Audio Vs. Visual Approach To Monitor The Critically Endangered Species *Atlapetes blancae*: Developing Deep Learning Models With Limited Data

Julian David Santamaria Pabon ,Bs
(SISTEMIC, Universidad de Antioquia)
Contact email: Julian.santamaria@udea.edu.co

Claudia V. Isaza, PhD. (SISTEMIC, Universidad de Antioquia)
Jhony Giraldo, PhD. (Télécom Paris, France)
Angélica Diaz, MSc. (Alexander Von Humboldt Institute, *Neotropical Innovation Corporation*)

Julian David Santamaria Pabon

2023 — Electronic Engineering of the Universidad de Antioquia

2023 — Conferences:
- VI Colombian Congress of Zoology
- International Conference on Computer Vision (ICCV 2023)

2024 — Pursuing a Master's degree in Engineering at Universidad de Antioquia

Topics of research interest

Topics of Interest:

Data classification in animal conservation, focus on camera traps and ecoacoustic data.

Current Projects:

Language Models and Contrastive Learning Techniques:

Applying techniques to improve the classification of conservation-related data. Exploring how natural language processing techniques can enhance data interpretation and model training.

Problematic

The species was believed extinct for many years.

In **2018**, rediscovery of *Atlapetes blancae* was reported

(Yovany Ochoa,2019)

Problematic

(Chiquito , Quiñones, 2019)

Problematic
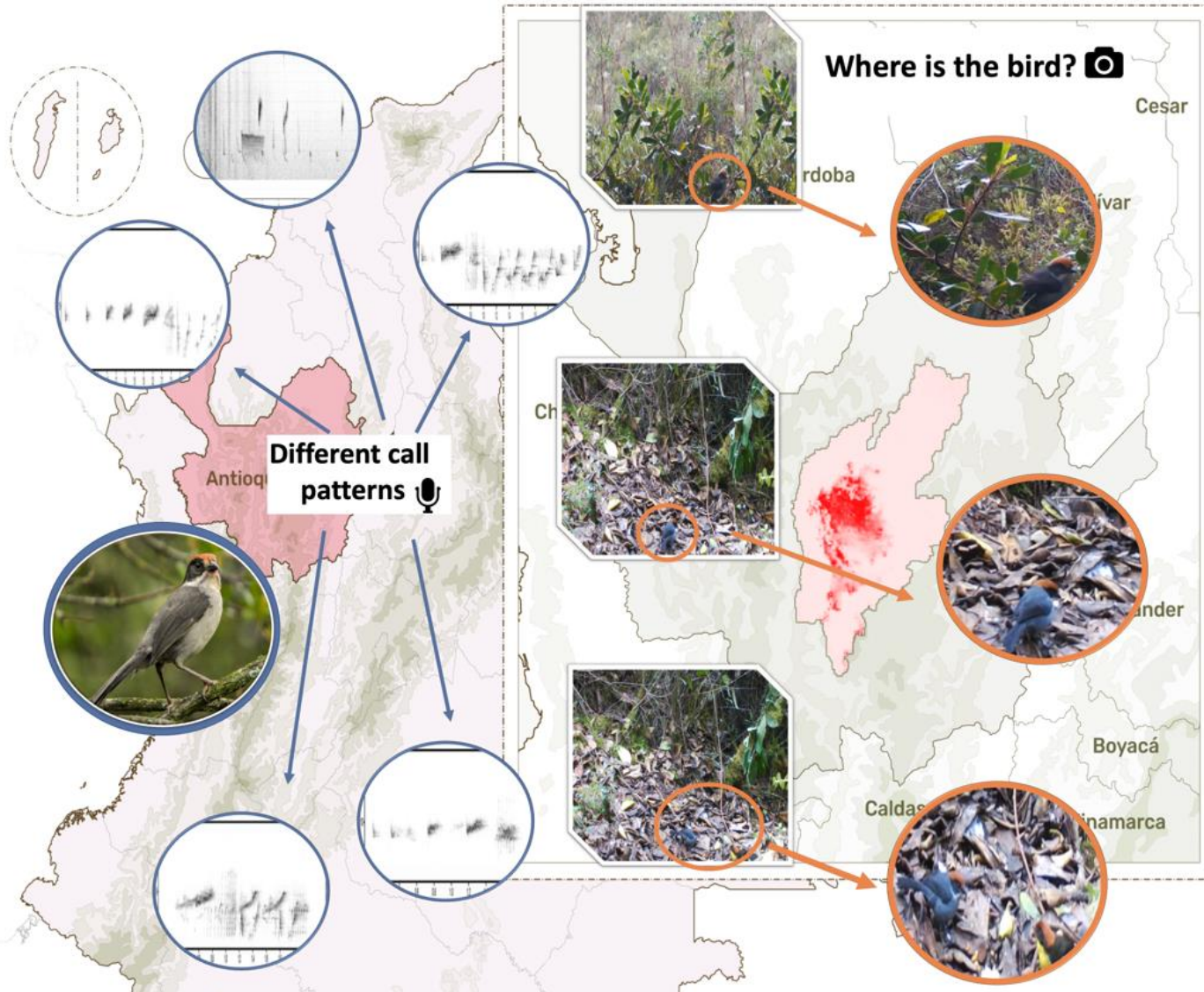
Materials

(Chiquito, 2019)
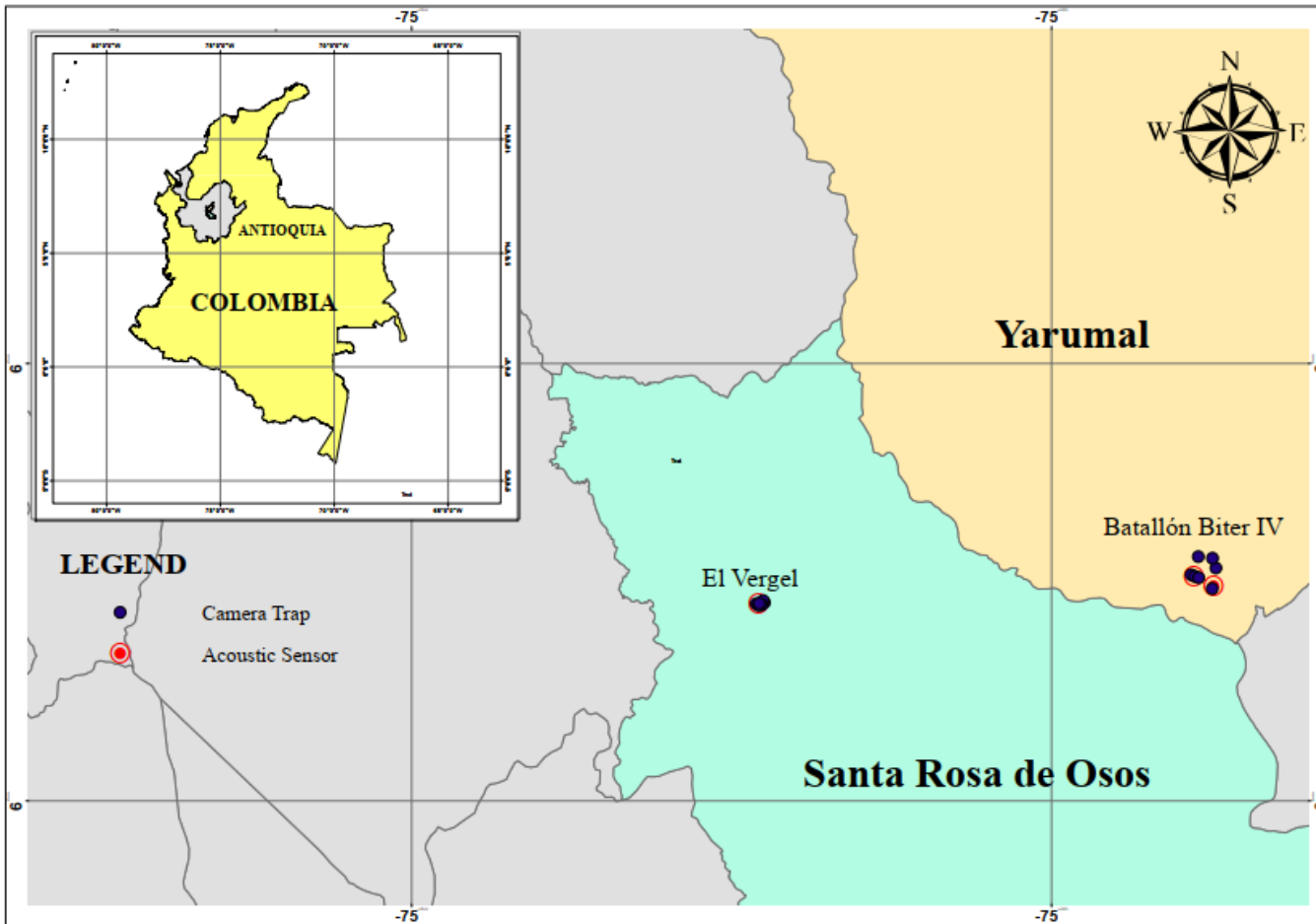
	Recollected data	Presence of A.blancae	Number of sensors
Audios	7147	11	3 Audio recordings
Videos	17159	48	13 Camera trap

Audio Methodology

Audio Methodology

Preprocessing

Segmentation (AED modified)

Feature Extraction (Linearly-spaced cepstral coefficients and frequency information)

$$y_i = [f_{peak}, f_{min}, f_{max}, f_1, \ldots, f_{24}]$$

Clustering
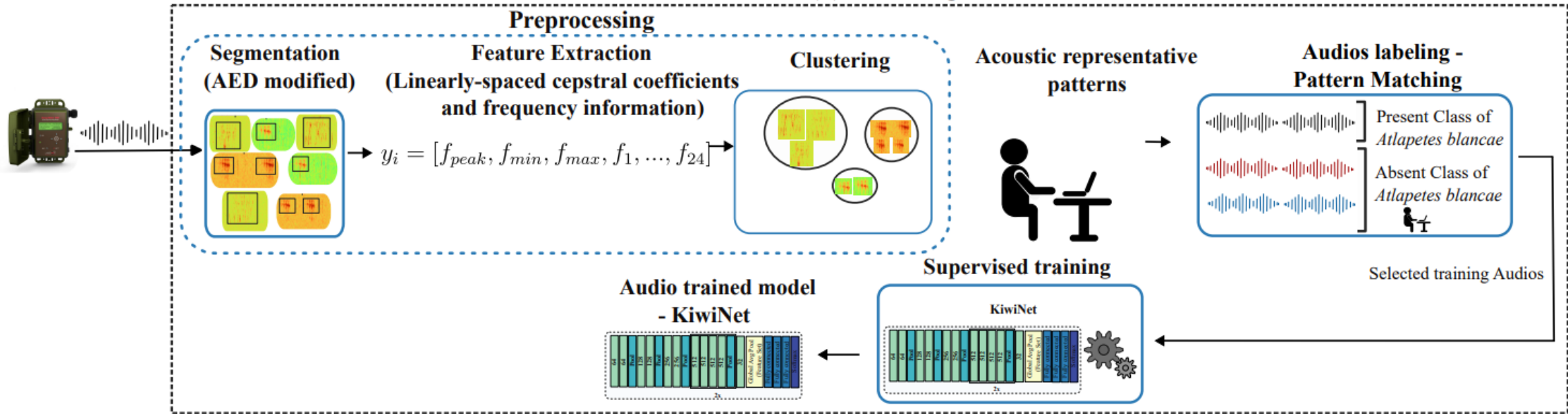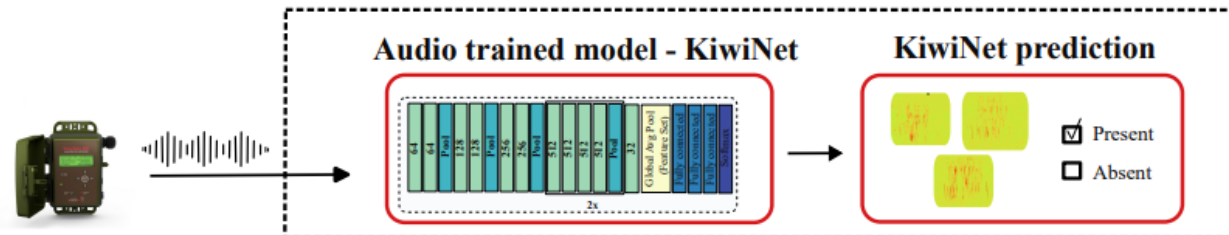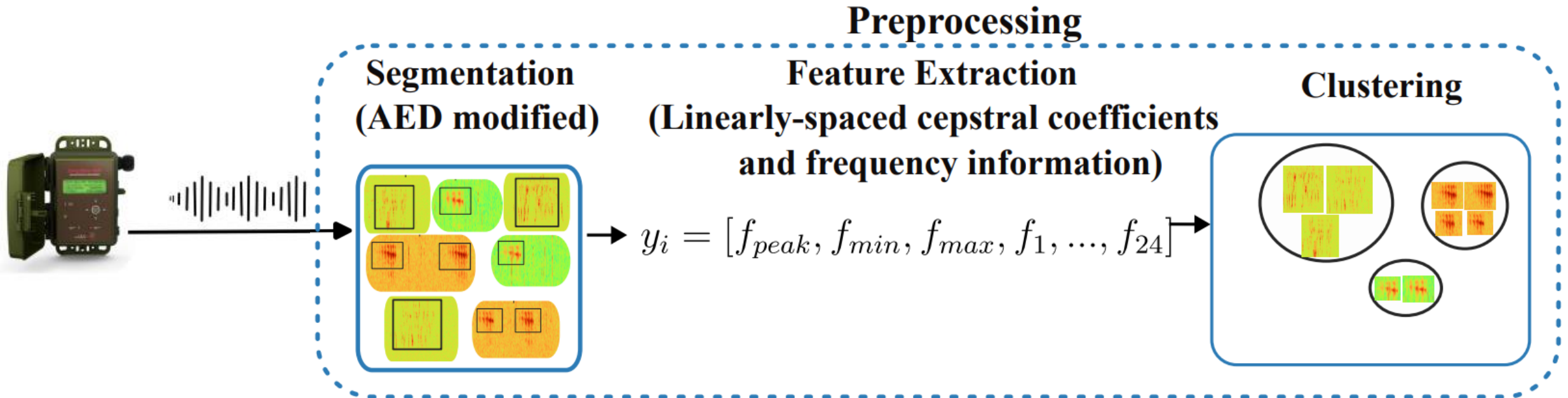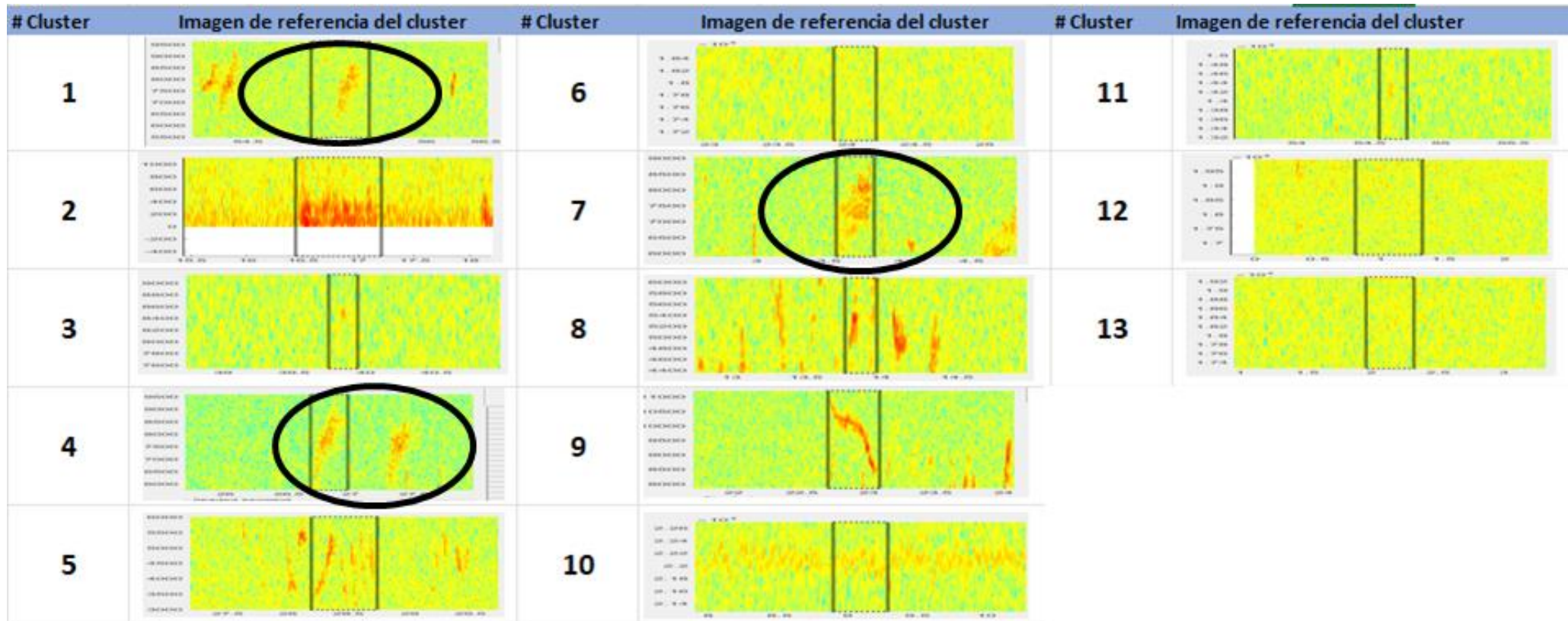
We move from analyzing the 17,147 omnidirectional audios to only analyzing a few specific examples.

Audio Methodology

Audio Methodology
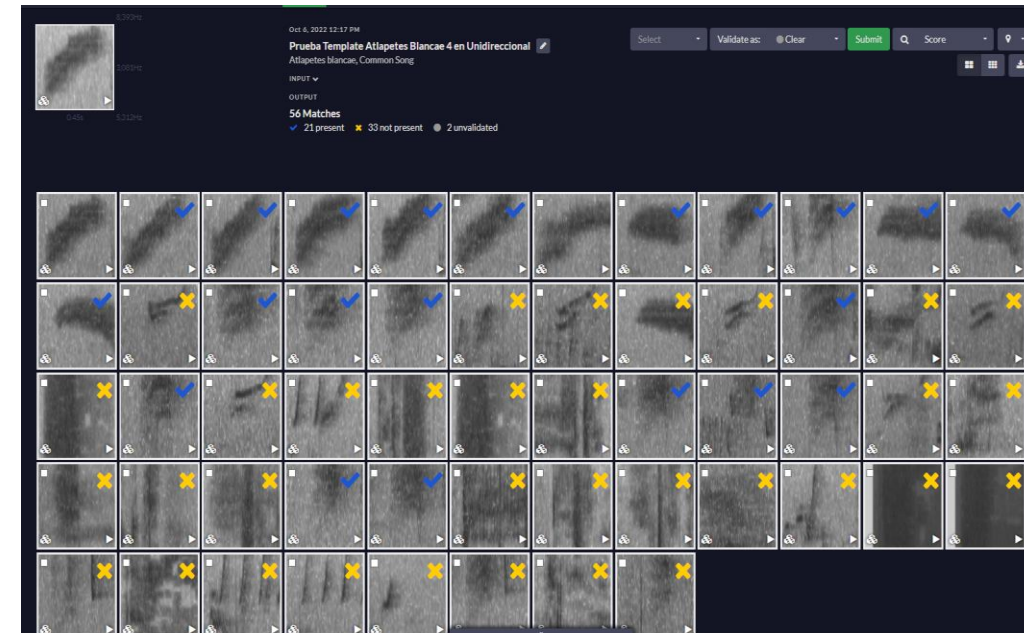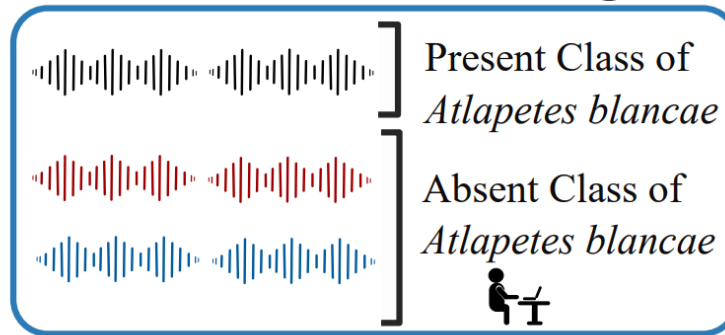

Acoustic representative patterns
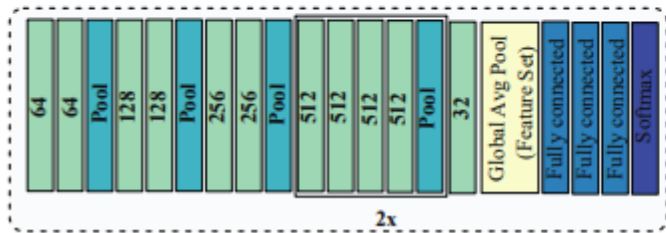
Audios labeling - Pattern Matching

Present Class of *Atlapetes blancae*

Absent Class of *Atlapetes blancae*

(Aide et al., 2013)

Audio Methodology

Video Methodology

Video Methodology

Video Methodology
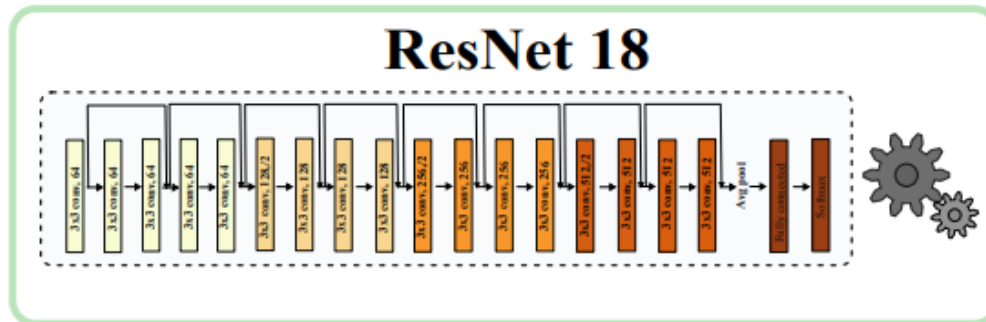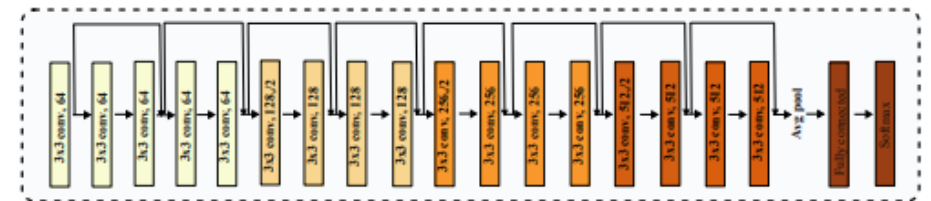
Results

Audio Results

COMPARING THE PERFORMANCE OF AUDIO MODELS ON TESTING DATA				
Model	**F-measure avg**	**Precision avg**	**Recall avg**	**Accuracy**
Our audio recognition model	**0.823**	**0.823**	**0.823**	**0.964**
Acoustic animal identification	0.743	0.690	0.805	0.929
Arbimon	0.794	0.981	0.667	0.964
ResNet-18	0.653	0.580	0.748	0.821

Video Results

COMPARING THE PERFORMANCE OF IMAGE MODELS ON TESTING DATA					
Model	**F-measure avg**	**Precision avg**	**Recall avg**	**Accuracy**	**Type of Data**
RPCA ResNet-18 (Ours)	**0.940**	**0.953**	**0.928**	**0.967**	**Cropped images**
RPCA ResNet-50 (Ours)	0.937	0.954	0.921	0.966	Cropped images
RPCA ResNet-101 (Ours)	0.926	0.947	0.905	0.956	Cropped images
RPCA ResNet-18 (Ours)	0.495	0.512	0.882	0.889	Videos
ResNet-18	0.473	0.475	0.472	0.467	Frames

Video Results

COMPARING THE PERFORMANCE OF IMAGE MODELS ON TESTING DATA					
Model	F-measure avg	Precision avg	Recall avg	Accuracy	Type of Data
RPCA ResNet-18 (Ours)	**0.940**	**0.953**	**0.928**	**0.967**	**Cropped images**
RPCA ResNet-50 (Ours)	0.937	0.954	0.921	0.966	Cropped images
RPCA ResNet-101 (Ours)	0.926	0.947	0.905	0.956	Cropped images
RPCA ResNet-18 (Ours)	0.495	0.512	0.882	0.889	Videos
ResNet-18	0.473	0.475	0.472	0.467	Frames

Conclusions and Future Perspectives

- Identifying patterns in images is easier than in audio, making deep learning directly applicable to visual data.

- Our findings indicating that the audio model is the preferred choice for processing the data. However, this model represents just the initial step in the development of a sufficiently robust tool for *Atlapetes blancae* recognition.

- Currently, the models use audio and visual information independently. Future work will focus on integrating data from both sensors to improve accuracy and robustness.

(Chiquito,2019)

Thank you very much

✉ **julian.santamaria@udea.edu.co**