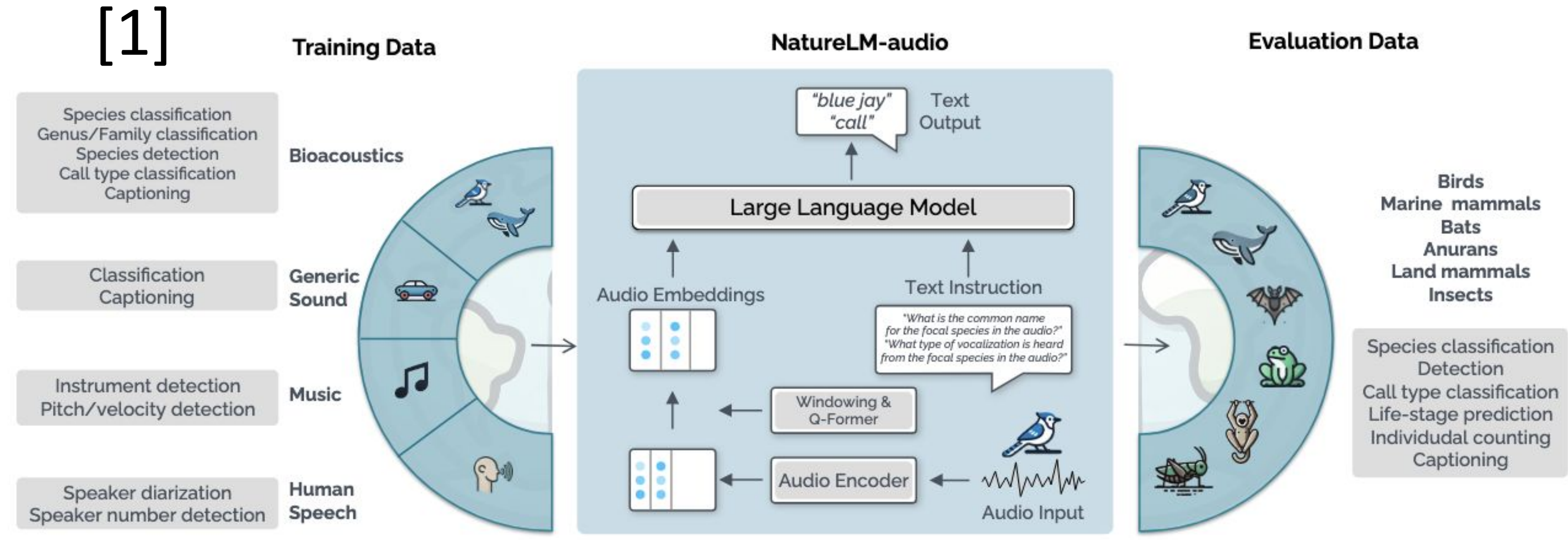
- Biodiversity can act as an indicator of an ecosystem's resilience to climate disturbances, and bioacoustic data can provide early warnings of extreme events like typhoons, spread of invasive species, and behavioral changes in animals from human activity.
- Bioacoustics can be used to study divergent response in birds to climate extremes like typhoons, assess biodiversity recovery after reforestation and detect nearby huan logging activities.
- Real-time, on-edge passive acoustic monitoring (PAM) systems can thus offer a vital tool for tracking ecosystem health and anticipating climate-related tipping points.



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.

Methodology

[1]

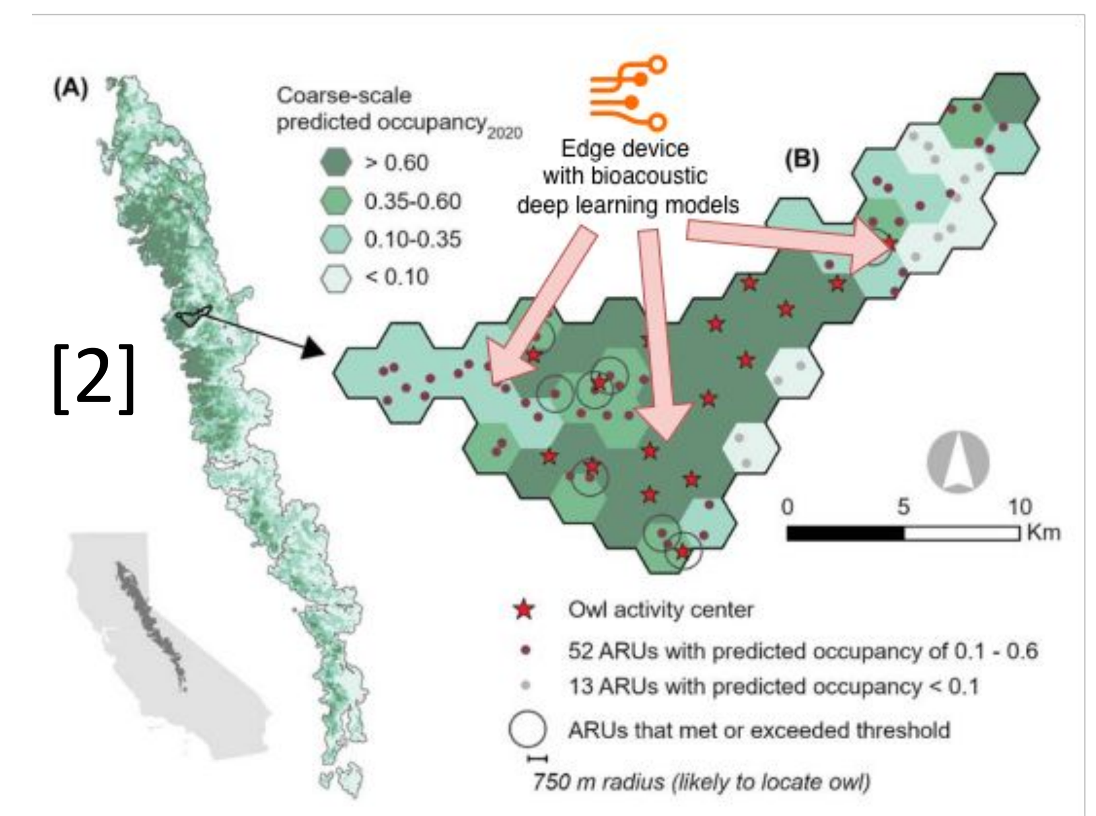


[1] Overview of NatureLM-audio. To measure the energy consumption metrics of this model we will fine-tune it using the BirdClef 2025 dataset, a bioacoustic dataset of labelled and unlabelled audio soundscapes spanning different families of species of animals.

- Using MLPerf Power metrics[3], we will evaluate the energy consumption of deep learning bioacoustic models during inference. These metrics stay relevant even as systems scale We will use the following metrics:
  - **Energy per inference:** Whenever direct instrumentation is not feasible we will measure the estimated energy consumed using total FLOPs and system GFLOPS per watt at full throughput.
  - **Samples per joule:** Measures the number of inferences per joule at full throughput, indicating device energy efficiency.
  - **Performance-constrained measurement:** Compares edge device versus cloud compute performance to assess performance degradation.
- We aim to benchmark baseline models using CNN + spectrogram architectures and state-of-the-art bioacoustic models, like NatureLM-audio on edge devices like NVIDIA Jetson boards and CUDA containers.
- Models will be trained and fine-tuned on the BirdCLEF 2025 dataset to assess technical feasibility and sustainability benefits for biodiversity monitoring.
- Phase 1 will estimate the energy use for a baseline ResNet50 and the first place position model for the BirdCLEF 2025 competition, by estimating FLOPs and power efficiency during inference. The architecture uses a self training with noisy student approach.
- Phase 2 will profile NatureLM-audio, a fine-tuned large audio-language model based on Llama 3.1-8B, using direct power instrumentation on edge hardware. For the fine-tuning process, we will take an audio-text pair as input, and encode the audio input with a BEATs encoder. The embeddings will then connected to the LLM using a Q-Former, and the LLM would then be fine-tuned with Low-Rank Adaptation(LoRa).

Pathway to Climate Impact

[2]



[2] Autonomous recording unit (ARU) study area on the Eldorado National Forest, California, USA, 2020 for detection of invasive spotted owl species. Edge devices can be deployed alongside the ARUs for real time or daily inferences, in contrast to 3 weeks of data collection per section.

- Deploying deep learning models for on-edge inference can significantly enhance ecological monitoring. As an illustration in the image above, detection of invasive owl detection currently relies on collecting audio data and processing it later at central locations.
- Running species classification models directly at recording sites would enable faster analysis, early warnings of climate extremes like typhoons, and more adaptive fieldwork planning.
- Successful implementation of such autonomous, continuous monitoring networks could accelerate research and provide critical real-time data for climate change adaptation and ecosystem preservation.

Limitations and Scope

- The energy costs of training and deploying bioacoustic models are beyond the scope of this proposal.
- The effects of temperature and humidity on energy efficiency are ignored.
- The energy cost of data transmissions are outside the scope of the proposal, although estimates were made to validate the feasibility of implementing such deployments.
- A successful deployment of trained models on edge devices would require periodic burst data transmissions of inference results using a satellite internet connection(high power consumption) or a low power long range radio transmitter(low bandwidth).

References

[1] 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.

[2] 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, pg. 16.

[3] Arya Tschand, Arun Tejusve Raghunath Rajan, Sachin Idgunji, et al. MlPerf power: Benchmarking the energy efficiency of machine learning systems from microwatts to megawatts for sustainable ai, 2025.