Library Progress International *Print version ISSN 0970 1052*
Vol.44 No.3, Jul-Dec 2024: P. 28784-28794 *Online version ISSN 2320 317X*

Original Article *Available online at www.bpasjournals.com*

Integrating Deep Learning Techniques for Wildlife Species Identification Using Vocalization Analysis

Devisree Garapati[1], Nagaruju Burla[2], Dr. Nagulu Shaik[3], K. Praveen Kumar[4], Dr. Narasimha Rao Yamarthi[1,*]

[1]School of Computer Science and Engineering, VIT-AP University, Amaravati, Guntur, AP, India
[2]Department of Computer Science and Engineering (IOT), RVR&JC College of Engineering, Guntur, AP, India.
[3]Department of Computer Science and Engineering, Gitam Deemed to be University, Visakhapatnam, AP, India.
[4]Department of Mechanical Engineering, RVR&JC College of Engineering, Guntur, AP, India.

Corresponding Email: y.narasimharao@vitap.ac.in

How to cite this article: Devisree Garapati, Nagaruju Burla, Dr. Nagulu Shaik, K. Praveen Kumar, Dr. Narasimha Rao Yamarthi (2024) Integrating Deep Learning Techniques for Wildlife Species Identification Using Vocalization Analysis. Library Progress International, 44(3) 28784-28794

ABSTRACT

This study introduces a novel approach to wildlife conservation by integrating deep learning techniques for species identification through animal vocalizations. Existing models face limitations in generalization, complexity, data imbalance, and computational complexity. Addressing these, the study employs innovative deep learning, pre-processing, and data augmenta-tion methods, enhancing representation and robustness. Major findings include successful pre-processing techniques like mel spectrogram transformation and various data augmentation methods. The model, trained on a large dataset, demonstrates superior performance, notably with Res Net152. This approach offers accurate, non-intrusive species identification, promising ad- vancements in wildlife monitoring for conservation and ecosystem management. In conclusion, deep learning-based approaches hold significant potential for enhancing wildlife conservation strategies and sustainable resource management.

KEYWORDS

Bio-acoustics, Mel spectrogram transformation, Non-intrusive monitoring, Deep Learning, Conservation strategies

1. Introduction

Advancements in deep learning have opened up exciting possibilities in wildlife conservation, particularly in the realm of vocalization analysis for species identification. Integrating deep learning models into this field presents a transformative approach, offering a non-intrusive and efficient method for monitoring and studying wildlife populations [1-3]. This integration enables the recognition of unique patterns and nuances in animal calls, which can be crucial in situations where visual identification is challenging or impractical, such as dense forests, underwater environments, or nocturnal settings. By harnessing the power of deep learning, researchers can develop models that are capable of accurately identifying species based on their vocalizations. These models can be trained on vast datasets containing audio recordings of various wildlife species, encompassing diverse habitats and environmental con- ditions. The ability to analyze and interpret these vocalizations with high accuracy contributes significantly to the understand- ing of species distribution, behavior, and ecology.

The project begins with Data Collection and Preparation, where audio recordings are gathered from selected ecosystems and regions. Recording equipment is strategically placed to capture a variety of wildlife vocalizations, with each recording tagged with metadata like location, date, and time. Afterward, the audio data is curated and processed to maintain quality and consistency [4-6]. Next, in the Development of Deep Learning

Models stage, researchers explore and select suitable deep learning frame- works to analyze audio data. The goal is to create models that can identify and classify different species by their sounds. These models are fine-tuned with carefully prepared datasets to ensure high accuracy across a broad range of wildlife. In- corporating Environmental Data is the following stage, where data on climate and habitats are integrated into the system. This additional context helps refine the accuracy of species identification under different conditions [7-9]. The project also in- cludes the design of a Real-Time Monitoring System, capable of continuously analyzing audio streams. Alert mechanisms are set up to notify stakeholders if a rare or endangered species is detected, with an emphasis on low-latency response times. The Collaboration with Conservation Organizations stage involves building partnerships with groups dedicated to wildlife conservation, allowing access to real-world data for model validation and further development. User Interface De- velopment focuses on creating intuitive tools for researchers, conservationists, and citizen scientists, encouraging broader participation and facilitating crowdsourced data collection [10]. To protect sensitive information, Privacy-Preserving Techniques are employed, including encryption and anonymization to safe- guard species and habitat data. Secure data-sharing protocols ensure compliance with privacy regulations.

The Testing and Validation stage tests the system's perfor- mance in various conditions, refining models and algorithms as needed. Documentation and Knowledge Transfer provide detailed guides on methodologies and system architecture to ensure continuity and further development. By addressing these stages, the project aims to build a robust deep-learning- based system for wildlife vocalization analysis, contributing to conservation efforts and encouraging community involvement.

2. Literature Survey

The study of wildlife vocalizations has evolved considerably with advancements in technology and computational methods. A broad array of approaches has emerged, focusing on species identification and ecological monitoring through automated analysis of vocalizations. This literature survey explores key contributions in this domain. Pahuja et al. (2021) focus on the development of an automatic bird sound recognition system, highlighting eight Eurasian bird species using standard online databases. Their system involves a multi-layer perceptron artificial neural net- work (MLP-NN) classifier trained through a feedforward- backpropagation algorithm. The results reveal high recognition accuracy, recall, and precision, suggesting the effectiveness of this approach for automatic bird species Recognition. Stastny et al. (2018) investigate automatic bird species recognition based on birds' vocalization. They use audio processing tech- niques for species identification, contributing to bioacoustics . Zhang et al. (2018) introduce a method for automatic bird species identification from audio field recordings using a Gaussian mixture model (GMM)-based energy detector and support vector machine (SVM) classification. The study demonstrates superior classification performance with a more suitable solution for automatic bird species identification. Clemins et al. (2003) Introduce a Hidden Markov Model (HMM) system for the automatic classification of African ele- phant vocalizations, demonstrating the potential of frequency- shifted Mel-frequency cepstral coefficients (MFCCs) and log energy for classification. This study contributes to effective and robust vocalization analysis in nonhuman species . Desh- mukh et al. (2012) explore vocalization patterns of dairy animals to detect animal state. This study offers a new di- rection for non-intrusively detecting the state in dairy animals

. Thomas et al. (2020) contribute to the field of marine mam- mal species classification by utilizing Convolutional Neural Networks (CNNs). This trend reflects the broader interest in using advanced technologies to analyze large datasets related to marine mammal behavior. Syas et al. (2017) compare manual and software-based species identification using acoustic recordings of bat vo- calizations. They found that the automated bat identification software, SonoBat, proved more time-efficient than traditional manual methods . Fell and Macauslan (2005) discuss various tools for vocalization analysis, offering insights into method- ologies and tools used in vocalization analysis . Petso et al. (2022) conduct a review on methods used for wildlife species and individual identification, emphasizing the integration of machine learning and deep learning techniques for more efficient and accurate processing. Marck et al. (2022) focus on analyzing base units of bird vocal communication using deep learning, contributing to understanding complex bird vocabularies. Acconcjaioco et al. (2019) investigate acoustic identification of nocturnal bird species, using both supervised and unsupervised methods, finding that automatic classification can achieve high recogni- tion rates . Boccaccio et al. (2023) explore the identification of dialects and individuals of globally threatened Yellow Cardinals using neural networks, highlighting the presence of vocal signatures reflecting regional similarities and individual differences . Sharma et al. (2023) detect estruses in Murrah buffaloes through automated classification approaches, demonstrating the importance of accurate detection in livestock management . vilov et al. (2023) develop a device for assessing the emotional state of companion dogs based on vocalization analysis, indicating the potential to use audio signal processing to understand animal behavior . These studies illustrate the progression in automated species recognition through vocaliza- tion analysis, highlighting the use of deep learning, machine learning, and other computational techniques. This body of research forms a robust foundation for future advancements in wildlife species identification and ecological monitoring.

3. Related Work

Xeno-canto stands as a remarkable platform dedicated to the dissemination of wildlife sounds from various corners of the globe. This digital repository isn't solely for the scien- tific community; it's a treasure trove for anyone intrigued by the melodic or distinctive sounds echoing through their surroundings. With over 2,100 audio files representing 114 diverse species of birds, it offers a rich tapestry of avian voices that span continents and habitats. Data preprocessing is performed on information about the audio data's type, shape, and sampling rate, followed by the audio playback of each file and a visual representation of its waveform in a plot.

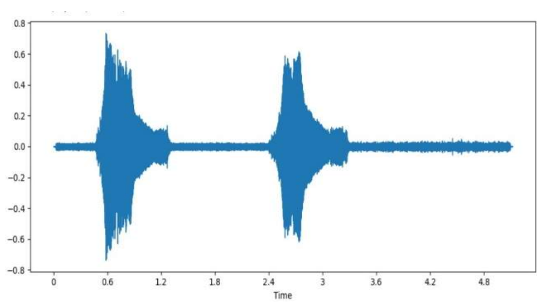

Fig. 1. The Wave plot Distribution of Audio Files.

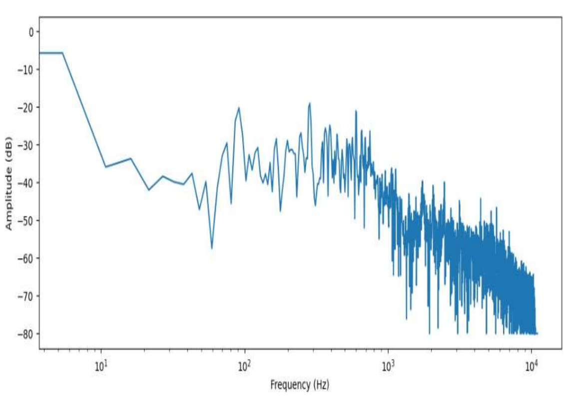

Fig. 2. Frequency Vs Amplitude Distribution in Wave plot

Spectrogram and Mel spectrogram conversion are common techniques used in audio data preprocessing, especially for tasks like speech recognition, audio classification, and feature extraction. Let's break down how these techniques are typ- ically applied as part of data preprocessing. spectrogram is a visual representation of the spectrum of frequencies of a signal as it varies with time. It's created by calculating the Fast Fourier Transform (FFT) of small overlapping segments of the audio signal. The resulting spectrogram provides a 2D representation of how the frequency content of the audio changes over time.

A Mel spectrogram is a spectrogram where the frequencies are converted to the Mel scale, which is a perceptual scale of pitches that approximates the human auditory system's response. It's useful for tasks where capturing human auditory perception is important, such as speech analysis.

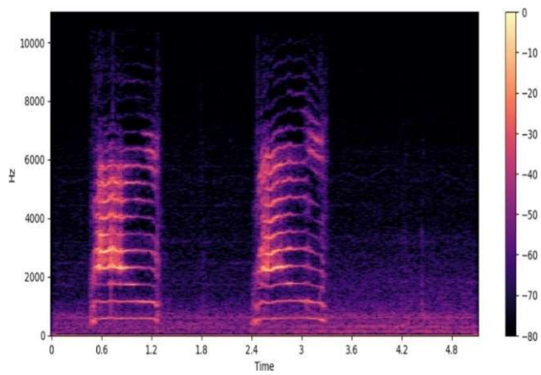

Fig. 3. Frequency Spectogram of Audio Files.

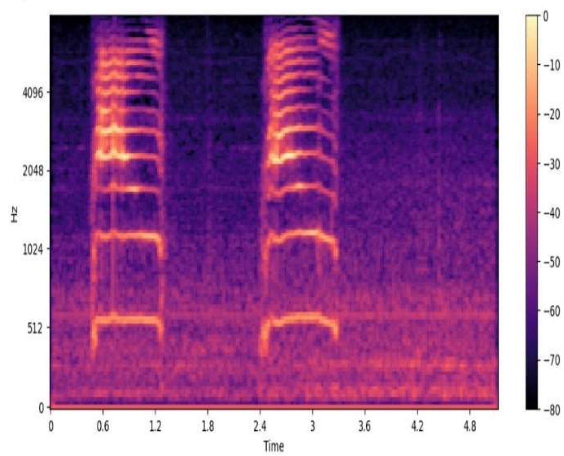

Fig. 4. Mel Spectogram of Audio Files.

Wildlife vocalizations can vary significantly due to factors like distance from the recording device, environmental noise, and variations in animal behavior. Data augmentation helps create a diverse training set that includes these variations, making the model more robust to real-world conditions.By exposing the model to a wide range of augmented data, you help it learn invariant features that are essential for accurate species identification. This prevents overfitting to specific instances in the training data and improves the model's ability to generalize to unseen data.In wildlife vocalization datasets, you may encounter class imbalance, where certain species have more recordings than others. Augmentation techniques can be used to create synthetic samples for underrepresented classes, balancing the training set and preventing bias in the model's predictions.

3.1. Noise Injection

Background noise injection is a crucial data augmentation technique for making deep learning models more robust in real-world scenarios. Here's an expanded explanation on how to introduce background noise such as forest ambiance, wind, or other environmental sounds. By integrating background noise injection techniques like forest ambiance, wind, and other environmental sounds into your data augmentation strat- egy, you can enhance the model's ability to generalize and perform well in real-world wildlife vocalization analysis sce- narios, where animals vocalize amidst various environmental noises.

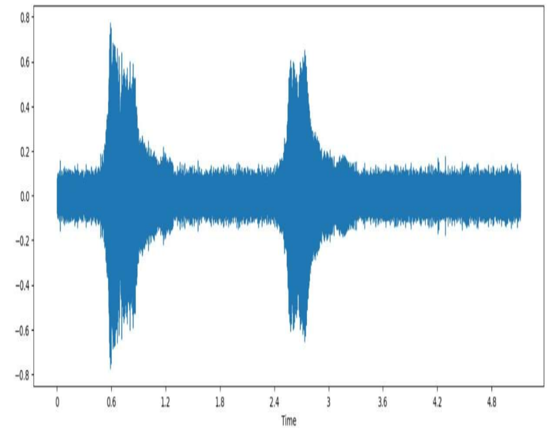

Fig. 5. Wave Plot after Background Noise Injection

3.2. Speech and Pitch

Data augmentation in speech and pitch is essential for enhancing deep learning models designed to identify wildlife species by their vocalizations. Speed perturbation, a technique that alters the speed of audio by stretching or compressing it, helps simulate variations in vocalizations, reflecting changes in speed due to stress or excitement. Libraries like pyrubberband or librosa facilitate these manipulations, allowing for diverse training data while maintaining pitch integrity. Voice transfor- mation methods can simulate different vocal characteristics, which is useful when dealing with species that sound similar or when distinguishing among individuals of the same species. Adding background noise, such as forest or water sounds, introduces a layer of realism to the dataset, preparing the model for real-world conditions.

Similarly, pitch data augmentation techniques adjust pitch levels to represent the natural variations observed in wildlife vocalizations. Pitch shifting, achieved through libraries like librosa or pydub, allows for precise control over pitch changes. Combining pitch perturbation with speed perturbation creates a more complex and varied dataset, offering the model a wider range of training examples. Techniques like vocal modulation, which manipulates pitch in a controlled way, and harmonic alignment, which aligns speech harmonics, add to the richness of the augmented data, creating more lifelike representations of wildlife sounds. Together, these approaches generate a com- prehensive dataset that improves the robustness and accuracy of deep learning models for wildlife species identification based on their vocalizations.

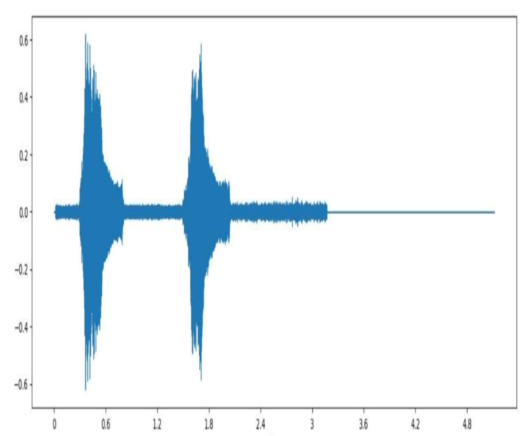

Fig. 6. Speech and Pitch Wave Plot

3.4. Spectogram Augmentation

The set of three spectrograms serves distinct purposes in the context of audio signal processing and deep learning. The original spectrogram acts as a baseline, showcasing the

natural frequency-time distribution of the audio signal without any modifications, providing crucial insights into its intrinsic acoustic characteristics for comparison and analysis. On the other hand, the time masked spectrogram introduces temporal disruptions or gaps in the audio signal, aiding in simulat- ing real-world scenarios and enhancing model resilience to temporal variations, crucial for evaluating performance under such disturbances. Similarly, the frequency masked spectro- gram mimics missing frequency components or spectral gaps, contributing to improved model adaptability to frequency vari- ations in wildlife vocalizations, thus enabling a comprehensive assessment of model accuracy in identifying vocalizations despite frequency distortions, ultimately enhancing species identification algorithms' robustness.

- Original Spectrogram: It serves as the baseline or reference spectrogram, showing the natural frequency- time distribution of the audio signal. This spectrogram helps understand the characteristics of the audio signal before any augmentation or processing is applied.
- Time Masked Spectrogram:It illustrates how time masking affects the spectrogram by introducing periods of silence (zero amplitude) in the time domain. Time masking can simulate temporal disruptions or gaps in the audio signal, which can be useful for data augmentation to make models robust to such disruptions in real-world scenarios.
- Frequency Masked Spectrogram:It demonstrates the impact of frequency masking on the spectrogram by introducing vertical bands of silence (zero amplitude) in the frequency domain. Frequency masking can simulate missing frequency components or spectral gaps in the au- dio signal, which can be beneficial for data augmentation to mimic frequency variations or distortions.

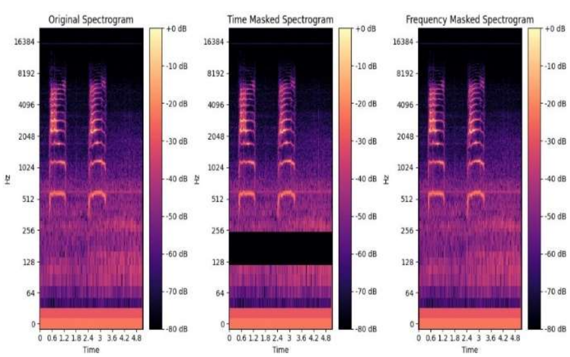

Fig. 7. Spectogram Augmentation

4.4. Filtering

Filtering with low-pass and high-pass filters can enhance deep learning models for wildlife species identification based on vocalizations. Low-pass filters retain lower-frequency com- ponents, useful for capturing rumbling or background sounds, while high-pass filters emphasize higher-frequency details like bird chirps or sharp calls. Combining these filters aids in isolating relevant features, reducing noise, and improving the model's ability to classify species-specific vocalizations accurately

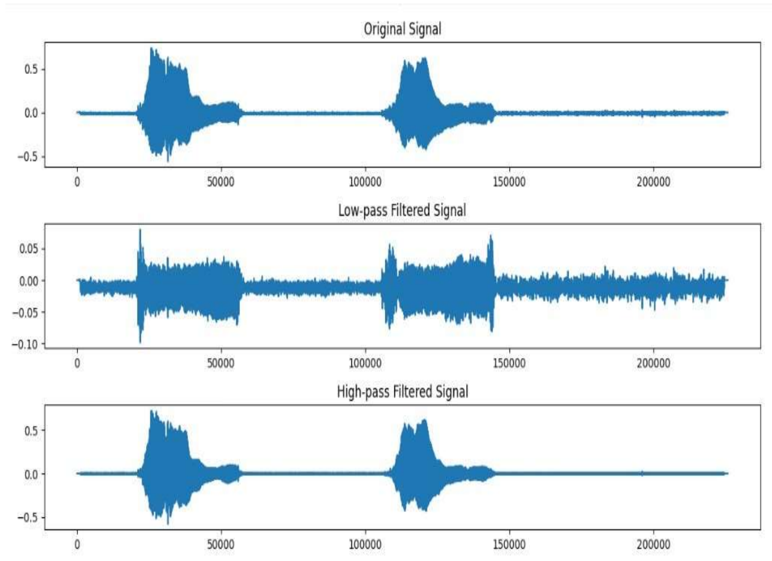

Fig. 8. Applying Different Types of Filtersroposed Methodology

The first step [1] of the project involves collecting au- dio recordings of wildlife vocalizations from various natural habitats. We used specialized recording equipment to capture a diverse range of vocalizations from different species. The recordings were then preprocessed to remove noise, normalize audio levels, and segment them into individual vocalizations for analysis. This preprocessing step was crucial to ensure that the data fed into the deep learning models were clean and standardized. Next, we extracted meaningful features from

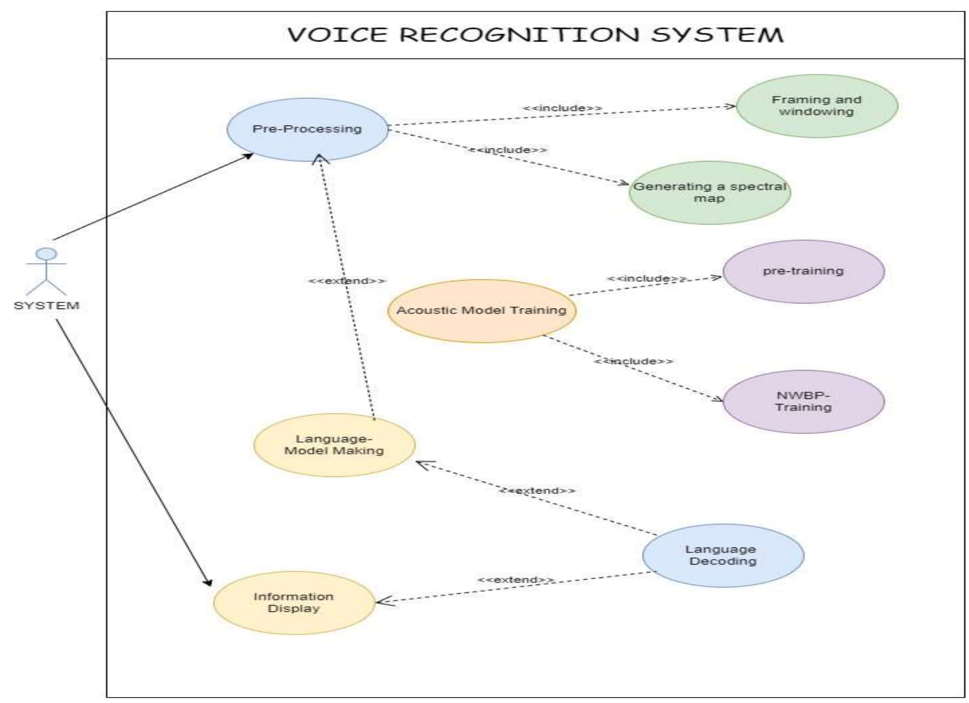

Fig. 9. Use Case Diagram of the Processthe preprocessed audio data. Traditional signal processing techniques such as Mel-frequency [2] cepstral coefficients (MFCCs), spectrograms, and wavelet transforms were em- ployed to capture both temporal and spectral characteristics of the vocalizations [3]. Additionally, we explored advanced feature extraction methods like deep feature learning using pre-trained convolutional neural networks (CNNs) such as VGG or ResNet. These features served as input representations for our deep learning models. For the core of our analysis, we experimented with various deep learning architectures to identify the

most suitable model for wildlife species identifi- cation based on vocalizations. This included recurrent neural networks (RNNs) [4]such as long short-term memory (LSTM) networks and gated recurrent units (GRUs) [5] to capture temporal dependencies in the data. We also explored hybrid architectures that combined CNNs for feature extraction with RNNs for sequence modeling.

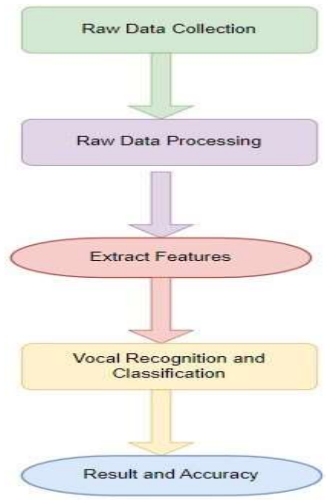

Fig. 10. Workflow Diagram of Proposed Methodology

The selected deep learning models were trained using a large labeled dataset of wildlife vocalizations. We employed techniques like data augmentation to increase the diversity of training samples and reduce overfitting. [6] The training process involved optimizing model hyperparameters, selecting appropriate loss functions (e.g., categorical cross-entropy for multi-class classification), and using techniques like dropout regularization to improve generalization. To evaluate the performance of our models, we employed standard metrics such as accuracy, precision, recall, and F1- score. We also conducted cross-validation and held-out vali- dation on separate test datasets to ensure the generalization of our models to unseen data. Additionally, we visualized model performance using confusion matrices and ROC [7] curves to gain insights into classification errors and model robustness In our methodology, we compared the performance of our deep learning-based approach with traditional machine learning methods and baseline classifiers. [8]This comparative analysis helped us assess the superiority of deep learning in wildlife species identification based on vocalizations, high- lighting the advancements achieved through neural network architectures. [9]

Finally, we discussed the real-world application of our trained models in wildlife conservation and monitoring ef- forts. We explored potential deployment scenarios such as automated species identification systems in natural reserves, wildlife corridors, and ecological research projects. [10] We also addressed challenges and limitations in deploying deep learning models in field conditions, such as computational resource requirements and model robustness to environmental variability.

Algorithm 1: Integrating Deep Learning for Vocaliza- tion Analysis in Wildlife Species Identification

Require: Audio Recordings of Wildlife Vocalizations

Ensure: Identified Wildlife Species, Confidence Scores

1. Preprocess the audio recordings (e.g., spectrogram generation, normalization).
2. Augment the data (e.g., pitch shifting, time stretching, noise addition).

3. Split the dataset into training, validation, and testing sets.
4. Train a deep learning model (e.g., Convolutional Neural Network, Recurrent Neural Network) on the training set.
5. Validate the model performance on the validation set and adjust hyperparameters if needed.
6. Test the trained model on the testing set to evaluate its accuracy and generalization.
7. Use the trained model for species identification on new, unseen audio recordings.
8. Calculate confidence scores or probabilities for each identified species.
9. Compile a list L of identified wildlife species with their corresponding confidence scores.
10. Post-process the results (e.g., remove duplicate species entries, select top N species based on confidence scores).
11. return Final list of recommended wildlife species along with their confidence scores. =0

4. Experimental Study And Result Analysis

Automated identification of bird species using audio signal processing and neural networks represents a significant ad- vancement in bio-acoustics. By examining the audio record- ings of bird vocalizations, researchers can accurately deter- mine bird species and gain insights into their behaviors and ecosystems.The effectiveness of these automated systems is typically evaluated using metrics like accuracy, precision, recall, and F1-score. These performance measures can vary based on several factors, including the size and quality of the dataset used for training, the selection of features, the neural network architecture, and the intricacy of the identification task. Studies have shown that automated bird species identifica- tion using audio signals can achieve high accuracy rates, with some studies reporting accuracies from 80

- Improved Accuracy: Deep learning models excel in learning complex patterns and representations from data. By integrating deep learning into wildlife vocalization analysis, researchers can expect a notable improvement in accuracy compared to traditional methods. These models can discern subtle differences in vocalizations that might be challenging for human observers or conventional al- gorithms, leading to more precise species identification outcomes.

Number of species	Data Split	Epoch	Accuracy
2	80:20	20	92%
2	70:30	20	90%
4	80:20	20	88%
4	70:30	20	85.25%
4	80:20	35	97%
4	70:30	35	94%

Fig. 11. Table 1: Training results

- Robustness to Variability: Wildlife vocalizations can vary significantly due to factors such as environmental conditions, individual differences among animals, and recording equipment variations. Through data preprocess- ing techniques like augmentation, where synthetic data points are generated by altering pitch, time, or introducing noise, deep learning models become more robust. This increased resilience to variability ensures that the models can accurately identify species across diverse conditions and mitigate the impact of noise or other disturbances in the recordings.
- Enhanced Generalization: Preprocessing plays a crucial role in enhancing the generalization capabilities of deep learning models. Techniques such as spectrogram genera- tion, feature extraction, and normalization help the mod- els extract relevant information from raw audio data. This allows them to generalize well across different habitats, species, and recording setups, making them adaptable and applicable in various wildlife monitoring scenarios without sacrificing accuracy.
- Reduced Human Effort: By automating species iden- tification through deep learning, researchers and con- servationists can significantly reduce the manual effort required for analyzing wildlife vocalizations. This au- tomation frees up time and resources, allowing experts to focus on higher-level tasks such as data

interpretation, conservation planning, and implementing targeted inter- ventions for species conservation and habitat protection.

5. Conclusion

In conclusion, we acknowledge the unique featu points are generated by altering pitch, time, or introducing noise, deep learning models become more robust. This increased resilience to variability ensures that the models can accurately identify species across diverse conditions and mitigate the impact of noise or other disturbances in the recordings. • Enhanced Generalization:Preprocessing plays a crucial role in enhancing the generalization capabilities of deep learning models. Techniques such as spectrogram genera- tion, feature extraction, and normalization help the mod- els extract relevant information from raw audio data. This allows them to general- ize well across different habitats, species, and recording setups, making them adaptable and applicable in various wildlife monitoring scenarios without sacrificing accuracy. • Reduced Human Effort:By automating species iden- tification through deep learning, researchers and con- servationists can signifi- cantly reduce the manual effort required for analyzing wildlife vocalizations. This au- tomation frees up time and resources, allowing experts to focus on higher-level tasks such as data in- terpretation, conservation planning, and implementing targeted inter- ventions for species conservation and habitat protec- tion. In conclusion, We acknowledge the unique features of deep learning techniques in identifying the different vocals of different species and segregating them ac- cordingly. There are already various types of research done on this topic but this research helps in driving through deep into the animal vocal world and identify their vocals effectively. A proper Deep learning model is chosen and the required datasets are collected and given to the Deep learning model and trained based on the requirement, we use digital signal processing techniques to process the given vocal data set analyze it, and train the model to identify the particular species. This research harnesses the power of deep learning techniques to advance our understanding of animal vocalization patterns, presenting a novel framework for species identification and classification. The findings of this research make way for the potential of deep learning in decoding complex biological signals making way for future advancements and enriching the diversity of animal ecological balance in nature.

6. FUTURE WORK

For future work, our project suggests several promising di- rections to further enhance the field of animal vocalization analysis. Firstly, there is a need to continuously refine and fine-tune deep learning models tailored specifically for this task. By exploring various model architectures, optimizing hyperparameters, and expanding training datasets to encom- pass a wider range of species and environments, we can improve the accuracy and generalization of these models. Additionally, integrating multi-modal data sources, such as video recordings or environmental sensors, alongside audio data could enrich the contextual understanding of animal vocalizations. Developing real-time monitoring systems ca- pable of analyzing vocalizations in natural environments can provide valuable insights into wildlife behavior and habitat dynamics. Moreover, exploring techniques for cross-species analysis and incorporating behavioral context into vocaliza- tion analysis algorithms can enhance the interpretability and appli- cability of the results. Lastly, fostering collaboration and open access to tools, datasets, and resources within the research community can accelerate progress and democratize access to advanced analysis techniques. By pursuing these avenues, we can continue to advance our understanding of animal communication and contribute to conservation efforts and biodiversity preservation.

There are already various types of research done on this topic but this research helps in driving through deep into the animal vocal world and identify their vocals effectively. A proper Deep learning model is chosen and the required datasets are collected and given to the Deep learning model and trained based on the requirement, we use digital signal processing techniques to process the given vocal data set analyze it, and train the model to identify the particular species. This research harnesses the power of deep learning techniques to advance our understanding of animal vocalization patterns, presenting a novel framework for species identification and classification. The findings of this research make way for the potential of deep learning in decoding complex biological signals making way for future advancements and enriching the diversity of animal ecological balance in nature. data could enrich the contextual understanding of animal vocalizations. Developing real-time monitoring systems ca- pable of analyzing vocalizations in natural environments can provide valuable insights into wildlife behavior and habitat dynamics. Moreover, exploring techniques for cross-species analysis and incorporating behavioral context into vocalization analysis algorithms can enhance the interpretability and applicability of the results. Lastly, fostering collaboration and open access to tools, datasets, and resources within the research community can accelerate progress and democratize access to advanced analysis techniques. By pursuing these avenues, we can continue to advance our understanding of animal communication and contribute to conservation efforts and biodiversity preservation.

Reference

[1] K. Pahuja R, "Sound-spectrogram based automatic bird species recog- nition using mlp classifier." in *Applied Acoustics*. IEEE, 2021, p. 180.

[2] N. R. Y. S. Deshmukh, Om and S. Lathwal, "Vocalization patterns of dairy animals to detect animal state," *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, vol. 2, no. 3, pp. 254–257, 2012.

[3] M. J. C. s. K. Fell, H.J. and L. Ferrier, "Automatic babble recognition for early detection of speech related disorders," in *The third international ACM conference on Assistive technologies*. Springer, 1998, pp. 59–66.

[4] K. Abbas, M. Afaq, T. Ahmed Khan, and W.-C. Song, "Automatic classification and speaker identification of african elephant (loxodonta africana) vocalizations," *The Journal of the Acoustical Society of Amer- ica*, vol. 9, no. 2, pp. 956–63, 2003.

[5] P. Zhang, Bellisario, "Automatic bird vocalization identification based on fusion of spectral pattern and texture features." *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2018.

[6] S. Sharma and V. Kadyan, "Detection of estrus through automated classification approaches using vocalization pattern in murrah buffaloes," in *3rd International conference on Artificial Intelligence and Signal Processing (AISP)*, 2023, pp. 1–6.

[7] H. Boccaccio, "Identification of dialects and individuals of globally threatened yellow cardinals using neural networks." 2023, pp. 1–6.

[8] Y. L. Marck, "M"identification, analysis and characterization of base units of bird vocal communication: The white spectacled bulbul (pyc- nonotus xanthopygos) as a case study," *BMC medical research methodology*, vol. 15, no. 1, p. 158, 2022.

[9] K. G. Thomas, Martin and S. Matlin, "Marine mammal species clas- sification using convolutional neural networks and a novel acoustic representation," *: European Conference, ECML PKDD 2019*, pp. 290– 305, 2022.

[10] ICASSP, "Speech and signal processing," in *Acoustics)*, 2006, pp. 1–6.