



03 a 05 de dezembro de 2025

As Engenharias na Transformação Digital das PMEs

Artificial Intelligence and the Internet of Things: Innovations in Environmental Reserve Monitoring

Eduardo Hartleib Galvão

Academic Department of Informatics(DAINF) - Federal University of Technology - Paraná. Ponta Grossa, PR, Brazil.

Itamar Iliuk

Academic Department of Informatics(DAINF) - Federal University of Technology - Paraná. Ponta Grossa, PR, Brazil.

José Manoel Balthazar

Academic Department of Production Engineering(DAENP) - Federal University of Technology - Paraná. Ponta Grossa, PR, Brazil.

Ângelo Marcelo Tusset

Academic Department of Production Engineering(DAENP) - Federal University of Technology - Paraná. Ponta Grossa, PR, Brazil.

José de Lima

Academic Department of Electrical Engineering(DAELE) - Federal University of Technology - Paraná. Ponta Grossa, PR, Brazil.

Abstract: The advancement of illegal practices, such as deforestation and predatory hunting, represents a serious problem for the conservation of natural areas worldwide, compromising biomes and affecting the climate balance. One of the main challenges in environmental preservation is the difficulty of continuously monitoring extensive forest areas, often located in remote regions, which limits the effectiveness of enforcement. Facing this problem, the objective of this paper is to analyze the application of modern technologies in environmental enforcement, investigating how the integration between Artificial Intelligence (AI) and the Internet of Things (IoT) enables the real-time identification of illegal human activities. The methodology is based on the use of acoustic sensors distributed in natural reserves for the detection of characteristic noises, such as those from chainsaws. For this, an embedded system was developed using a XIAO ESP32S3 Sense module, and a Convolutional Neural Network (CNN) was trained on the Edge Impulse platform, utilizing the Mel Frequency Cepstral Coefficients (MFCC) algorithm and data from the Google AudioSet. As a main result, the model achieved an F1-Score of 88.0% in validation tests, demonstrating high accuracy. It is concluded that the integration of AI and IoT shows significant potential for conservation, presenting a transformative capability in ecosystem protection and the promotion of environmental sustainability.

Keywords: Environmental Enforcement, Deforestation, Artificial Intelligence, Internet of Things.

1. Introduction

The advancement of illegal practices in environmental preservation areas, such as deforestation, uncontrolled mining, and predatory hunting, represents a growing threat to ecosystems worldwide. The absence of effective prevention and enforcement methods compromises entire biomes, with direct impacts on biodiversity loss, climate balance, and society. In Brazil, millions of hectares of native vegetation have been lost in the last decade, especially in the Legal Amazon, where illegal logging continues to expand even in protected areas. Such activities not only destroy the habitat of countless species, like the Hyacinth Macaw (*Anodorhynchus hyacinthinus*), but also create conditions for the emergence and spread of diseases, such as malaria and viral zoonoses.

The main problem faced by environmental enforcement agencies lies in the difficulty of carrying out continuous and comprehensive monitoring of extensive and hard-to-access territories. Traditional methods, such as the use of satellite imagery or ground patrols, have significant limitations, including cloud cover, staff shortages, and logistical challenges, which prevent a swift response to environmental crimes.

In this scenario, technologies based on acoustic monitoring emerge as a promising alternative. Recent literature demonstrates the feasibility of using Artificial Intelligence (AI) to detect specific sound events for preservation purposes. For instance, works such as (ZHANG *et al.*, 2024) have explored the use of deep learning for the detection of forest fires from their acoustic signatures, while (ZIMMERMANN *et al.*, 2024) focused on the classification of noises to identify illegal mining activities. Although these approaches validate the use of acoustic analysis, this work distinguishes itself by focusing specifically on the detection of chainsaws—a direct indicator of deforestation—through a low-cost embedded system optimized for real-time operation.

The objective of this paper is, therefore, to develop and analyze a low-cost solution that integrates Artificial Intelligence and the Internet of Things (IoT) for environmental enforcement. The proposed system aims to identify characteristic sound patterns of chainsaws in natural ecosystems, enabling the real-time detection of illegal activities for a faster response from authorities. The methodology involves the use of a XIAO ESP32S3 Sense microcontroller with an integrated microphone and the training of a Convolutional Neural Network (CNN) on the Edge Impulse platform.

This paper is structured into five sections. The first section, this introduction, contextualizes the problem of deforestation and the limitations of current monitoring methods, presenting the objectives of the work. The second section details the methodology used for data collection and the development of the AI model. The third section presents and analyzes the results obtained. The fourth section discusses the implications of these results, while the fifth section presents the conclusions, limitations, and suggestions for future work.

2. Methodology

To create and train the proposed Artificial Intelligence model, the Google AudioSet sound database was utilized, which offers a wide variety of audio samples extracted from public videos on the YouTube platform. AudioSet is widely recognized for its comprehensiveness and quality in representing different sound events, making it ideal for environmental sound classification and detection tasks.

The model development process was conducted on the Edge Impulse platform, which allows for the creation of machine learning models optimized for embedded devices. The platform was used for feature extraction, such as Mel Frequency Cepstral Coefficients (MFCCs), construction of the neural network architecture, and the final training and validation of the model.

2.1. Dataset Construction and Augmentation

The audio samples used for training have a uniform duration of 5 seconds and were processed to ensure consistency in size and quality. The dataset was structured with two main classes relevant to the context of environmental monitoring: "environment" and "chainsaw". The final dataset comprises approximately 1930 audio files, totaling 2 hours and 40 minutes of data, with a balanced distribution between the two classes to ensure model uniformity.

2.1.1. Environment Class

The "environment" class represents the set of natural sounds typical of a preserved forest, which is fundamental for training detection models that need to distinguish between normal acoustic patterns and sounds indicative of human or illegal activities. This class is composed of recordings that simulate the real soundscape of forest areas, including both biotic sounds (from living beings) and abiotic sounds (from non-living natural phenomena). Biotic sounds include vocalizations of birds, insects, small mammals, and amphibians, while abiotic sounds include the sound of wind, moving leaves, streams, and rain.

2.1.2. Chainsaw Class

The "chainsaw" class represents the characteristic sounds of illegal deforestation activities. To build this class, data was initially sourced from Google AudioSet. However, as the source videos often contain long periods without the sound of interest, a filtering process was necessary to extract only pure audio segments. For this, a Python script was developed using the TensorFlow and Librosa libraries.

The script automated the validation of each downloaded audio file. First, the audios were segmented into 5-second clips. Each clip was then analyzed by YAMNet, a pre-trained sound event classification model available from TensorFlow Hub. A segment was only kept for the dataset if the YAMNet model classified the sound as "Chainsaw" with a minimum confidence threshold of 30%. This process ensured the creation of a clean, high-quality dataset containing exclusively the target chainsaw sound.

After the filtering stage, the dataset was expanded through a multifaceted data augmentation strategy, implemented with Python scripts and the pydub library. The approach was divided into two complementary fronts to simulate different field conditions:

- 1. Simulation of Noisy Environments:** In the first process, each pure chainsaw audio was mixed with an ambient noise sound (such as wind, rain, or fauna) randomly selected from a noise library. Before overlaying, the chainsaw audio's volume was reduced by a random factor between 5 and 10 decibels (dB). This technique generated "dirty samples" that simulate realistic scenarios where the sound of interest is partially masked by the forest's natural background noise.

2. **Simulation of Distance:** In the second process, to simulate the effect of distance between the sound source and the sensor, the pure chainsaw audio files were systematically duplicated and attenuated at fixed volume levels. New samples were created with reductions of 5, 10, 15, 20, and 25 dB, representing the weakening of the acoustic signal as the distance increases.

This two-step approach—first filtering for data purity and then diversifying through augmentation—was fundamental to creating a robust and representative dataset, enhancing the model's ability to generalize to unpredictable field conditions.

2.2. Audio Processing and Model Architecture

The audio data processing and the construction of the classification model were performed entirely on the Edge Impulse platform. Initially, Mel Frequency Cepstral Coefficients (MFCCs) were extracted from each 5-second audio sample. MFCCs are robust features that represent the short-term power spectrum of a sound and are widely used in audio recognition tasks. To further increase the model's robustness, additional data augmentation techniques were applied, including the addition of Gaussian Noise (stddev=0.2) and the use of SpecAugment on the spectrograms.

The architecture of the implemented Convolutional Neural Network (CNN) is as follows:

1. A GaussianNoise layer with a standard deviation of 0.2,
2. A Reshape layer to adapt the input dimensionality,
3. A first Conv1D layer with 16 filters, a kernel size of 3, and ReLU activation,
4. A MaxPooling1D layer with a pool size of 2 and strides of 2,
5. A Dropout layer with a rate of 0.2,
6. A second Conv1D layer with 8 filters, a kernel size of 3, and ReLU activation,
7. A second MaxPooling1D layer with a pool size of 2 and strides of 2,
8. A second Dropout layer with a rate of 0.2,
9. A Flatten layer to convert the feature maps into a 1D vector,
10. A final Dense output layer with softmax activation to produce classification probabilities;

The model was trained using the Adam optimizer with a learning rate (LEARNING_RATE) of 0.0005 and categorical crossentropy loss function. Training was conducted for 150 epochs (EPOCHS) with a batch size (BATCH_SIZE) of 32.

The platform also allows for quantization of the trained model to optimize its size and inference speed on microcontrollers, with the option to disable per-channel quantization (disable_per_channel_quantization) configurable to balance RAM and accuracy.

3. Results and Discussion

The trained model was deployed on the XIAO ESP32S3 Sense module. Real-time inference tests were conducted by capturing audio directly from the board's integrated microphone and processing the samples through the embedded model.

Class predictions (environment or chainsaw) and their respective probabilities were monitored via the microcontroller's serial port.

During the model development and training process on the Edge Impulse platform, several iterations were performed to optimize performance. Initially, with a more limited dataset

and less diverse environment sounds, the model achieved a validation accuracy of approximately 75.46% (March training).

3.1 Initial Model Version Results

The initial results of the classifier and model testing, using training data from March/April, are presented in Figure 1 and Figure 2. This figure illustrates the classifier's confusion matrix and corresponding performance metrics.

Figure 1. Confusion matrix of Classifier Version 1 (initialdata)

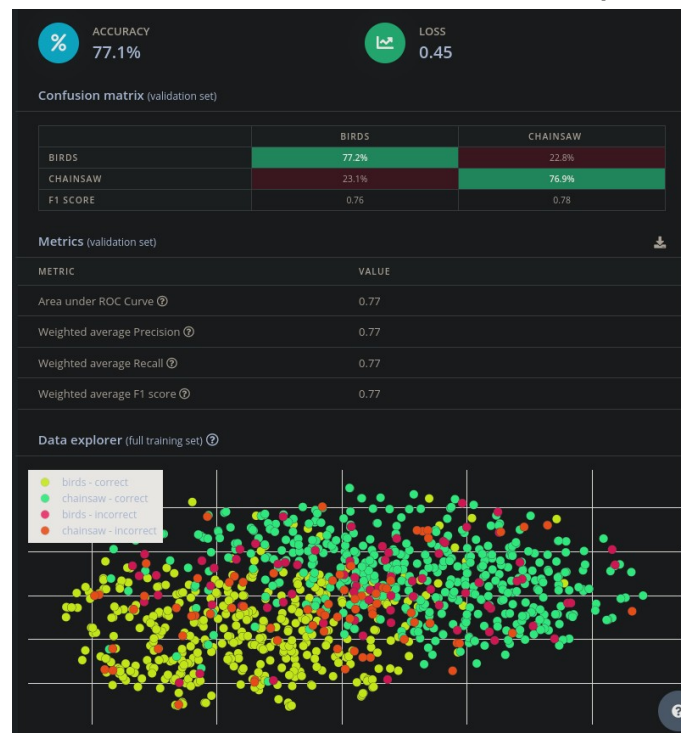
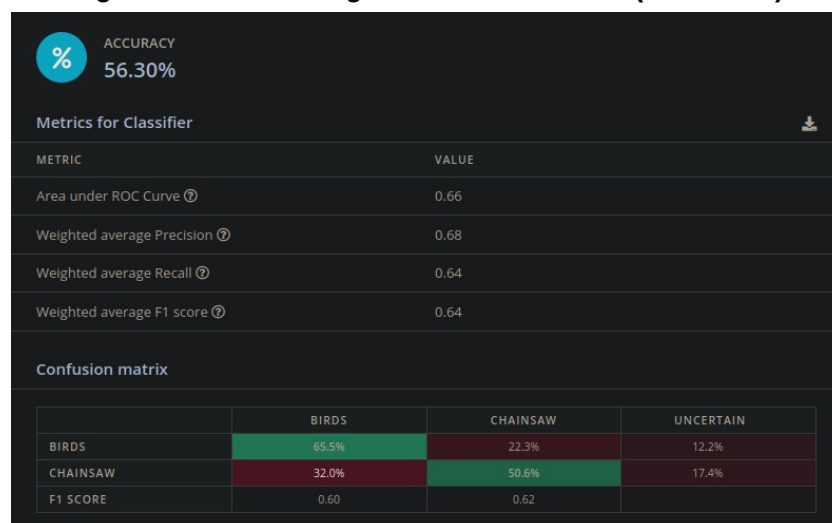


Figure 2. Model testing results for Version 1 (initial data)



3.2 Final Model Version Results

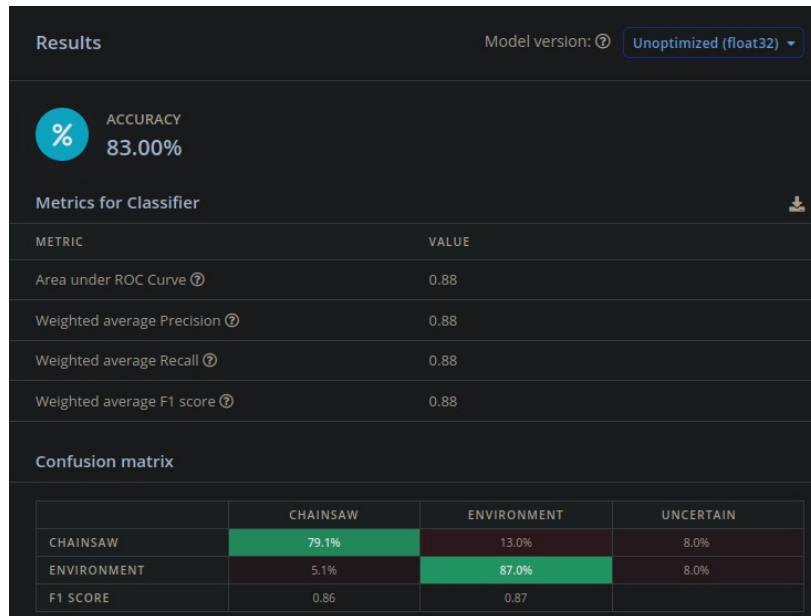
With the evolution of the dataset, including more variations of chainsaw sounds (with natural noise and at different decibel levels) and greater diversity in the 'environment' class (with other natural noises besides birds), the model's performance improved significantly. The final version of the model, after refining the dataset and architecture, demonstrated robust performance in the platform's internal validation. Based on the Edge Impulse platform's training logs, the model achieved a training accuracy of 86.65% and a loss of 0.3130. On the validation set, the metrics were even more promising, with a validation accuracy of 88.00% and a validation loss (val_loss) of 0.2893.

To evaluate the model's generalization capability, a test was performed on an independent test dataset ('Model testing') on the Edge Impulse platform, using the quantized (Int8) model. In this test, the model obtained an overall accuracy of 83.00%. The class-weighted metrics were notable: Area Under the ROC Curve (AUC-ROC) of 0.88, Precision of 0.88, Recall of 0.88, and weighted F1-Score of 0.88. Figure 4 and Figure 5 presents the confusion matrix and model testing results for the final version of the classifier.

Figure 4. Confusion matrix of the Final Classifier Version



Figure 5. Model testing results for the Final Version (quan-tized)



The confusion matrix (detailed in Figure 4) revealed the following results for the 'chainsaw' and 'environment' classes:

- chainsaw: 79.1% true positives, 13.0% incorrectly classified as 'environment', and 8.0% uncertainty;
- environment: 87.0% true positives, 5.1% incorrectly classified as 'chainsaw', and 8.0% uncertainty.

The individual F1-Scores for 'chainsaw' and 'environment' were 0.86 and 0.87, respectively, confirming the model's high capability to distinguish between the sounds of the two classes in a controlled environment and with unseen test data. The model optimization and quantization performed by the Edge Impulse platform ensured that its size and processing requirements were suitable for the ESP32S3 microcontroller, not constituting a limiting factor for speed or inference capability on embedded devices. The model's performance itself, as validated in software tests with clean datasets, remained robust and demonstrated the solution's potential for acoustic monitoring of illegal activities. In addition to the validation on a controlled test dataset, the model underwent a practical evaluation with more recent data, reflecting a continuous monitoring scenario. In these tests, its performance proved exceptional, achieving an accuracy of **98.8%** and a loss of just **0.03**. The corresponding confusion matrix (Figure 6) reinforces this robustness, with correct classification rates of **99.3%** for the 'chainsaw' class and **98.3%** for 'noise', resulting in F1-Scores of **0.99** for both. An analysis of the classification distribution in a real-time scenario revealed that, out of 162 analyzed samples, the model identified **55.56%** as 'chainsaw' and **44.44%** as 'noise' (Figure 7). This predominance of 'chainsaw' detections, coupled with the extremely high precision demonstrated, confirms the model's effectiveness and sensitivity in the field, thus validating its practical application for detecting acoustic events of interest.

Figure 6. Final model performance results on the validation set

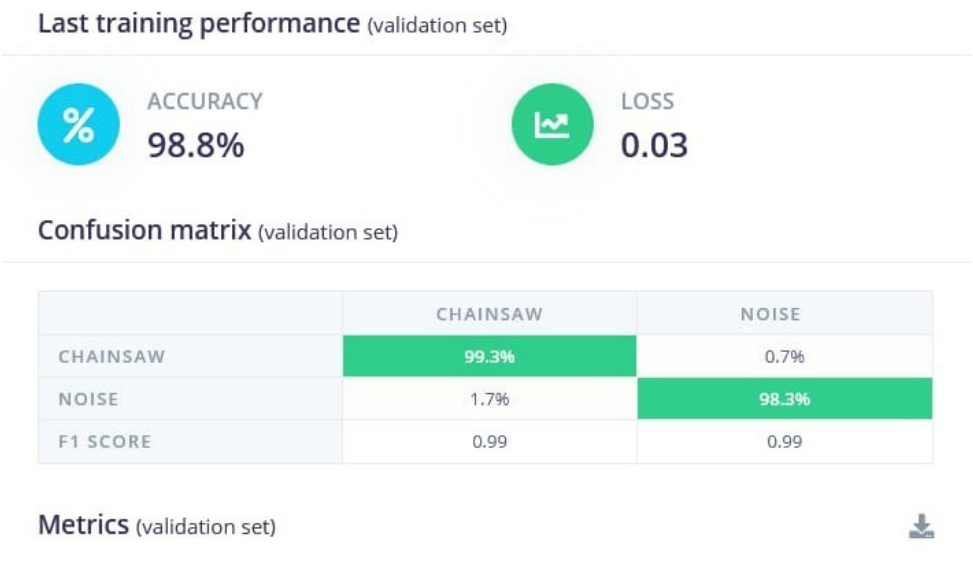
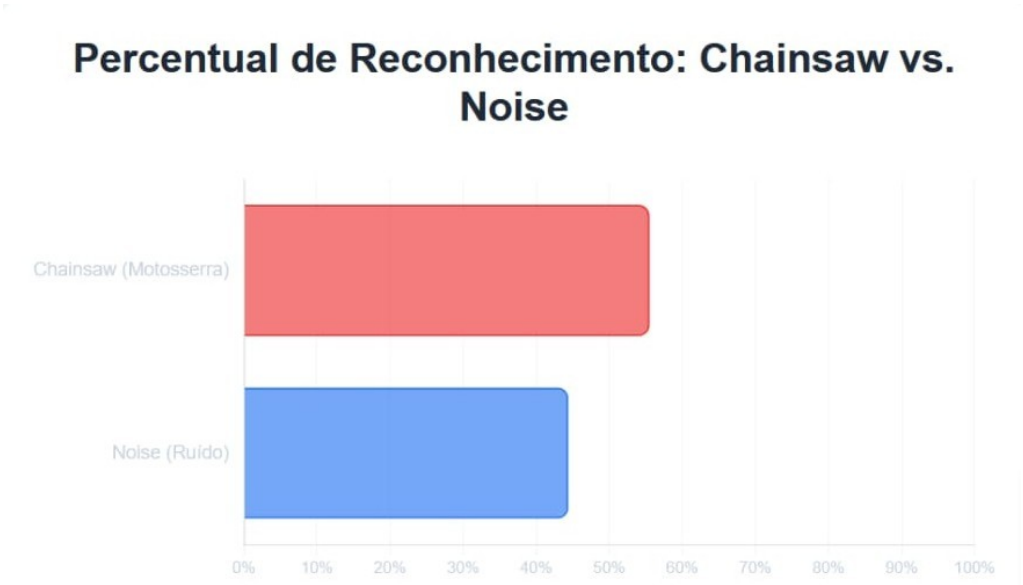


Figure 7. Classification distribution from the practical field test.



4. Discussion

The model's evaluation on the XIAO ESP32S3 Sense module under real hardware conditions revealed computationally efficient behavior, with average Digital Signal Processing (DSP) times of approximately 73.5 ms and extremely fast classification times of only 1 ms. This agility ensures real-time inference capability, crucial for continuous monitoring applications.

During the most recent practical tests, the model demonstrated a significant leap in performance, validating the effectiveness of the refined dataset and architecture. The system achieved an exceptional accuracy of **98.8%** with a minimal loss of **0.03**. An analysis of the confusion matrix revealed an outstanding ability to distinguish between classes, correctly identifying **99.3%** of 'chainsaw' instances and **98.3%** of 'noise' instances. This high degree of precision and recall is encapsulated by the F1-Score of **0.99** for both classes, indicating a robust and well-balanced classifier.

This level of performance overcomes the limitations of earlier iterations, which tended to produce low-confidence predictions. In contrast, the final model operates with high certainty. Furthermore, the analysis of the classification distribution (55.56% 'chainsaw' vs. 44.44% 'noise') demonstrates that the model is not only accurate but also highly sensitive to the target acoustic event in a real-world environment, without a bias towards the noise class. In summary, the model's behavior on the hardware, validated by these strong empirical results, confirms its robustness and viability as a reliable tool for acoustic monitoring.

5. Conclusions

This work demonstrated the development and effectiveness of an Artificial Intelligence model for the acoustic detection of chainsaws, crucial for the environmental monitoring of conservation areas. The model's robustness was evidenced by the training metrics on the Edge Impulse platform, which showed improved performance with the diversification of the dataset, underscoring the fundamental importance of a solid and representative database for the success of AI solutions.

The optimization of the model through quantization made it intrinsically suitable for deployment on low-power embedded systems, such as the XIAO ESP32S3 Sense microcontroller. Hardware tests confirmed its ability to perform fast and consistent inferences (classification times of 1 ms and DSP of 73.5 ms), enabling real-time detection. The model demonstrated the capability to identify chainsaw sounds with significant probabilities, confirming its expected behavior and the validity of the approach.

The implementation of this system represents a significant advancement in the efficient monitoring of environmental reserves, allowing for the early detection of environmental crimes and agile decision-making by authorities. Looking ahead, the model's scalability potential is notable. It can be extended to include the detection of other relevant acoustic events, such as those associated with illegal mining and poaching, by adding specific sound data. Additionally, integration with technologies like mesh networks for communication between devices can enhance the system's coverage and reliability, further strengthening environmental preservation strategies.

Acknowledgements

The author thanks LABRIOT (Robotics and IoT Laboratory) and UTFPR (Federal University of Technology – Paraná) for the provided infrastructure.

References

- AMAZON CONSERVATION.** *Put science and tech to work.* 2025. Available at: <https://www.amazonconservation.org/what-we-do/put-science>. Accessed on: Sep. 14, 2025.
- BARBOSA, I.** More than 361 million trees were cut down in the legal amazon in 2022, reports greenpeace. *Portal Amazônia*, 2022. Available at: <https://portalamazonia.com/meio-ambiente/mais-de-361-milhoes-de-arvores-foram-cortadas-na-amazonia-le>. Accessed on: Sep. 14, 2025.
- INSTITUTO ARARA AZUL.** *Hyacinth Macaw - instituto arara azul.* 2025. Available at: <https://www.institutoararaazul.org.br/especies/arara-azul>. Accessed on: Sep. 14, 2025.
- MAPBIOMAS ALERTA.** *Map - deforestation and degradation.* 2024. Available at: <https://brasil.mapbiomas.org/2024/08/21/em-2023-a-perc>. Accessed on: Sep. 14, 2025.
- NATIONAL GEOGRAPHIC BRAZIL.** Deforestation is causing an increase in infectious diseases in humans. 2019. Available at: <https://www.nationalgeographicbrasil.com/meio-ambiente/2019/12/desmatamento-esta-causando-a>. Accessed on: Sep. 14, 2025.
- RAINFOREST CONNECTION.** *Ecoacoustics - listening to nature.* 2025. Available at: <https://rfcx.org/eoacoustics>. Accessed on: Sep. 14, 2025.
- ZHANG, Tianyu et al.** A novel sound event detection method using deep learning for forest fire detection. *Applied Acoustics*, v. 215, 109720, 2024.
- ZIMMERMANN, Rafael et al.** Acoustic classification of artisanal and small-scale mining soundscapes for identifying illegal mining activities in the Amazon rainforest. *MethodsX*, v. 12, 102604, 2024.