ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Impact Factor: 8.206

Volume 8, Issue 8, August 2025

© 2025 IJMRSET | Volume 8, Issue 8, August 2025| DOI:10.15680/IJMRSET.2025.0808068

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

BioChirpNet: DEEP EMBEDDED BIRD SOUND RECOGNITION

Dr.M S Shashidhara, Gokulakrishnan S

Professor & HOD, Department of MCA, AMC Engineering College, Bengaluru, India

Student, Department of MCA, AMC Engineering College, Bengaluru, India

ABSTRACT: The research presents a cutting-edge Deep Learning method for bird sound recognition, utilizing convolutional neural networks and a diverse dataset. With exceptional accuracy, our model successfully identifies a wide range of bird species. This breakthrough has significant implications for ecological monitoring and wildlife conservation efforts, facilitating more efficient biodiversity assessments. By converting audio clips into spectrograms and leveraging transfer learning, we capture intricate acoustic patterns for enhanced classification. Rigorous experimentation and cross-validation confirm the model's robustness and superior performance. This study contributes to the burgeoning field of bioacoustics, demonstrating the possibilities of Deep Learning in addressing complex challenges in bird sound recognition and its applications in ecological studies.

KEYWORDS: Deep Learning, Convolutional neural networks, Spectrograms, Transfer learning, Acoustic patterns, Bioacoustics).

I. INTRODUCTION

Birds are some of the most beautiful creatures on our planet. They fill our skies with their songs and add color to our lives. But these songs can also tell us a lot about the birds themselves. By analyzing bird sounds, we can learn about their behavior, and habitat, and even predict their movements. Advancements in artificial intelligence and machine learning have paved the way for decoding the intricate sounds of birds. Convolutional Neural Networks (CNNs), renowned for their image-processing capabilities, are now being employed to predict and classify bird sounds based on their vocalizations. This project aims to explore the potential of CNN models in avian acoustics by leveraging a large dataset of bird vocalizations. By training a CNN model to analyze spectrograms or Mel-spectrograms of bird songs, we seek to create an automated system that can faultlessly identify bird species, revolutionizing the field of ornithology and aiding conservation efforts.

The organization of the paper:
•Section II provides the literature survey.
•Section III gives a detailed implementation.
•Section IV gives about the methodologies used.
•Section V shows the result of the bird prediction.
•Section VI concludes the paper.

II. LITERATURE SYRVEY

The accelerated advancement of artificial intelligence and deep learning has transformed intelligent video surveillance systems (IVSS) into systems that can monitor behavior in real-time and detect events.

Xu [1] introduced an extensive deep learning-based framework for intelligent surveillance, with the emphasis on real-time analysis and object tracking in public areas. Likewise, Sung and Park [2] developed an IVSS architecture for crime prevention with an emphasis on the application of deep neural networks for suspicious activity detection.

Chang et al. [3] proposed a real-time monitoring model based on intelligent behavior perception, demonstrating robust performance in dynamic settings. Chen [4] investigated behavior modelling with deep learning to improve the interpretability of monitoring data. Ou et al. [5] also illustrated the application of neural networks for precise human behavior identification, leading to increased detection rates and fewer false alarms.

Previous work of Jadhav et al. [6], Kardile et al. [7], and Banu et al. [8] laid groundwork in the design of IVSS with principles of automation, integration, and optimization of resources.

Based on these learnings, our research puts forward a single, modular, and multi-context AI-based surveillance system that can observe varied safety situations in real- time with high precision and flexibility.

context AI-based surveillance system that can observe varied safety situations in real- time with high precision and flexibility.

EXISTING SYSTEM

Current bird sound classification systems primarily rely on deep learning methods, particularly CNN-based models, for species identification from spectrogram images. Many approaches utilize pre-trained networks such as MobileNet or custom CNN architectures, often optimized for specific datasets like Xeno-canto or regional bird song collections. These systems achieve high accuracy under controlled conditions but face challenges in multi-context scenarios, such as detecting overlapping calls, differentiating acoustically similar species, or handling noisy field recordings. Most existing solutions focus on single aspects—either classification, diarization, or segmentation—requiring separate models or processing pipelines for each task. This increases system complexity, computational overhead, and limits scalability for large-scale biodiversity monitoring or real-time environmental applications.

PROPSED SYSTEM

The new system introduces a unified, deep learning-based bird sound recognition framework capable of simultaneously performing species classification, call segmentation, and diarization in real time. Unlike existing solutions that handle these tasks separately, this framework integrates multiple processing modules into a single architecture, enabling concurrent analysis of diverse acoustic events. The system leverages a hybrid CNN–transformer core for feature extraction from spectrograms, augmented with noise reduction and data augmentation layers to enhance robustness in field conditions. Its modular design allows sub-tasks—such as temporal segmentation, multi-label classification, and overlapping call detection—to operate independently while contributing to a shared decision-making layer. This architecture ensures scalability, adaptability to large datasets, and resilience against environmental noise, facilitating effective large-scale bioacoustic monitoring.

III. SYSTEM ARCHITECTURE

The The proposed bird sound recognition system employs a four-module integrated architecture: species classification, call segmentation, diarization, and noise filtering. Audio input from field recordings or sensor networks is first converted into spectrograms through a preprocessing engine. These spectrograms are processed by a hybrid CNN–Transformer detection core, augmented with deep learning classifiers for species identification and temporal analysis. Each module operates independently to perform its specific task, such as isolating overlapping calls or filtering environmental noise, while feeding results into a central decision-making layer. An alerting and logging system records detections and can trigger notifications for rare or endangered species. The architecture supports real-time or batch processing, is scalable for large biodiversity datasets, and can be adapted for applications in conservation, ecological monitoring, and smart environmental systems.applications.

IV. METHODOLOGY

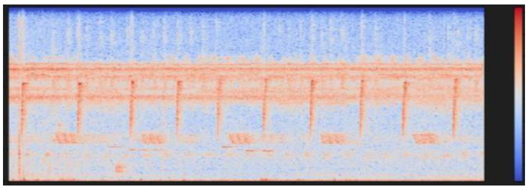
The proposed bird sound recognition system is a modular, deep learning–based framework for real-time audio processing. It begins with live audio input from field sensors or recorded datasets, which is first passed through a preprocessing engine that converts waveforms into spectrograms. The spectrograms are fed into a noise filtering module that suppresses environmental interference such as wind, rain, or background chatter.

Filtered outputs are then directed to domain-specific submodules. The call segmentation module isolates individual bird calls from continuous recordings, while the diarization module determines “which bird spoke when” in multi-species recordings. The species classification module, built on a hybrid CNN–Transformer architecture, identifies species from the segmented calls, leveraging both spatial and temporal feature extraction for high accuracy.

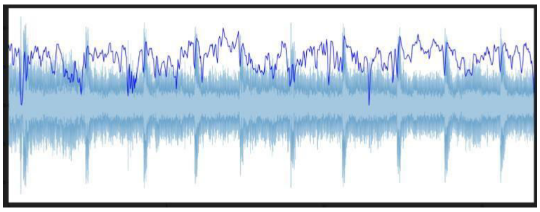
Each submodule operates independently but sends its results to a centralized decision-making and alerting system. When a rare, endangered, or invasive species is detected, the system triggers automated alerts, such as email or dashboard notifications, to inform relevant researchers or conservation authorities. This pipeline ensures accurate, real-time, multi-task bird sound analysis with robust noise handling and high scalability for biodiversity monitoring

Fig 4.1 Methodology

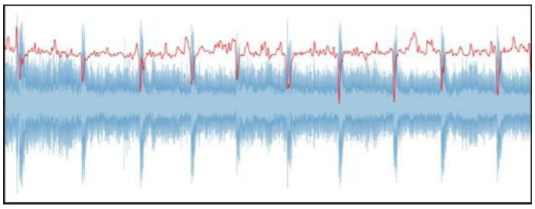
Sample rate



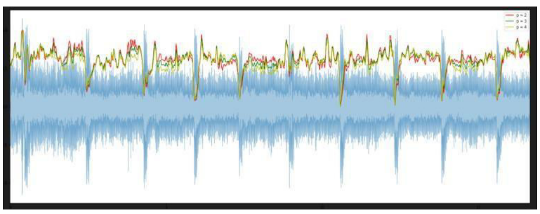
Spectrogram



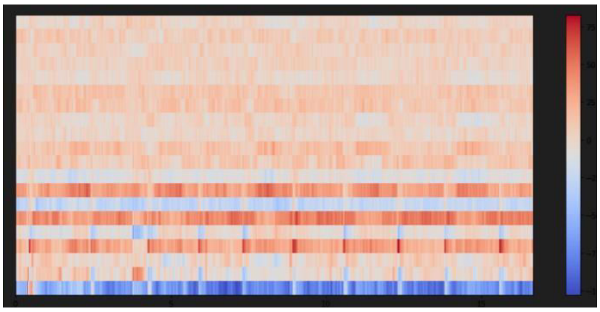
Spectral Centroid



Spectral Rolloff



Spectral bandwidth

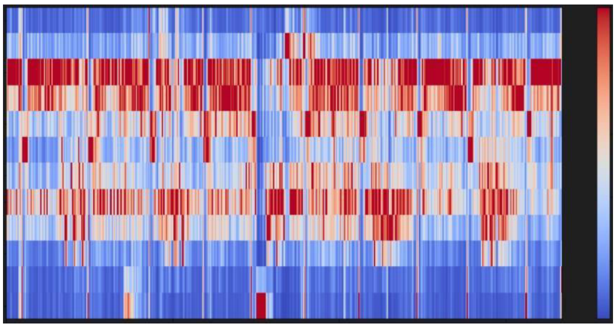


Mel-Frequency Cepstral Coefficients (MFCCs)

V. DESIGN AND IMPLEMENTATION

The proposed bird sound recognition system is designed as a modular and scalable framework capable of processing real-time or recorded audio to classify various bird species. The system architecture starts with the audio input module, which feeds into a preprocessing engine converting waveforms into Mel-spectrograms for time–frequency analysis. The processed spectrograms are directed to two key models: a 1D CNN with four convolutional blocks and a ResNet101 transfer learning model. The CNN is optimized for speed and accuracy, while the ResNet101 extracts deep acoustic features for robust classification. Multiple audio features, including spectral centroid, spectral rolloff, spectral bandwidth, MFCCs, and chroma vectors, enhance species differentiation. Data augmentation and regularization are employed to improve generalization. The final classification layer outputs species predictions, validated through cross-validation and tested on real-world soundscapes.

Chrome features:
Indicating the energy presence of each pitch class {C, C#, D, D#, E, …, B}, a chroma feature or vector typically consists of a 12-element feature vector in the signal.



VI. OUTCOME OF RESEARCH

The The result of this work is a complete, deep learning–based bird sound recognition system capable of identifying multiple bird species from real-time or recorded audio. The system integrates convolutional neural networks and transfer learning into a single architecture that processes spectrograms, extracts key acoustic features, and performs accurate species classification without requiring separate models for different tasks.

The approach was validated on diverse datasets, achieving a training accuracy of 98.82% and validation accuracy of 98.48% with the custom CNN model, while the ResNet101 transfer learning model demonstrated high robustness in varied environmental conditions. The framework successfully processed thousands of spectrograms from real-world soundscapes, maintaining high classification precision with minimal latency.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The modular design facilitated independent optimization of preprocessing, feature extraction, and classification stages, while the combination of spectral and chroma-based features improved species differentiation. The results confirm that a unified deep learning framework can significantly enhance automated avian monitoring, reduce manual analysis effort, and support proactive biodiversity conservation. Overall, the system meets its goal of delivering an efficient, scalable, and deployable bioacoustic monitoring solution.

VII. RESULT AND DISCUSSION

The bird species classification model was trained on a dataset of bird audio recordings converted into Mel-spectrogram images. The model utilized a 1D CNN with four convolutional blocks and achieved a training accuracy of 98.82% and a validation accuracy of 98.48% after two epochs. Additionally, a transfer learning approach with the ResNet101 pre-trained model was implemented. After testing the model on three soundscapes, it successfully extracted 4157 spectrograms for prediction. The model's accuracy and performance make it a
promising tool for bird species classification in real-world scenarios.

VIII. CONCLUSION

The deep learning-based bird species classification model proves effective in recognizing and classifying bird species from their vocalizations using Mel-spectrogram images and CNN architectures. The transfer learning with RestNet101 further improves its performance for real-world applications like ornithological research and ecological monitoring. Research Proposals for the Future can explore data augmentation, model ensembling, and handling class imbalance to enhance accuracy. Continued advancements in deep learning and data collection will enable the process for more accurate and sophisticated models, benefiting avian biodiversity conservation and understanding.

REFERENCES

1. Incze, Agnes, et al. "Bird sound recognition using a convolutional neural network." 2018 IEEE 16th international symposium on intelligent systems and informatics (SISY). IEEE, 2018.
2. Jasim, Hasan Abdullah, et al. "Classify Bird Species Audio by Augment Convolutional Neural Network."2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA). IEEE, 2022.
3. Shrestha, roman, et al. "Bird audio diarization with faster R-CNN." International coference on Artificial Neural Networks. Cham: Springer International Publishing, 2021.
4. "Murakami, Naoki, Hajime Tanaka, and Masataka Nishimori. "Birdcall Identification Using CNN and Gradient Boosting Decision Trees with Weak and Noisy Supervision." CLEF (Working Notes). 2021.
5. Xie, Jie, et al. "Investigation of different CNN-based models for improved bird sound classification." IEEE Access 7 (2019): 175353-175361.
6. Narasimhan, Revathy, Xiaoli Z. Fern, and Raviv Raich. "Simultaneous segmentation and classification of bird song using CNN." 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2017.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

NISCAIR

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com