WHAT'S THAT SOUND?

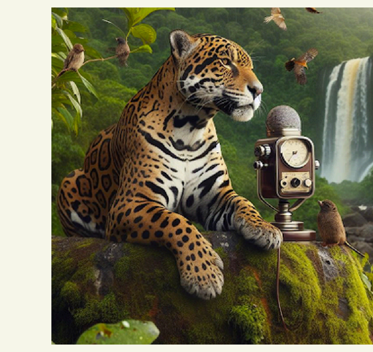
USING AI TO IDENTIFY WILDLIFE FROM SOUND RECORDINGS

American Museum of Natural History
Center for Biodiversity and Conservation

JADE VELAZQUEZ[1], NICO WEISBERGER[2], MOHAMMED CHOWDHURY[3], DANIEL LOPEZ-LOZANO[4]

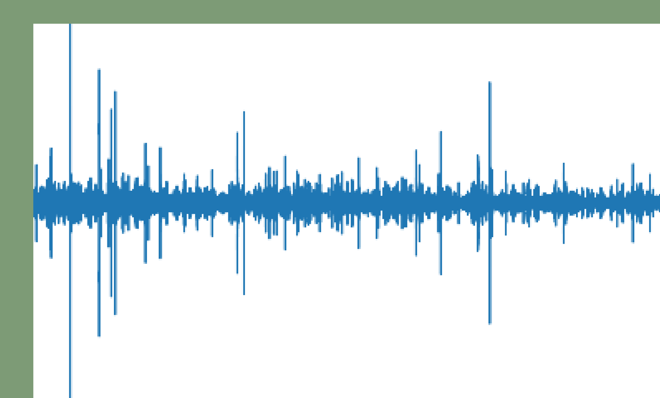
[1]THE YOUNG WOMEN'S LEADERSHIP SCHOOL OF MANHATTAN, [2]HUMANITIES PREPARATORY ACADEMY, [3]TRINITY SCHOOL, [4]CENTER FOR BIODIVERSITY AND CONSERVATION - AMERICAN MUSEUM OF NATURAL HISTORY

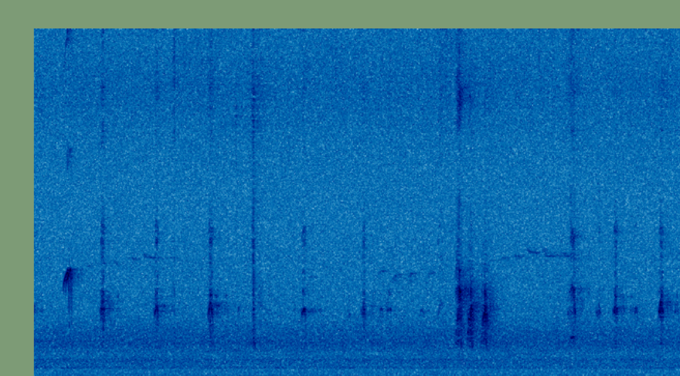
SCIENCE RESEARCH MENTORING PROGRAM



INTRODUCTION

- Climate change and habitat destruction are critically endangering species and disrupting ecosystems, which may cause animals to leave their natural habitats or go extinct.
- By analyzing sound recordings, scientists can localize, identify species presence, and detect changes in animal populations. These results can provide crucial data for the government to make informed decisions that can foster biodiversity and address ongoing environmental crises.
- Sound recorders are cost effective and provide expansive data coverage. However, the challenge lies in the exhaustive hours scientists spend analyzing audio files to identify the presence of animals.
- Existing AI models for wildlife sound detection often face limitations. Many are not open-source, holding back collaboration and development. Additionally, some models are highly specialized for specific species or regions, limiting their broader applicability.
- Our goal is to implement and train a versatile machine learning model using a real-world dataset to identify wildlife sound segments in audio recordings.
- Our approach involves working with spectrograms (visual representations of strength of a signal over time). We will then use convolutional neural networks (CNNs) to classify these spectrograms.





RESULTS

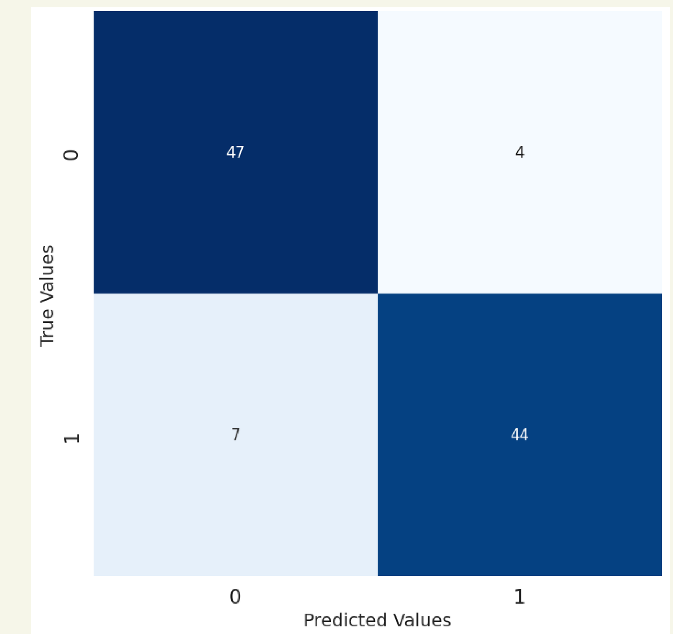


Fig. 1a

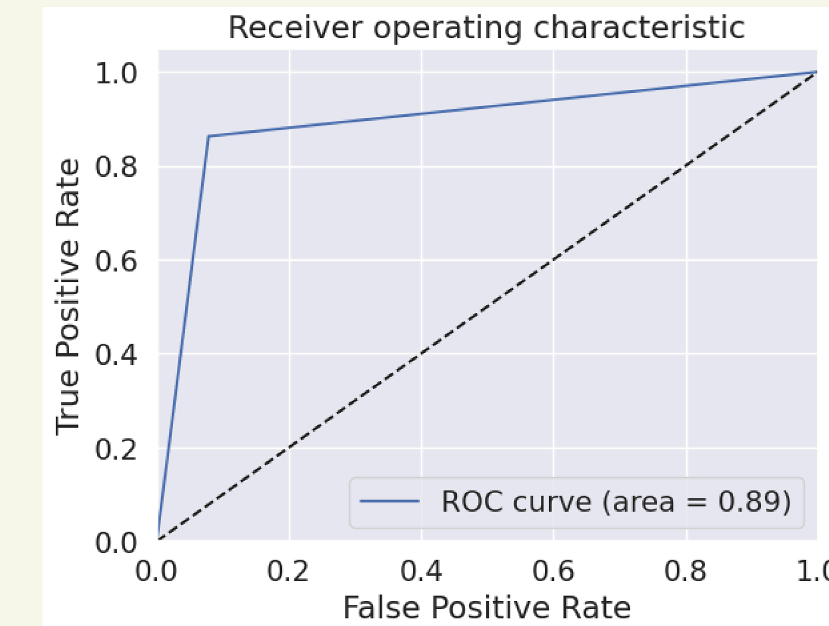


Fig. 2a

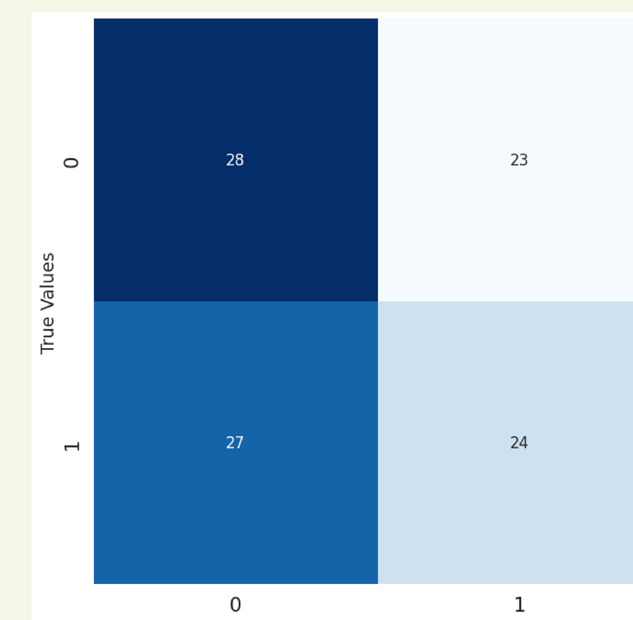


Fig. 1b

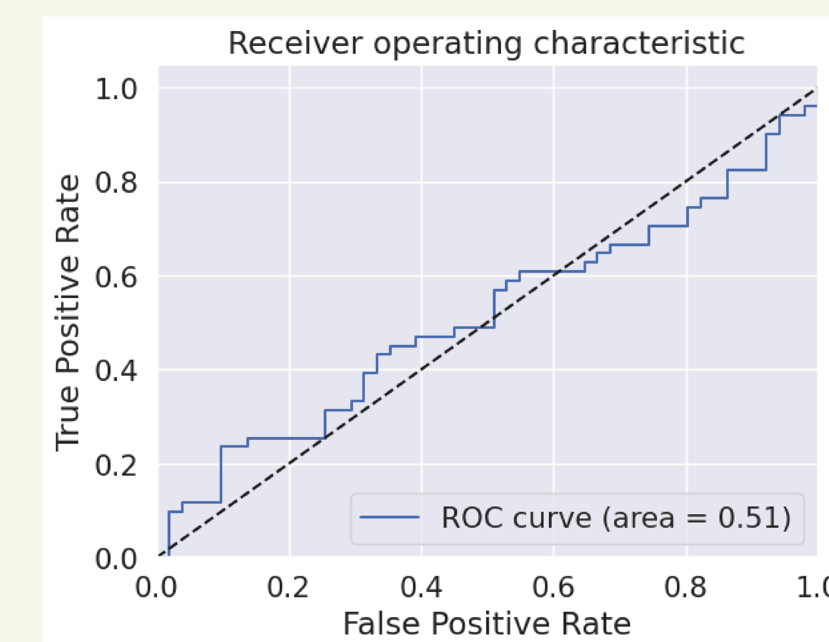


Fig. 2b

Fig 1: Confusion Matrix with the True Positive, True Negative, False Positive and False Negative values for a) Model 1 and b) Model 2. Fig 2: ROC curve. for a) Model 1 and b) Model 2.

	Model 1	Model 2
Accuracy	0.89	0.51
Precision	0.92	0.51
Recall	0.86	0.47
F1-Score	0.89	0.49

Table 1. Performance Metrics on Wildlife Sound Detection for Each Model

python™

METHODS

1. Data Collection
- 6 Audio Datasets (646 audio files)
- Over Time
- Single Location
- Au Sable Forks, NY

2. Preprocessing
- Select samples from each dataset.
- Audio split into 5-second segments.
- Converted audio segments to spectrogram images.

3. Data labeling
- 1012 segments manually labeled (animal sounds 1, background 0)
- Split for training & testing

4. CNN Architecture selection

Model 1: CNN
- Used a Convolutional Neural Network (CNN)
- Trained with balanced labeled data (background vs. wildlife)
- Supervised learning approach

Model 2: Autoencoder
- Trained on background only
- Reconstructs spectrograms
- High reconstruction error for wildlife sounds vs. background

5. Training
- Trained 2 models (50 epochs each)
- Model 1: 1022 segments using 92 for validation.
- Model 2: 460 segments.

6. Tuning
- Optimized model parameters (batch size, epoch number, learning rate)
- Improved accuracy & loss metrics

7. Evaluation
- Tested on unseen data
- Evaluated: Accuracy & Loss (training), Confusion Matrix & ROC (final)
- Ensured accurate segment classification

DISCUSSION

- Our first model showed promising results for wildlife sound detection. However, incorporating more diverse data from different regions and recorders could enhance its versatility for researchers.
- Despite underperforming, the second model has room for improvement through additional training data (matching the first model), more rigorous hyperparameter tuning, and incorporating preprocessing steps.
- Exploring unsupervised or semi-supervised learning is a promising future direction, as data labeling is expensive and time-consuming.
- This research demonstrates the potential of machine learning for wildlife sound detection when using a diverse open data set to train a versatile model. By incorporating more sophisticated preprocessing techniques and exploring advanced architectures, future research can further enhance model performance.

ACKNOWLEDGEMENTS

We would like to thank Dr. Jemima Georges and the American Museum of Natural History Student Research Mentorship Program (SRMP); The Center for Biodiversity and Conservation at AMNH, especially Peter Ersts for his inspiration and facilitation of the data, and Dr. Mary Blair for her support. We would also like to thank all of our parents for supporting us on this journey <3

REFERENCES

- STowell D. 2022. Computational bioacoustics with deep learning: a review and roadmap. PeerJ 10:e13152
- Penar, W., Magiera, A., & Klocek, C. (2020). Applications of bioacoustics in animal ecology. Ecological Complexity, 43, 100847. https://doi.org/10.1016/j.ecocom.2020.100847