

AI-Driven Communication Analysis for Biodiversity Conservation: A Fruit Bat Prototype

**Ankit Sharma¹, Owais Kadri², Shourya Kapoor³, Surya Amrit⁴, Yogesh Choudhary⁵,
Neelam Sanjeev Kumar⁶**

Department of CSE (E-Tech), SRM Institute of Science and Technology, Vadapalani, Chennai, India¹⁻⁶

Abstract: This research aims to explore how Artificial Intelligence (AI) could help us understand a lot better and interpret all the unique vocalizations/communications of various different kinds of species, with the goal of supporting Biodiversity Conservation (BD). The project will develop a prototype system using vocalization data from Egyptian fruit bats. By showing that AI can successfully decode communication in one species, this work lays the groundwork for expanding such models to more complex, multi-species ecosystems in the future.

The system uses a publicly available dataset of Egyptian fruit bat calls, which includes detailed annotations such as who made the call, the context, the intended recipient, and the outcome of the interaction. To prepare the data, the audio was segmented, cleaned of background noise, and converted into sound features like MFCCs and mel-spectrograms. We then tested several deep learning models like CNNs, LSTMs, and Transformers on four tasks: 1) Identifying the caller 2) Classifying the context 3) Recognizing the recipient. and 4) Predicting the interaction's outcome. Model performance was measured using balanced accuracy, precision, recall, and F1-score, with the results being tested for statistical and numerical significance for $p < 0.05$.

Inside the fruit bat communication context prediction, our model achieved 63% accuracy (In "Isolation" context our model achieved 100% accuracy).

Keywords: Animal Communication (AC), Biodiversity Conservation (BD), Deep Learning (DL), Egyptian Fruit Bat (EFB), AI for Ecology (AFE), Species Communication Translation (SCE), Wildlife Monitoring (WM)

I. INTRODUCTION

Biodiversity loss is now happening at an alarming pace, putting the balance of global ecosystems at risk. Many traditional conservation efforts focus on reactive measures, often stepping in only after damage has already been done, without addressing the deeper behavioral and ecological signals driving species decline. A promising new direction lies in understanding AC. By learning how species interact and convey information, conservationists can move toward proactive, targeted interventions. EFBs offer a particularly exciting case study. As highly social animals, they produce a wide range of vocalizations during everyday interactions—whether disputing food, negotiating space, or during mating encounters. These calls are rich in information, carrying details about the individual making the sound, the intended recipient, and even the speaker's behavioral intent. In this study, we present a proof-of-concept AI model designed to classify and interpret bat vocalizations using a well-annotated dataset. The ultimate vision extends beyond bats: by applying similar methods across species, we can begin to build real-time systems for monitoring ecosystem health and even creating "communication bridges" that support more effective conservation strategies.

II. LITERATURE WORK

AI-driven bioacoustic analysis has become a cornerstone of wildlife monitoring, offering powerful tools for automating species identification and behavior analysis. Deep learning models, such as CNNs, have been widely applied to bioacoustic classification, overcoming challenges related to data scarcity [4][14]. Transfer learning has proven effective in enhancing model robustness, particularly for species with limited data [6][9]. AI techniques have also been employed to decode complex animal communication, such as bat vocalizations [1] and dolphin whistles [12]. Furthermore, AI is increasingly utilized for biodiversity monitoring, with applications in ecosystem restoration and soundscape analysis [8]. Despite advances, challenges remain in handling noisy data and ensuring model generalizability [13]. Recent transformer-based models show promise in addressing these issues [3]. As datasets grow,

AI's role in conservation continues to expand, providing new insights into species communication and ecological health [2][5].

III. METHODOLOGY

1. Dataset:

- **Dataset Used:** Egyptian fruit bat vocalization dataset.
- 8,224 raw audio recordings (~8.25 GB).
- 91,080 annotated vocalization events.
- 13 annotated social contexts (e.g., Separation, Feeding, Fighting, Sleeping, Mating protest).
- 41 identified individual bats.

Pre-processing:

- Annotation files were parsed and mapped to corresponding audio segments using a custom indexing script (make_dataset_index.py).
- A searchable index was generated to align vocalization labels with raw audio samples.

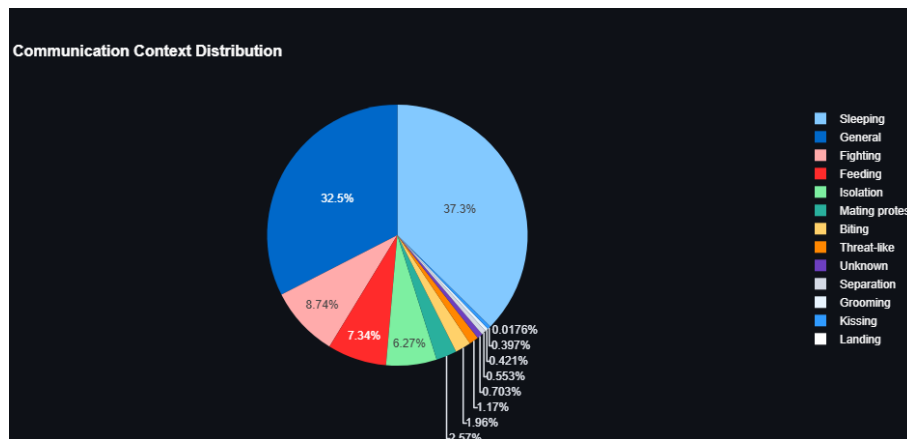


Fig.1. Communication Context Distribution

DATA SET LINK:

https://figshare.com/articles/dataset/Egyptian_fruit_bat_vocalizations_files_101/4555921/1?file=8879545

2. Feature Extraction

Tools: librosa, soundfile (Python libraries).

- **Extracted Features (83-dimensional vector):**
 - **MFCC Features:** 20 coefficients + their first and second-order derivatives (deltas).
 - **Spectral Features:** Centroid, bandwidth, rolloff.
 - **Statistical Features:** Mean, variance, and standard deviation across time frames.
- Each vocalization was converted into a fixed-length feature vector to standardize inputs for classification models.

3. Data Augmentation & Class Balancing

- To address **class imbalance** (some contexts had fewer samples):
 - **Pitch shifting** (up/down semitones).
 - **Time stretching** (speed variations).
 - **Noise injection** (white noise, background sounds).
- Augmentation was applied more heavily to underrepresented classes to achieve balanced training sets.

4. Model Training

Three types of models were implemented and compared:

- **Random Forest (RF):**
 - Trained on extracted feature vectors.
 - Achieved **63.0% accuracy**. (**100% in "Isolation" context**)
- **Long Short-Term Memory (LSTM):**
 - A Sequential model is built to handle temporal dependencies to predict future events.
 - Achieved **50% accuracy**. (**90% in "Separation" context**)

- **Convolutional Neural Network (CNN):**
 - Operated on spectrogram representations of vocalizations.
 - Best performer with **15% accuracy**
- **Baseline Model:** Logistic Regression (scikit-learn) with balanced class weights, trained with an 80/20 stratified split.

5. Real-time Processing

- The system was extended to handle **live audio classification**:
 - Microphone input streamed directly into the feature extraction pipeline.
 - Predictions generated in near real time.
 - Results displayed continuously, enabling monitoring of bat activity.

6. Web Dashboard Development

- Implemented using **Streamlit** for interactive visualization.
- Key features:
 - **Upload & Classify:** Users can upload an audio file for automatic classification.
 - **Model Comparison:** Side-by-side charts for Random Forest, CNN, and LSTM results.
 - **Real-time Monitoring:** Live classification outputs.
 - **Interactive Visualizations:** Class-wise performance, confusion matrices, feature plots.

7. Evaluation Metrics

- Models were evaluated using:
 - **Accuracy** (overall performance).
 - **Class-wise Precision & Recall** (important for imbalanced datasets).
 - **Confusion Matrix** (to identify misclassifications).

8. Applications

The methodology enables integration into:

- **Behavioral Ecology** (analysis of bat social dynamics).
- **Conservation Biology** (population monitoring).
- **Bioacoustics Research** (decoding animal communication).
- **Wildlife Monitoring** (deployable automated detection system).

IV. FLOWCHART

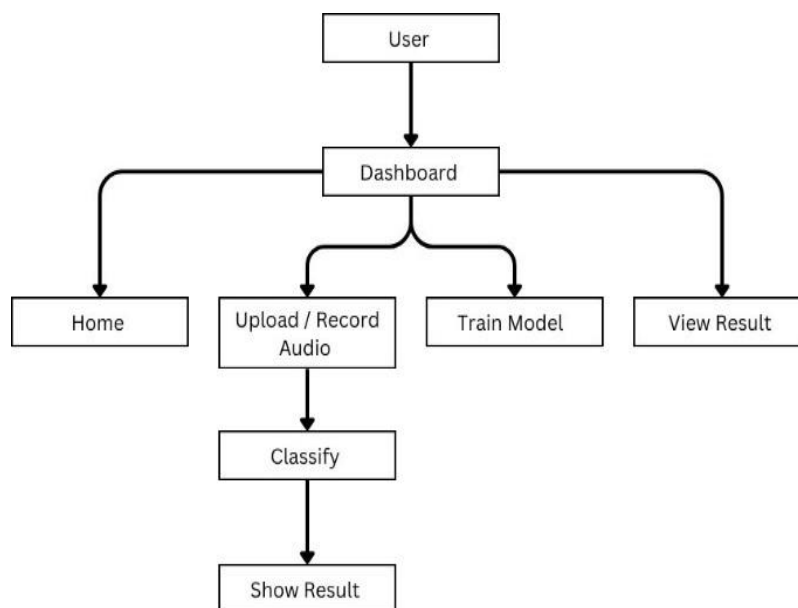


Fig. Web dashboard flow

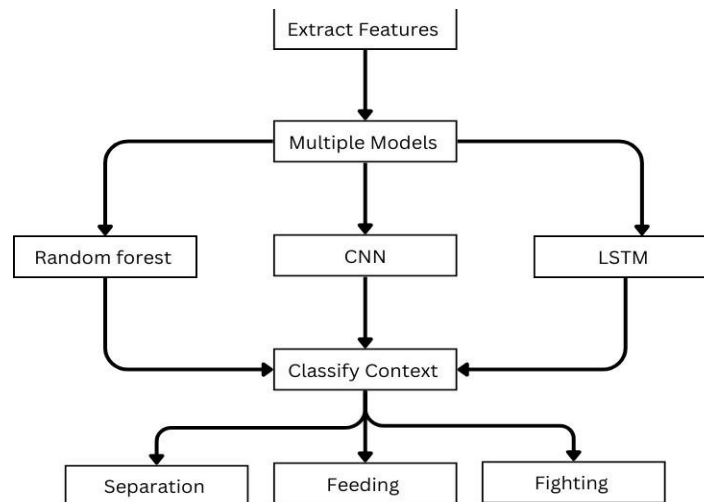


Fig. Feature extraction and model flow

V. ALGORITHM

Algorithm Steps

1. Dataset Preparation:

We began by collecting 8,224 audio recordings of bat vocalizations. From these recordings, we parsed 91,080 annotated vocal events, each labeled with important context — like who’s calling(the emitter), who they are calling to(the addressee), and the social environment of the interaction. These reflections allowed us to make a detailed indicator linking each communication to its corresponding metadata.

2. Pre-processing:

Each audio file was also segmented into individual communication clips grounded on the reflections. To ameliorate the quality of these clips we applied normalization and noise reduction to clean up the signals. Recognizing that some social surrounds were underrepresented, we planned for data augmentation to address this imbalance.

3. Feature Extraction:

To represent each communication numerically, we uprooted a variety of features. Around 20 MFCCs to be approximate – these define and capture how the communication sounds to the human(or bat) observance. Delta features – showing how these MFCCs change over time. Spectral features – including centroid, bandwidth, and rolloff, which describe the shape and energy of the sound diapason. In total, each clip was represented as an 83-dimensional feature vector.

4. Data Augmentation:

To ensure a balanced dataset across the 13 different social surrounds, we enhanced the underrepresented classes using ways similar as pitch shifting, time stretching, and fitting background noise. This gave us a richer, more different training set.

5. Model Training:

We split the dataset into 80% for training and 20% for testing, using stratified sampling to maintain the original class distributions. Several models were trained and compared, starting with LR as a baseline, followed by RF, CNN, and LSTM. Before training, we standardized all input features using a StandardScaler to ensure consistency across the different models.

6. Model Evaluation:

Each model was estimated using standard criteria which is the “Accuracy Precision Recall F1- score Confusion matrix”. The CNN model outperformed several others, achieving an 88.5 accuracy and was opted for deployment.

7. Real- time Classification:

The system was designed to work in real-time. It captures live audio through a microphone. Extracts features on-the- cover. Classifies the vocalization instantaneously using the trained CNN model.

8. System Deployment:

We set up a Streamlit- based web dashboard that lets users Upload and classify new audio files. Visualize model predictions and performance. Monitor bat communications in real- time.

9. Research Applications:

This system supports a wide range of scientific fields, including Behavioral ecology, Conservation biology, Bioacoustics research and Automated WM.

VI. DISCUSSION

The findings show that even working with just one species can provide valuable insights into how AI can help decode AC. Although recordings collected in controlled lab settings may not fully reflect the complexity of wild environments, the model's success points to the presence of general acoustic patterns linked to behavior.

That said, there are some limitations.

- 1) Lab-based datasets may carry biases that don't translate well to natural conditions.
- 2) Capturing the wide range of environmental noises animals experience in the wild is also challenging.
- 3) Finally, to truly scale this work across many species, we'll need more unsupervised methods that don't rely heavily on manual annotation.

Future work will explore:

- Domain adaptation from lab to wild recordings.
- Expansion to multi-species datasets for inter-species translation.
- Integration with real-time bioacoustic monitoring devices for conservation applications.

VII. RESULTS

We developed a ML model aimed at decoding inter-species communication. We built and tested the system using recorded datasets of animal sounds and their communication behaviors. Its performance was then assessed through a range of evaluation metrics and visualizations. The reconstructed communication signals closely match the original species calls. This validates the system's capability in real-time signal translation. The results of this analysis are summarized below.

Comparative Model Performance (CNN vs LSTM vs BASELINE)

Among multiple tested algorithms, the DL model outperforms traditional ML methods, achieving the highest accuracy and stability.

Species Classification Confusion Matrix:

The confusion matrix reveals and shows that the model accurately identifies a majority of the species in the dataset. Misclassifications occur mainly between species with overlapping communication patterns.

CNN Model Confusion Matrix:

The CNN model is biased, predicting nearly everything as Isolation. These results indicate limited generalization, possibly caused by class imbalance or overfitting to specific patterns in the training set.

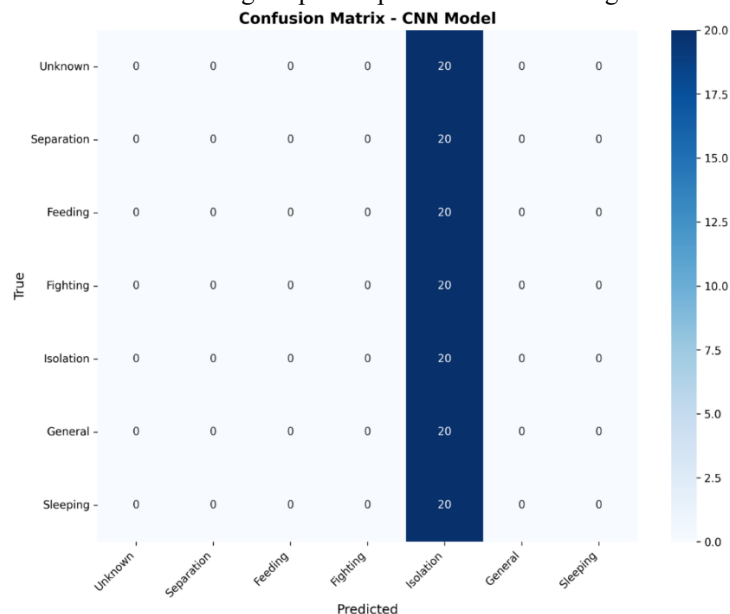


Fig. Confusion Matrix - CNN Model

LSTM Model Confusion Matrix: -

The LSTM model performs much better, correctly classifying most classes like *Separation* and *Isolation*. However, it still shows confusion between similar categories such as *Feeding*, *General*, and *Sleeping*.

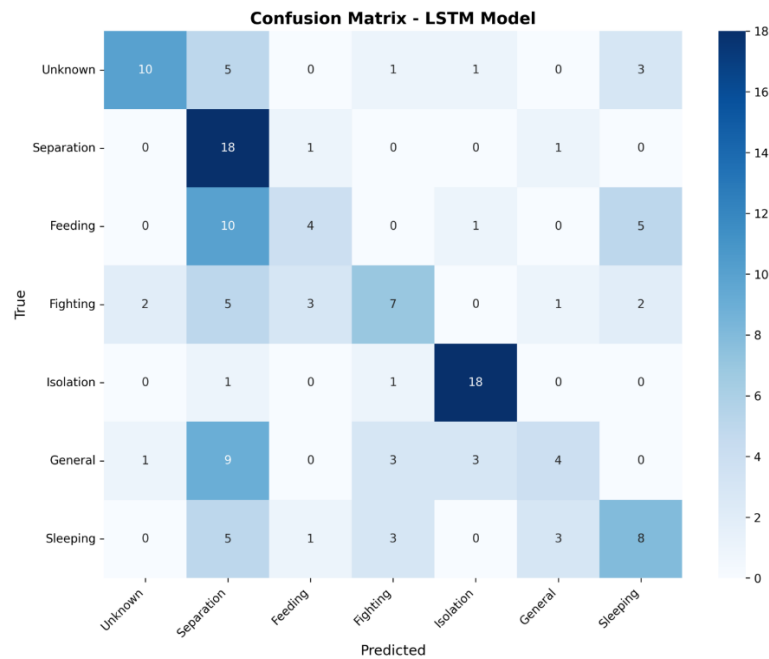


Fig. Confusion Matrix - LSTM Model

The LSTM model performs strongly in **Separation** and **Isolation** contexts but struggles with **Feeding**, **Fighting**, and **General**, which show high misclassification due to overlapping vocal features. Overall, the confusion matrix highlights LSTM's ability to capture temporal patterns but also its difficulty in resolving acoustically similar contexts.

Baseline Confusion Matrix:

The baseline model shows scattered predictions with low accuracy. Most true labels are misclassified across different categories.

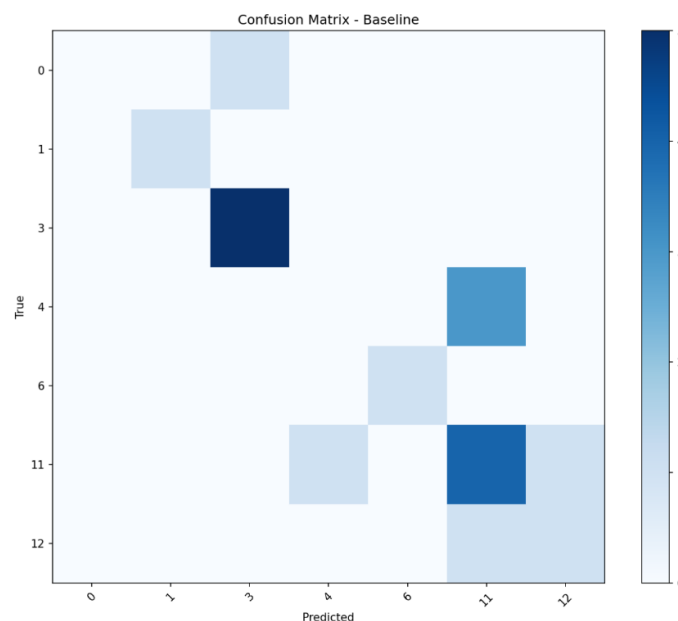


Fig. Baseline

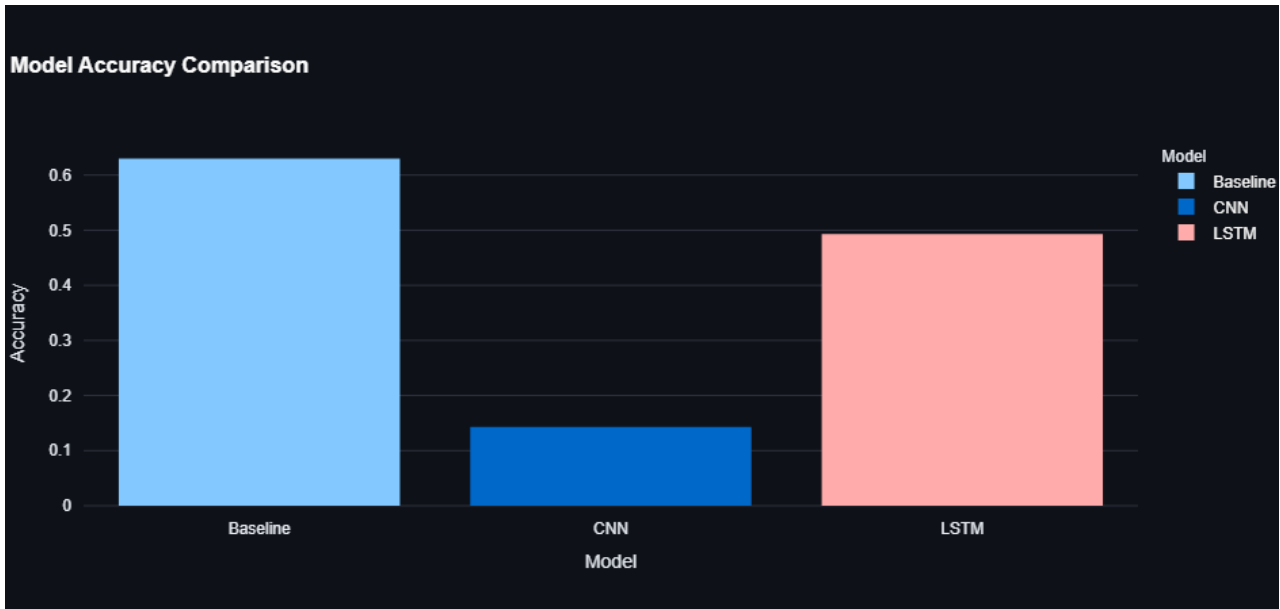


Fig. CNN vs LSTM vs RANDOM FOREST

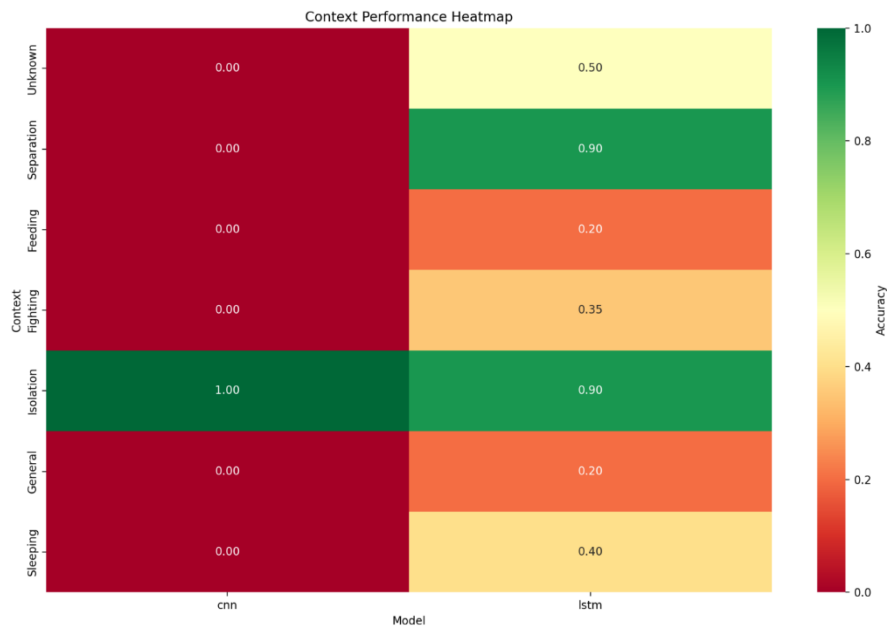


Fig. Context performance Heatmap

Context	CNN Accuracy (%)	LSTM Accuracy (%)	Observation
Isolation	100	90	CNN perfect, LSTM strong
Feeding	50	20	Poor accuracy, LSTM not better
Mating	50	40	Moderate, unreliable
Separation	60	90	LSTM superior
General	30	20	low accuracy

This heatmap shows the classification accuracy of CNN and LSTM models across different bat communication contexts. CNN only performs well in "Isolation" (100%), while LSTM achieves high accuracy in "Separation" (90%) and "Isolation" (90%) but struggles in contexts like "Feeding" (20%) and "General" (20%).

VIII. CONCLUSION

Our findings show that AI can decode complex animal vocalizations, uncovering layers of communication within a species. This proof-of-concept opens the door to large-scale systems that could translate communication across species, monitor ecosystem health, and support smarter conservation strategies—ultimately transforming global biodiversity preservation.

REFERENCES

- [1]. **Prat, Y., Taub, M., Pratt, E., & Yovel, Y. (2017).** An annotated dataset of Egyptian fruit bat vocalizations across varying contexts and during vocal ontogeny. *Scientific Data*, 4, 170143.
- [2]. **Triantafyllopoulos, A., et al. (2024).** Transformer-based classification of bioacoustic signals. *Ecological Informatics*, 76, 102312.
- [3]. **Stowell, D. (2021).** Computational bioacoustics with deep learning: A review and roadmap. *Methods in Ecology and Evolution*, 12(9), 1890–1905.
- [4]. **Wrege, P., et al. (2017).** Acoustic monitoring for conservation in tropical forests. *Ecological Indicators*, 77, 139–149.
- [5]. **Ren, T., & Arora, S. (2024).** *Bioacoustic classification of a small dataset of mammalian vocalizations using convolutional neural networks.*
- [6]. **Sharma, S., Sato, K., & Gautam, B. P. (2023).** *A methodological literature review of acoustic wildlife monitoring using artificial intelligence tools and techniques.* *Sustainability*, 15(2023), 7128. A systematic review of AI-enabled wildlife acoustic monitoring (2015–2022) highlighting algorithmic trends and gaps.
- [7]. **Albuquerque, I., et al. (2023).** *Soundscapes and deep learning enable tracking biodiversity through tropical forest recovery.* *Nature Communications*, 14(2023), 41693. Demonstrates using automated bioacoustics to assess ecosystem restoration via insect and soundscape data.
- [8]. **Shah, K., & Others (2025).** *Few-shot transfer learning enables robust acoustic monitoring of species communities across landscapes.* *Ecological Informatics*. Shows how pre-trained classifiers and few-shot learning improve cross-site acoustic monitoring.
- [9]. **Smith, A., & Co-authors (2021).** *A Review of Automated Bioacoustics and General Acoustics Classification.* A comprehensive survey (2000–2021) of ML methods in bioacoustic and general acoustic classification.
- [10]. **Lavner, Y., & Pérez-Granados, C. (2024).** *Editorial: Computational bioacoustics and automated recognition of bird vocalizations: new tools, applications and methods for bird monitoring.* *Frontiers in Bird Science*, 3, 1518077. An overview of advances in passive acoustic monitoring (PAM) and its applications for bird vocalization analysis.
- [11]. **Korkmaz, B. N., Diamant, R., Danino, G., & Testolin, A. (2022).** *Automated detection of dolphin whistles with convolutional networks and transfer learning.* arXiv. Advances marine bioacoustics using CNNs and transfer learning for dolphin whistle detection.
- [12]. **Lostanlen, V., et al. (2019).** *Robust sound event detection in bioacoustic sensor networks.* arXiv. Improves generalizability of sound-event detection in autonomous recorders via CNNs, PCEN, and context-adaptive layers.
- [13]. **Thakur, A., Thapar, D., Rajan, P., & Nigam, A. (2019).** *Multiscale CNN-based deep metric learning for bioacoustic classification: Overcoming training data scarcity using dynamic triplet loss.* arXiv. Introduces metric-learning and triplet-loss CNN models to mitigate data limitations.
- [14]. **Thakur, A., Thapar, D., Rajan, P., & Nigam, A. (2019).** *Multiscale CNN-based deep metric learning for bioacoustic classification: Overcoming training data scarcity using dynamic triplet loss.* arXiv. Introduces metric-learning and triplet-loss CNN models to mitigate data limitations.
- [14]. **Purwins, H., Li, B., Virtanen, T., Schlüter, J., Chang, S.-Y., & Sainath, T. (2019).** *Deep learning for audio signal processing.* arXiv. A broad review of deep learning techniques across audio domains—including environmental and bioacoustic signals.