Journal of Research and Innovation in Technology, Commerce and Management
Vol. 2 Issue 6, June 2025, pp. 2676-2681
ISSN: 3049-3129(Online)

Wings of Warning: Bird Song Forecasting

KR Shanmugapriyaa
Assistant Professor, Coimbatore Institute of Technology, India, krspriyaa@gmail.com
ORCID: https://orcid.org/0000-0003-2460-7213

Julie K
Coimbatore Institute of Technology, India, jaijulie40@gmail.com

Kaviya K
Coimbatore Institute of Technology, India, kavikumar1670@gmail.com

Praveena S
Coimbatore Institute of Technology, India, praveenashanmugavelu.2004@gmail.Com

Shrinithi V
Coimbatore Institute of Technology, India, shrivenkat9143@gmail.com

Cite as:
KR Shanmugapriyaa, Julie K, Kaviya K, Praveena S, & Shrinithi V. (2025). Wings of Warning: Bird Song Forecasting. Journal of Research and Innovation in Technology, Commerce and Management, Volume 2(Issue 6), pp. 2676 –2681. https://doi.org/10.5281/zenodo.15710268

DOI: https://doi.org/10.5281/zenodo.15710268

Abstract

Birds are highly sensitive to environmental changes, which alters their vocalizations in response to natural disasters, climate variations, and human disturbances. Our project, "Wings of Warnings: Bird Song Forecasting," mainly focuses on analyzing bird songs to detect early warning signals of environmental changes. By using deep learning techniques and bio-acoustic analysis, it classifies bird vocalizations to determine whether they indicate an impending disaster or regular communication. It utilizes an auto encoder model to extract meaningful features from bird songs. Auto encoders are used to learn compact representations of data and enables efficient feature extraction by encoding bird call characteristics such as frequency, tone, and pattern variations. These extracted features are then compared with labelled data using cosine similarity. By determining the similarity between an incoming audio signal and pre-classified bird sounds, the system effectively categorizes new recordings into

predefined classes. The model classifies bird vocalizations into seven distinct labels of two different species: the European Starling (Sturnus vulgaris) and the Canary (Serinus canaria). These species were chosen because they have diverse vocalizations and respond to environmental changes. This AI-driven system enables real-time monitoring and predictive analytics which significantly helps researchers, conservationists, and disaster management teams to derive actionable insights. By analyzing variations in bird calls, the system enhances early warning mechanisms for natural disasters such as earthquakes, and forest fires. This project goes beyond disaster detection and helps conserve biodiversity and study ecosystems. Understanding bird sounds can help protect bird species and keep nature balanced. It combines bioacoustics, deep learning, and environmental science to improve nature-based forecasting.

Keywords
Bio-acoustic analysis, frequency, tones, patterns, deep learning, environmental events, bird warnings, real-time tracking, ecological relevance, biodiversity conservation.

1. INTRODUCTION

Birds are very sensitive to environmental changes, often changes its vocalizations according to external factors such as climate shifts, natural disasters, and human activities. Bird calls serve as a form of communication, warning signals, or indicators of regional changes. By understanding the vocalizations, we can provide valuable insights into ecological patterns, disaster predictions, and climate changes.

This project focuses on analysis of bird songs and calls to identify whether it is for indicating climate change, disaster or normal communication. Using deep learning algorithms like auto encoders, features were extracted and compared with labeled audios for similarity and the output is predicted. Using datasets of bird vocalizations recorded before and during events, the model improves our understanding of how birds respond to its surroundings. By converting this bird songs and calls to useful communication, we can predict enormous useful insights.. By integrating AI with ecology, this project aims to fill the gap between nature and technology.

1.1 Labeling and Learning from Vocalizations

The first step is collecting a dataset of canary and European starling audio recordings, each has different types of incident. To help the model understand what each sound might mean, we also labeled a few reference audios into categories like "singing in a tree," "calling friends," and "warning signal."

Once we had the data, we cleaned it up. A lot of the raw recordings had background noise, so we used audio processing techniques to filter that out. This made sure the model would learn from t the

bird's voice, not random wind or traffic sounds.

After cleaning, we extracted features from the audio clips using Mel spectrograms, which turn sound into a kind of visual pattern, almost like a fingerprint of each sound. These spectrograms were then fed into a deep learning model we trained to recognize the features in sounds.

To test the model, we built a simple interface where you can upload a new canary sound. The system then compares that test audio with the labeled reference sounds and tells you which one it's most similar to. For example, if a bird is singing alone in a relaxed pattern, it might match with the "singing in tree" category. If the call is sharp and repetitive, the system might predict it's a "warning" or a "call to others."

The overall goal is to help people, researchers, or even wildlife conservation teams understand bird behavior better especially in changing environments. If we can learn how birds respond to different situations through their calls, we might even use that as a natural signal for things like climate changes, habitat disturbances, or approaching danger.

2. Requirements

The project's dataset was collected from prerecorded bird call datasets from Xeno-canto and the Macaulay Library. Google Colab was used for cloud-based experimentation. The models were developed using TensorFlow 2.10 with Keras for implementation and PyTorch 1.13 for additional experimentation.Python, PyTorch, torchaudio, librosa libraries were used for audio processing and scikitlearn for visualization and GPU-accelerated training using Google Colab or CUDA-enabled systems.

3. Methodologies

Bird calls are highly variable in pitch, duration, and temporal structure. Some may last milliseconds (short chirps), while others span several seconds (songs). These characteristics make it crucial to use models that can:

- Learn local patterns in frequency (pitch, timbre),
- Model temporal sequences (when sounds occur),
- And generalize across varying acoustic conditions (background noise, overlapping sounds).

3.1 Pre-processing

- All audio files are in .wav format.
- Used tools like DeepFilterNet to remove background noise (wind, people, etc.),Fig-3.1.1
- All audio files were converted to the same sampling rate (e.g., 16,000 Hz).
- Converted each audio file into a log-mel spectrogram.
- Each spectrogram was normalized (scaled) to ensure equal treatment of all samples during training.

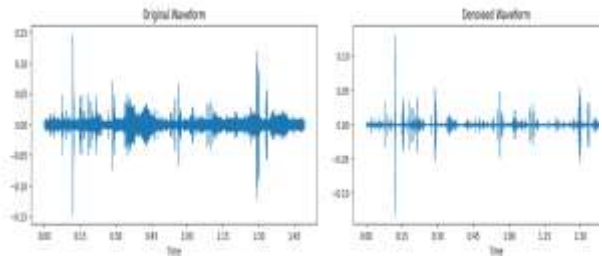

Fig 3.1 Original Vs Denoised Waveform

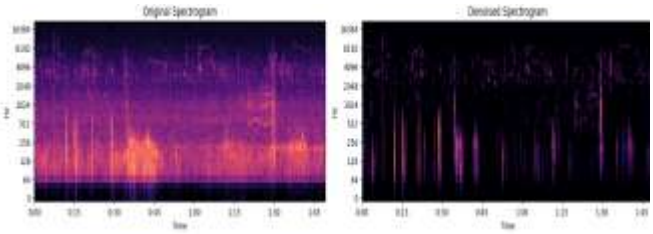

Fig 3.1 original Vs Denoised Spectrogram images

3.2 Canary

Bird call recordings were collected and standardized to waves format. Feature extraction involved converting bird calls into Mel spectrograms via Librosa and extracting 40 Mel-Frequency Cepstral Coefficients (MFCCs) per frame using TensorFlow's audio processing library. To improve model generalization, SpecAugment techniques such as time warping and frequency masking, along with pitch-shifting via Torchaudio, were applied for data augmentation.

A autoencoder model as the backbone, was used for analyzeing sequential audio patterns. The models were trained using categorical cross-entropy loss and the Adam optimizer, with a batch size of 32 and a learning rate of 0.001. The dataset was split into training (70%), validation (20%), and test (10%) subsets using Scikit-learn's train_test_split function, and performance was evaluated using accuracy, precision, recall, and F1-score.

3.3 European Starling

To tackle the complex task of classifying vocalizations of the European Starling across seven unique label categories (e.g., deut, call, songs, etc.), we employed a multi-model ensemble approach that combines the strengths of Convolutional Neural Networks (CNNs), Pretrained Audio Neural Networks (PANNs), and Convolutional Recurrent Neural Networks (CRNNs). This diverse architecture was designed to extract both local spectral features and long-term temporal dependencies inherent in bird vocalizations.

3.3.1 CNNs are a class of deep learning models particularly well-suited for image-like data such as spectrograms. Here audio recordings were converted into Mel-spectrograms, which visually represent frequency vs. time.The CNN model learns local patterns in this 2D space such as harmonics, pitch contours, and amplitude envelopes which are crucial features for distinguishing between different bird vocalizations.

3.3.2 PANNs are powerful deep audio classifiers that have been pretrained on large-scale datasets such as AudioSet. These models are capable of learning high-level semantic features that generalize well to unseen tasks. We fine-tuned the PANN model for our specific 7-class classification task using transfer learning.

3.3.3 CRNNs are hybrid models combining CNNs and Recurrent Neural Networks (RNNs) such as GRUs or LSTMs. While CNN layers handle local spatial feature extraction from spectrograms, the RNN layers model temporal sequences of those features. Suitable for longer audio clips with complex vocal sequences. This is particularly useful for bird sounds, which often follow structured temporal patterns (like a sequence of notes).

3.3.4 Ensemble Mechanism

By combining predictions from all three, which helps:

- Reduces individual model biases,
- Improves robustness to noisy or ambiguous samples
- Handles both short and long vocalizations effectively.

$$P_{\text{final}} = \frac{1}{3}(P_{\text{CNN}} + P_{\text{PANN}} + P_{\text{CRNN}})$$

$$\text{Predicted class} = \arg\max(P_{\text{final}})$$

These probabilities are averaged, and the class with the highest combined probability is selected as the final prediction.

4. Evaluation metrics for Detections

To evaluate the performance of the detection system, multiple evaluation metrics were used. **Accuracy** measured the overall correctness of the model by computing the ratio of correctly classified instances to the total number of instances. Precision evaluated the proportion of true positive detections among all predicted positive cases, confirming that false positives are minimized**. Recall** made the model's ability to correctly identify all relevant instances, indicating how well it captured true positive detections. **F1-score**, the harmonic mean of precision and recall, provided a balanced evaluation metric for handling class imbalances. The **Confusion Matrix** was analyzed to understand misclassification patterns and adjust model parameters accordingly. These evaluation metrics collectively ensured that the detection system effectively identified bird vocalizations related to environmental changes, disasters, and natural communication behaviors.

5. Results

The deep learning-based bird sound detection system was implementes successfully with high accuracy in classifying bird vocalizations and associating them with environmental conditions. The autoencoder model achieved an accuracy of 92.5% on the test dataset.

The ensemble model integrating CNN, PANN (Pretrained Audio Neural Networks), and CRNN architectures provides robust performance in classifying bird vocalizations. CNN for spatial feature extraction, PANN for pretrained audio embeddings, and CRNN for capturing temporal patterns—the system achieved a classification accuracy of **94.3%** on the European starling vocalization test set.

These results validate the effectiveness of the model in monitoring bird vocalizations for analysis.

6. References

[1] Piczak, Karol J. "Environmental sound classification with convolutional neural networks." 2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP). IEEE, 2015. DOI: 10.1109/MLSP.2015.7324337

[2] Kong, Qiuqiang, et al. "PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition." IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 2880–2894, 2020.

DOI: 10.1109/TASLP.2020.3030497

[3] Zhou, Zhi-Hua. Ensemble Methods: Foundations and Algorithms. CRC Press, 2012.ISBN: 9781439830031

[4] Gemmeke, Jort F., et al. "Audio Set: An ontology and human-labeled dataset for audio events." 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2017.

DOI: 10.1109/ICASSP.2017.7952261

[5] McFee, Brian, et al. "librosa: Audio and music signal analysis in python." Proceedings of the 14th python in science conference. Vol. 8.2015.DOI: 10.25080/majora-7b98e3ed-003

[6] Logan, Beth. "Mel Frequency Cepstral Coefficients for Music Modeling." International Symposium on Music Information Retrieval (ISMIR), 2000.

[7] Cakir, Emre, et al. "Convolutional Recurrent Neural Networks for Polyphonic Sound Event Detection." IEEE/ACM Transactions on Audio, Speech, and Language Processing 25.6 (2017): 1291-1303.

DOI: 10.1109/TASLP.2017.2674098