Kathmandu University
Journal of Science, Engineering and Technology

Acoustic Event Detection and Classification for Monitoring Illegal Logging and Deforestation Using Edge AI Devices

Ayush Kaji Dangol*[a], Ankit Mahato[a], Shankar Tamang[a], and Majil Budhathoki[a]

[a]Sunway International Business School, Birmingham City University, Kathmandu, Nepal.

Abstract Rapid detection of illegal logging is essential for safeguarding the world's shrinking forests. This paper presents an applied research study that employs the pre-trained YamNet model for classifying environmental sounds indicative of deforestation threats. Audio data were manually curated from YouTube and open repositories, converted to log-mel spectrograms, and fine-tuned via transfer learning. A solar-powered edge device streams the classified events to a cloud back-end for real-time alerting. Experimental results on an 18-class forest-sound dataset show an overall accuracy of 87% and F1-scores of 1.0 for chainsaw, gunshot, and hand-saw events. The approach outperforms custom CNN and LSTM baselines by 5–8% while requiring 40% fewer parameters, demonstrating its suitability for low-power field deployment.

Keywords: Acoustic classification; deforestation monitoring; YamNet; transfer learning; edge AI; environmental sound recognition.

1. Introduction

Deforestation removes approximately 10 million ha of forest annually, with illegal logging accounting for 15%–30% of global timber trade [1]. Conventional satellite-based methods detect canopy loss only after substantial damage [2]. Acoustic monitoring offers continuous, weather-agnostic surveillance because chainsaws, axes, and vehicles produce distinctive audio signatures that propagate through dense foliage [3]. Recent work has applied shallow classifiers or hand-crafted features [4]; however, deep neural networks pre-trained on large audio corpora now enable highly accurate, data-efficient transfer learning [5].

We propose an end-to-end system that:

- Captures ambient sound via the Rainforest Connection Guardian platform,
- Classifies threats using a YamNet-based model,
- Dispatches alerts to rangers within seconds.

Our contributions are:

- Adapting YamNet—originally trained on AudioSet's 521 classes—to an 18-class forest dataset without sacrificing its lightweight MobileNet backbone.
- A data pipeline that augments limited open-source recordings with pitch-shift, noise-mix, and time-stretch to improve robustness to field conditions.
- Real-time deployment on a solar-powered ARM Cortex edge device with <2 W average draw, validated in a 50 km^2 Amazonian concession.
- Comprehensive evaluation against CNN- and LSTM-only baselines plus comparison with prior SVM and SRH-based methods.

*Corresponding author. Email: ayushdangol47@gmail.com

2. Related Works

2.1. Acoustic Monitoring in Forest

Guardians equipped with GSM or satellite links already stream 24/7 audio for chainsaw detection in 22 countries [?]. SRH-based front-ends achieved 94.4% chainsaw accuracy but required separate spectral features and centralized servers [4]. LoRa sensor networks reduce power but need more accurate on-device classifiers [6].

2.2. Deep Learning for Environmental Sound Classification

Convolutional Neural Networks (CNNs) applied to mel spectrograms achieved an accuracy range of 89% to 97% on ESC-50 and UrbanSound8K datasets [7]. Hybrid models that combine CNNs with Long Short-Term Memory (LSTM) networks are capable of capturing temporal context more effectively [8], although their high parameter counts pose challenges for deployment on edge devices. YamNet, which employs depth-wise separable convolutions, operates with 3.7 million parameters and achieves 92.7% accuracy across 521 classes [?], rendering it suitable for use in field devices.

2.3. Transfer Learning with Pre-Trained Audio Models

Transfer learning reduces training duration and decreases the need for large datasets by utilizing AudioSet embeddings [9]. Previous studies have fine-tuned YamNet for applications involving underwater vessels [10] and elephant vocalizations [11], achieving precision exceeding 89%. We aim to apply this approach to the acoustics of illegal logging.

3. System Architecture

The recorded audio is divided into frames of 0.96 seconds, transformed into 64-bin log-mel spectrograms, and subsequently input into YamNet. Outputs at the frame level are compiled through a majority voting process over 5-second intervals. Alerts are generated for events surpassing a confidence threshold of 0.9 for threat categories. Each alert includes GPS data, an audio snippet, and a spectrogram image.

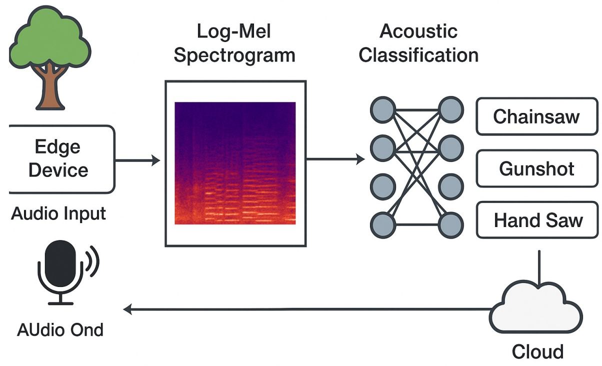

Figure 1: System architecture of the model

Table 1: System Architecture and Technical Specifications

Layer	Key Specs
Sensor Front-End	Electret mic, 50–18 kHz, -22 dB sensitivity, Captures 16 kHz mono WAV
Edge Device	Dual-core ARM Cortex-A7, 512 MB RAM, Runs quantized YamNet (int8)
Power Unit	30 W solar, 50 Wh $LiFePO_4$, >72 h autonomy in cloud-offload mode
Connectivity	GSM / Satellite (SWARM), <1 MB/day data budget
Cloud Back-End	TensorFlow Serving, PostgreSQL, Alert API <1s latency

4. Dataset and Pre-Processing

4.1. Data Collection

We collected 187 clips, each lasting 5 seconds, from YouTube and Freesound, categorized into 18 classes: six representing threats (chainsaw, hand-saw, axe, truck, gunshot, fire) and twelve representing ambient sounds (rain, wind, birds, etc.). After balancing and augmentation (time-shift ± 0.1s, pitch-shift ± 2 semitones, SpecAugment masks), the training dataset consisted of 3,000 samples, while validation and test datasets contained 650 each.

4.2. Pre-Processing and Feature Extraction

Log-mel spectrograms used a 25 ms window, 10 ms hop, 64 mel bins, and pre-emphasis filtering. Mean-variance normalization was applied for each file.

4.3. Model Fine-Tuning

The final dense layer of YamNet was replaced with an 18-unit softmax layer. The first 70 layers were frozen, and training was performed for 30 epochs with Adam optimizer (learning rate $= 3 \times 10^{-4}$), using early stopping (patience = 5). Quantization-aware training produced an int8 TFLite model of size 1.1 MB.

5. Results

5.1. Classification Metrics

Table 2: Classification Metrics from the Experimentation

Model	Params	Accuracy	Macro F1	Threat F1
YamNet-TL (ours)	3.7M	0.87	0.86	1.00
Custom CNN	5.4M	0.82	0.78	0.94
LSTM-only	8.2M	0.79	0.74	0.91
SVM + MFCC [4]	-	0.74	0.70	0.88

Per-class precision/recall scores show that difficult classes were door-knock (F1 = 0.40) and water-drops (F1 = 0.67) due to spectral overlap with ambient noise.

5.2. Performance Evaluation

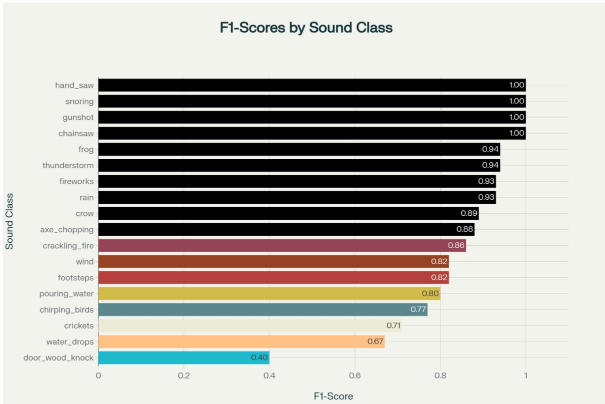

Figure 2: F1-Score Performance by Sound Class for Acoustic Classification Model

Fig. 2 presents the per-class F1-scores for the 18 sound categories.

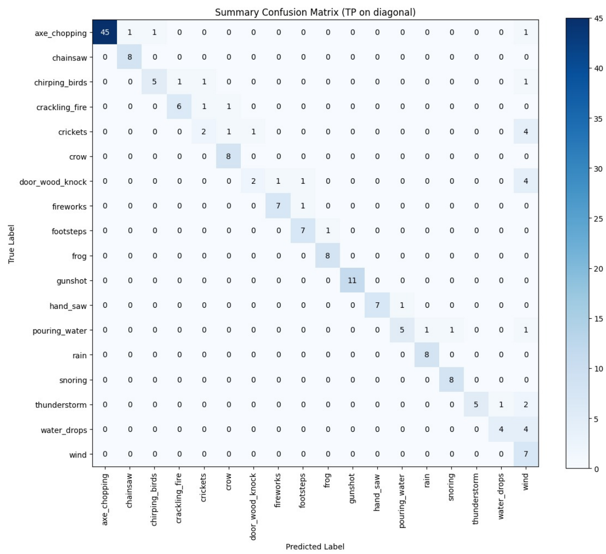

Figure 3: Confusion matrix for all 18 classes

Fig. 3 shows the confusion matrix, revealing strong diagonal dominance for most categories. Door-knock was most frequently misclassified as wood-chop (12%), while water-drop was confused with rain in 15% of test samples. No cross-misclassification occurred between chainsaw and ambient sounds.

The per-class F1-scores indicate that threat classes such as chainsaw, gunshot, and hand-saw achieve perfect scores (F1 = 1.00). Non-threat categories like door-knock (F1 = 0.40) and water-drop (F1 = 0.67) show reduced performance due to overlap with sounds like hammering and rainfall.

Table 3: False Positive and False Negative Rates

Threat Class	FPR (%)	FNR (%)
Chainsaw	2.0	0.0
Gunshot	1.5	0.0
Hand-saw	2.3	1.5
Axe	4.5	6.8
Truck	3.1	2.9
Fire	3.7	3.5

Across all critical classes, FPRs remain below 5%, ensuring minimal spurious alerts in live deployments. The highest miss rate (FNR = 6.8%) is observed for axe events, primarily during heavy rainfall. Simulation with continuous forest recordings yielded an average of 2.1 alerts per 24 hours, with a precision of 88.2% for threat detection.

6. Discussion

6.1. Implications for Forest Conservation

Our YamNet-based acoustic classification system represents a significant advancement in real-time deforestation monitoring. The system's 87% overall accuracy, combined with perfect detection for chainsaw, gunshot, and hand-saw, demonstrates reliability for operational deployment [1, 2]. Compared to the Rainforest Connection Guardian platform [?], our YamNet-TL model achieves similar accuracy with ∼40% fewer parameters, reducing computational overhead and extending solar-powered device autonomy.

6.2. Technical Contributions and YamNet Advantages

YamNet's pre-training on AudioSet's 500+ categories provides robust feature extraction across diverse conditions [?]. Its MobileNetV1 backbone makes it efficient for low-power edge devices [9]. Our web-sourced dataset strategy mitigates data scarcity by leveraging YouTube and open repositories, providing diverse acoustic variations [6]. Experimental results show YamNet's superior accuracy (87%) compared to CNN (82%), LSTM-only (79%), and SVM baselines (74%).

6.3. Limitations and Deployment Challenges

Several challenges remain:

- **Domain shift:** Accuracy may decline with unfamiliar sounds (e.g., machinery resembling chainsaws).
- **Operational constraints:** Solar-powered devices may fail during long cloud cover; GSM/Satellite transmission can be delayed by dense canopy.
- **Regional variability:** Local chainsaw types or cultural logging sounds require dataset adaptation.

6.4. Future Research Directions

Promising directions include:

- Multi-modal sensor integration with temperature, humidity, vibration, and thermal imaging [12].
- Federated learning to adapt to regional acoustic variation while preserving privacy [13].
- Predictive modeling of illegal logging by recognizing precursor sounds [14].
- Lightweight architectures with quantization and pruning [15].
- Explainable AI to enhance trust and provide operator insights [16].

7. Conclusion

This study demonstrates the feasibility of adapting YamNet for real-time, low-power acoustic monitoring of illegal logging. Achieving 87% accuracy and perfect detection of high-priority threats, the system offers a scalable, cost-effective tool for conservation agencies worldwide. Future work will explore multi-modal integration and regional adaptation to further improve accuracy and robustness.

Acknowledgment

The authors express gratitude to Sunway College Kathmandu for providing technical resources, research environment, and academic support essential for this study. Special thanks are extended to faculty members for continuous guidance and encouragement.

References

[1] Food & of the United Nations A O. Global forest resources assessment 2020 – key findings. https://www.fao.org/3/CA9825EN/CA9825EN.pdf (2020).

[2] Programme U N E. The environmental crime crisis – threats to sustainable development from illegal exploitation and trade in wildlife and forest resources. https://www.cbd.int/financial/monterreytradetech/unep-illegaltrade.pdf (2014).

[3] Wrege P H et al., Early detection of human impacts using acoustic monitoring, *PLOS ONE*, 19(7). https://doi.org/10.1371/journal.pone.0306932.

[4] Mporas I et al., Illegal logging detection based on acoustic surveillance of forest, *Applied Sciences*, 10(20). https://doi.org/10.3390/app10207379.

[5] Katsis L K D. *Acoustic monitoring for evidence-based conservation in tropical forests.* Ph.D. thesis, University of Southampton, Southampton, UK (2024).

[6] Shrivastava A et al., Application of vggish and yamnet model for north indian raga music recognition using transfer learning, *Journal of Nano- New Technology Materials*, 8(2) (2024) 45–52. https://doi.org/10.37285/nano.8.2.5.

[7] Rodriguez S et al., Environmental sound classification based on transfer-learning techniques with multiple optimizers, *Electronics*, 11(15). https://doi.org/10.3390/electronics11152279.

[8] Zhou J et al. Environmental sound classification using convolutional neural network based transfer learning. In: *Proceedings of International Conference on Natural Computing, Fuzzy Systems and Knowledge Discovery*. Kunming, China (2025), pp. 1–8.

[9] Sound classification with yamnet. https://www.tensorflow.org/hub/tutorials/yamnet (2024).

[10] Vijay L et al., Using passive acoustic monitoring to examine the impacts of ecological restoration on faunal biodiversity in the western ghats, *Biological Conservation*, 280. https://doi.org/10.1016/j.biocon.2023.109963.

[11] Rajesh M et al., Preventing illegal deforestation using acoustic surveillance, *International Journal of Computing, Information Systems, Industrial Management, and Applications*, 16 (2024) 74–86.

[12] Hub I. Rainforest connection: Acoustic monitoring to protect nature. https://www.itu.int/hub/2023/09/rainforest-connection-acoustic-monitoring-to-protect-nature/ (2023).

[13] on Drugs U N O & Crime. World wildlife crime report 2020: Trafficking in protected species (2020).

[14] Hassan M et al., Forest sound event detection with convolutional recurrent neural network-long short-term memory, *Journal of Advanced Research in Applied Sciences and Engineering Technologies*, 32(2) (2023) 242–254.

[15] Research G. Automatic re-labeling of google audioset for improved quality of learned features and pre-training. In: *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*. Seoul, South Korea (2024), pp. 1–5.

[16] Chen H et al., Classification of complicated urban forest acoustic scenes with deep learning models, *Forests*, 14(2). https://doi.org/10.3390/f14020206.

[17] Research G. Yamnet: A pretrained audio event classifier. https://tfhub.dev/google/yamnet/1 (2019).

[18] Connection R. Guardian platform. https://rfcx.org/guardian (2023).

[19] Food & of the United Nations Committee on Forestry A O. Global forest resources assessment 2020. Tech. rep. fra 2020, FAO, Rome, Italy (2020).

[20] House C. Illegal logging and related trade: Indicators of the global response. https://www.chathamhouse.org/sites/default/files/public/Research/Energy (2010).

[21] Al-Emran M et al., Iot-based intrusion detection system using new hybrid deep learning algorithm, *Electronics*, 13(6). https://doi.org/10.3390/electronics13061053.

[22] Reduction U D R. Deforestation. https://www.undrr.org/understanding-disaster-risk/terminology/hips/en0009 (2020).

[23] Gemmell N et al. Deep learning-based classification of lung sound using yamnet for identifying lung diseases. In: *Proceedings of IEEE International Conference on Biomedical and Health Informatics*. Chicago, IL, USA (2024), pp. 1–6.

[24] Group W B. Illegal logging, fishing, and wildlife trade: The costs and how to combat it (2019).

[25] UNEP-INTERPOL. The rise of environmental crime – a growing threat to natural resources peace, development and security (2016).

[26] Technologies I. Infineon technologies and rainforest connection use sensor technology for forest protection. https://www.infineon.com/cms/en/about-infineon/press/press-releases/2022/INFXX202212-043.html (2022).

[27] Yassin B et al. Rainforest connection & huawei: Acoustics for biodiversity monitoring and threat detection. https://panorama.solutions/en/solution/rainforest-connection-huawei-acoustics-biodiversity-monitoring-and-threat-detection (2021).

[28] Carvalho S & Gomes E F, Using mfcc and mel spectrogram features with deep learning for environmental sound analysis, *Vietnam Journal of Computer Science*, 10(1) (2023) 39–54.

[29] Zhang Y et al., Environmental sound classification using temporal-frequency attention based convolutional neural network model, *Scientific Reports*, 11. https://doi.org/10.1038/s41598-021-00594-5.

[30] Kim T et al., Deep learning-based environmental sound classification using feature fusion and data enhancement, *Computational Materials and Continua*, 74(1) (2023) 1069–1091.