An Analytical Study of Birds Identification Through the Classification Techniques Based on Birds Call And Song

Ms. K. R. Shanmugapriyaa
Research Scholar Assistant Professor,
Department of Artificial Intelligence and Data Science,
Coimbatore Institute of Technology, Coimbatore
krspriyaa@gmail.com

Dr. Valliappan Raman
Professor
Department of Artificial Intelligence and Data Science,
Coimbatore Institute of Technology, Coimbatore
valliappan@cit.edu.in

Abstract— This study focuses on a convolutional neural network-based deep learning approach for bird sound classification. The primary objective is to interpret the communication patterns of wild bird's species. It involves a comprehensive study on the field of bird bioacoustics, specifically targeting the segmentation of bird syllables from audio recordings and predicting the associated bird species for each identified syllable. Avian vocalizations are captured and transformed into mel-spectrograms, which are well-known for their similarity to human auditory perception. The CNN is trained to adeptly classify various bird sounds, allowing for predictive analysis of anticipated bird vocalizations. The study involves addressing challenges such as background noise, simultaneous vocalizations, and distinguishing between mating calls and songs. The procedural framework involves the initial stages of collecting and converting bird vocalizations into spectrograms. Mel Frequency Cepstral Coefficients (MFCCs) are applied to translate these spectrograms into human auditory signals, known as Mel-spectrograms. The study aims to extend the implementation of the trained model in sanctuaries, zoos, and wildlife areas. This contributes to the preservation and enhancement of biodiversity, bird migration, and mating periods based on their distinctive sounds, and potential threats leading to bird extinction.

Keywords—Spectogram, Mel Frequency Cepstral Coefficients (MFCCs), Convolutional Neural Network (CNN), Predictive analysis

INTRODUCTION

For centuries, humans have marveled the avian vocalizations, but it is only in recent times that technology has enabled us to decipher the nuanced language of our feathered companions. In this paper, we embark on a cinematic journey into the world of bird sound recognition, where the harmonies of nature meet the precision of cutting-edge technology. In recent years the field of bird sound recognition has shown a remarkable witness by using digital signal processing and neural network methodologies. Bird sound recognition holds immense promise not only for the scientific community but also for conservation efforts, ecological monitoring, and citizen science initiatives. Bird sound recognition transcends the boundaries of traditional ornithology, inviting us to explore the hidden dimensions of our surroundings. Through the lens of audio signal processing and machine learning, we endeavor to decode the diverse lexicon of bird calls that echo through canopies and across landscapes. Each chirp, whistle, and trill represents a unique expression, a coded message in the intricate communication network of the avian realm. Bird sound recognition serves as a critical tool for researchers, ornithologists, and conservationists, enabling them to identify and monitor avian species based on their distinct vocalizations. The inherent complexity of bird calls, influenced by species-specific nuances and environmental variables, poses challenges that require sophisticated computational approaches. Delving deeper, we navigate through the algorithmic landscapes where classic signal processing techniques harmonize with the neural intricacies of deep learning models. The paper sheds light on the challenges posed by the diversity of bird calls – the regional dialects, the improvisational solos, and the subtle variations that distinguish one species from another. We explore how these challenges fuel innovation in recognition technologies, pushing the boundaries of what is possible in decoding the language of our avian counterparts. This paper aims to elucidate the key methodologies employed in bird sound recognition, ranging from traditional signal processing techniques to state-of-the-art deep learning models. The traditional methods of manual bird identification, often reliant on the expertise of seasoned ornithologists, are being augmented and, in some cases, replaced by automated recognition systems. This transition has been catalyzed by the surge of inaccessible audio recording devices, coupled with the exponential growth of audio datasets, enabling the training of machine learning models. However, the avian tapestry is not without challenges. We pause to explore the regional dialects, individual variations, and the elusive improvisations that defy simplistic categorization. In doing so, we uncover how these challenges fuel innovation, driving the development of recognition technologies that adapt and evolve, echoing the dynamic nature of avian communication. This paper aims to provide an overview of the current landscape of bird sound recognition and broaden the views on cutting-edge technologies such as neural networks, predictive analysis, and other AI technologies. Through this exploration, we hope to inspire further research and innovation in the dynamic field of technology and ornithology.

The evolution of bird sound recognition has been propelled by advancements in digital signal processing and neural network methodologies. These cutting-edge technologies have revolutionized the way we perceive and analyze avian vocalizations, offering unprecedented insights into the complex language of birds. By leveraging computational algorithms and machine learning models, researchers and conservationists can now discern patterns, identify species, and monitor populations with unprecedented accuracy and efficiency.

One of the most compelling aspects of bird sound recognition is its interdisciplinary nature, bridging the gap between ornithology, computer science, and ecology. It is a testament to the power of collaboration and innovation, as experts from diverse fields come together to unravel the mysteries of avian communication. By combining expertise in signal processing, machine learning, and field biology, researchers are pushing the boundaries of what is possible in bird sound recognition, opening new avenues for exploration and discovery.

At the heart of bird sound recognition lies the challenge of capturing the diversity and complexity of avian vocalizations. From the melodic trills of songbirds to the haunting calls of raptors, each species possesses a unique vocal repertoire shaped by evolutionary, ecological, and behavioral factors. This intricate tapestry of sounds presents a formidable challenge for recognition systems, requiring them to discriminate between subtle variations and nuances that distinguish one species from another.

To meet these challenges, researchers have developed a plethora of computational techniques and methodologies, ranging from traditional signal processing algorithms to state-of-the-art deep learning models. These approaches draw inspiration from both biological principles and technological innovations, mimicking the auditory processing mechanisms of the human brain while harnessing the computational power of modern computing systems.

In recent years, deep learning models, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have emerged as powerful tools for bird sound recognition. These models excel at capturing complex patterns and temporal dependencies in audio data, enabling them to achieve remarkable accuracy in species identification and classification tasks. By training these models on large-scale audio datasets, researchers can leverage the collective wisdom of millions of bird calls to enhance the performance and robustness of recognition systems.

However, the journey into the world of bird sound recognition is not without its challenges and limitations. Despite the tremendous progress made in recent years, recognition systems still struggle to cope with certain aspects of avian vocalizations, such as background noise, overlapping calls, and species with similar vocalizations. Moreover, the availability of high-quality labeled datasets remains a bottleneck for training accurate and reliable recognition models.

Despite these challenges, the potential of bird sound recognition as a tool for scientific inquiry and conservation efforts is immense. By harnessing the power of technology, we can gain unprecedented insights into avian behavior, ecology, and evolution, paving the way for more effective conservation strategies and sustainable management practices. As we continue our exploration of the dynamic intersection between nature and technology, the possibilities for innovation and discovery are limitless.

RELATED WORK

"Bird Sound Classification based on Machine Learning Algorithms"[1]. This paper presents a structured methodology comprising four stages: data collection, preprocessing, feature extraction, and classification. Notably, the study explores the intriguing aspect of predicting emotions in bird vocalizations by comparing them with the female voice. Through the implementation of three distinct algorithms—Naïve Bayes, J4.8, and Multilayer Perceptron—the paper evaluates their efficacy in classifying bird sounds. Remarkably, the J4.8 algorithm demonstrates the highest accuracy rate of 78.40%. Additionally, the paper references another literature source where an MC-LS-VM Classifier achieved an accuracy rate of 85.43%, indicating the variability in performance across different classification techniques.

"Classification of Bird Sounds as an Early Warning Method of Forest Fires using Convolutional Neural Network (CNN) Algorithm"[2]. The paper examines the notable approach in leveraging bird vocalizations for forest fire detection. The key takeaway from the study is the exploration of two distinct conditions: the normal state scheme and the threatened state scheme, the latter aimed at detecting forest fires. The methodology involves the conversion of bird sound data into spectrogram images or digital signals utilizing the MEL-scale log-amplitude spectrogram technique. As the CNN method inherently processes data in image form, two architectures—ResNet and the Inception model—are employed. Notably, the paper highlights that the combination of these approaches, coupled with deep learning, yields the most promising results. This underscores the effectiveness of merging advanced deep learning techniques with traditional spectrogram analysis for early detection of forest fires using bird vocalizations, providing valuable insights for forest monitoring and conservation efforts.
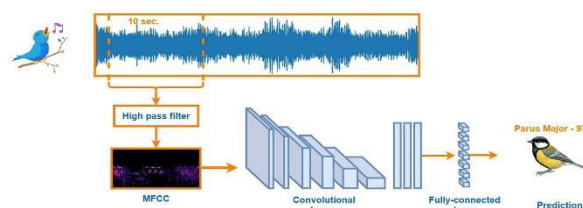

Figure 1: Classification of Spectrograms

"AI-based Bird Sound Classification Using Convolutional Neural Networks"[3]. The paper underscores the utilization of deep learning algorithms, specifically convolutional neural networks (CNNs), for both model training and classification tasks. Notably, the study emphasizes the implementation of a sound recognition system employing CNNs. The methodology outlined involves the conversion of bird sounds into their respective spectrograms, facilitated by the Librosa inbuilt Python package for audio and music processing. The classification process employs various techniques, including the Random Forest classifier with an ECC classifier (Ensemble of Probabilistic Classifier Chains), Theano and Lasagne deep learning models, the LIBSVM classifier utilizing the one-against-one decomposition technique, and the Support Vector Machine. Additionally, the Hidden Markov model is utilized in the classification pipeline. Through the integration of these methodologies, the paper aims to demonstrate the effectiveness of CNNs in accurately classifying bird sounds, contributing to advancements in automated bird sound recognition systems and fostering applications in biodiversity monitoring and conservation efforts.

"A Comparison Study to Identify Bird Species based on Bird Song Signals"[4]. The key focus lies on utilizing various features, including Mel-Frequency Cepstral Coefficients (MFCCs) and geo-related meta-features, for bird species identification. Technical aspects such as Linear Discriminant Analysis, Support Vector Machines, K-nearest neighbor, and Ensemble Learning are explored to assess their effectiveness in classifying bird species based on their song signals. The methodology involves establishing the mapping between physical frequency scale and perceptual frequency, with mathematical equations provided for this mapping process. Notably, the XENO-CANTO dataset serves as the primary

source of bird song signals for the study, facilitating comprehensive analysis and comparison of different classification techniques. Through this comparative analysis, the paper aims to provide insights into the most effective methods for identifying bird species based on their unique song signals, contributing to advancements in avian bioacoustics research and conservation efforts.

"Wavelets Recognition of Bird Sounds"[5]. The paper introduces a novel approach that treats bird sounds as wavelets, employing artificial neural networks (ANN) to recognize distinctive patterns within these wavelet representations. Notably, the study utilizes two algorithms: the unsupervised Self-Organizing Map (SOM) and the supervised Multilayer Perceptron (MLP). The methodology involves categorizing bird sounds into harmonic and inharmonic classes, with a particular emphasis on the effectiveness of inharmonic sounds in the classification process. The species identification is accomplished by analyzing the arrangement of nodes within the SOM, which acts as a wavelet function. Four key features—maximum energy, position, speed, and width—are utilized to characterize and differentiate bird sounds, facilitating their recognition and classification. Through the integration of wavelet analysis and ANN-based pattern recognition techniques, the paper aims to provide a robust framework for the automated identification of bird species based on their unique vocalizations, contributing to advancements in avian bioacoustics research and conservation efforts.

PROPOSED WORK

The proposed system aims to develop an advanced framework for bird sound classification using convolutional neural networks (CNNs). This system is designed to address key challenges in avian bioacoustics, including the segmentation of bird syllables, accurate species identification, and the interpretation of communication patterns. Leveraging recent advancements in deep learning and signal processing techniques, the proposed system offers a comprehensive approach to analyzing and classifying bird vocalizations.

At the core of the proposed system is a CNN architecture tailored specifically for bird sound classification. The architecture is designed to effectively capture both temporal and spectral features of bird vocalizations, thereby enabling accurate classification across various species and environmental conditions. Convolutional layers are strategically stacked to extract hierarchical features from Mel-spectrogram representations of audio segments. Pooling layers are incorporated to reduce spatial dimensions and computational complexity, facilitating efficient feature extraction. Dropout layers are utilized to prevent overfitting and enhance the generalization capability of the model.

To train the CNN model, a diverse dataset of bird sounds is collected from different habitats and environmental contexts. These recordings undergo preprocessing, including segmentation into short segments and conversion into Mel-spectrogram representations using Mel Frequency Cepstral Coefficients (MFCCs). Each segmented audio clip is labeled with its corresponding bird species, forming the basis for supervised learning. The training process involves optimizing the CNN model parameters using suitable optimization algorithms, such as stochastic gradient descent (SGD) or Adam.

Throughout the training phase, the performance of the CNN model is continuously monitored and evaluated using metrics such as accuracy, precision, recall, and F1-score. Confusion matrix analysis is conducted to assess the model's ability to distinguish between different bird species and identify any potential misclassifications. Fine-tuning and hyperparameter optimization techniques are employed to enhance the model's performance and robustness.

One of the key strengths of the proposed system lies in its ability to handle challenging scenarios commonly encountered in bird sound analysis, such as background noise, simultaneous vocalizations, and variations in recording conditions. Advanced signal processing techniques, including noise reduction algorithms and data augmentation methods, are integrated into the system to improve the robustness and generalization capability of the CNN model.

The implementation of the CNN architecture tailored specifically for bird sound classification requires careful consideration of various factors, including model architecture, hyperparameters, and training methodologies. Researchers and practitioners must select appropriate CNN architectures based on their specific requirements and computational resources. Techniques such as transfer learning, where pre-trained CNN models are fine-tuned on bird sound datasets, can also be leveraged to expedite the training process and improve model performance.

In addition to CNN architecture, the preprocessing steps play a crucial role in enhancing the accuracy and robustness of the classification system. The segmentation of bird vocalizations into short segments and the conversion into Mel-spectrogram representations using Mel Frequency Cepstral Coefficients (MFCCs) are essential preprocessing steps that facilitate feature extraction and model training. Furthermore, data augmentation techniques, such as time stretching, pitch shifting, and adding background noise, can be employed to augment the training dataset and improve the generalization capability of the CNN model.

The training process involves optimizing the CNN model parameters using suitable optimization algorithms, such as stochastic gradient descent (SGD) or Adam. Hyperparameter tuning techniques, such as grid search or random search, can be utilized to fine-tune the model hyperparameters and optimize performance metrics such as accuracy, precision, recall, and F1-score. Continuous monitoring and evaluation of the CNN model performance throughout the training phase are essential to identify potential issues and refine the model accordingly.

Confusion matrix analysis is a valuable tool for assessing the CNN model's ability to distinguish between different bird species and identify any potential misclassifications. By analyzing the confusion matrix, researchers can gain insights into the model's strengths and weaknesses across various classes and identify areas for improvement. Fine-tuning and hyperparameter optimization techniques can be employed based on the results of confusion matrix analysis to enhance the model's performance and robustness.

The proposed system's ability to handle challenging scenarios commonly encountered in bird sound analysis, such as background noise, simultaneous vocalizations, and variations in recording conditions, is a key strength that sets it apart. Advanced signal processing techniques, including noise reduction algorithms and data augmentation methods, are

integrated into the system to improve the robustness and generalization capability of the CNN model. By effectively mitigating these challenges, the proposed system can achieve more accurate and reliable bird sound classification results.

Moreover, the scalability and adaptability of the proposed system make it suitable for deployment in various real-world applications, including wildlife sanctuaries, ecological research projects, and citizen science initiatives. The trained CNN model can be seamlessly integrated with existing bird monitoring and conservation initiatives, providing automated and accurate classification of bird vocalizations. This enables researchers and conservationists to gain valuable insights into avian behavior, population dynamics, and ecosystem health, ultimately contributing to biodiversity conservation efforts.

METHODOLOGY

The methodology employed in this study encompasses a meticulous process for bird sound classification using convolutional neural networks (CNNs). Initially, a comprehensive dataset of bird sounds is gathered from diverse habitats and environmental conditions. These recordings undergo preprocessing, wherein they are segmented into short segments, typically around 5 seconds in duration. This segmentation aims to enhance the model's accuracy by focusing on smaller units of audio data.

Following segmentation, the audio segments are transformed into Mel-spectrograms. This transformation involves the application of Mel Frequency Cepstral Coefficients (MFCCs), a widely used technique for capturing the spectral characteristics of audio signals. The MFCC formula serves as a pivotal step in converting the raw audio data into a format suitable for CNN-based classification. The formula calculates the MFCC coefficients for each segment, capturing important features of the audio signal:

$$MFCC(i) = \frac{1}{t}\left(\sum_{f=1}^{N} |X(t,f)|^2 . H(f) . \cos\left[i.\frac{\Pi}{N}.(f - 0.5)\right]\right)$$

Here, MFCC(i) represents the ith MFCC coefficient, T is the total number of time frames, X(t, f) denotes the Short-Time Fourier Transform (STFT) magnitude spectrum of the audio signal at time t and frequency f, H(f) represents a Mel filter bank, and N is the number of frequency bins.

Each segmented audio clip is then labeled with its corresponding bird species, forming the basis for supervised learning in training the CNN model. The architecture of the CNN is carefully designed to consider both temporal and spectral aspects of bird sounds. Increasing the number of convolutional layers allows the model to capture intricate patterns present in the spectrograms. Additionally, pooling layers are integrated to reduce spatial dimensions and computational load, aiding in efficient feature extraction. Dropout layers are incorporated to prevent overfitting during the training process.

The model is trained using a suitable algorithm, with continuous monitoring of its performance throughout the training iterations. Techniques such as confusion matrix analysis are employed to evaluate the model's effectiveness in classifying different bird species. This analysis provides insights into the model's strengths and weaknesses across various classes, guiding further refinement.

Continuous improvement of the model is ensured by regularly updating it with new datasets. This iterative process enhances the model's performance over time, enabling it to adapt to a wider range of bird sounds and environmental conditions. Moreover, monitoring the model's performance in real-world applications allows for adjustments and fine-tuning as needed.

Ultimately, the trained CNN serves as a powerful tool for classifying bird sounds, enabling predictive analysis of anticipated vocalizations. By leveraging the distinctive features of bird vocalizations, the model contributes to the understanding of avian communication patterns, migration behaviors, and mating periods. Implementation of the model in sanctuaries, zoos, and wildlife areas facilitates biodiversity preservation efforts by providing valuable insights into bird populations and potential threats to their habitats.

CONCLUSION

In conclusion, the development of a convolutional neural network (CNN)-based system for bird sound classification represents a significant advancement in the field of avian bioacoustics. Through the systematic integration of advanced signal processing techniques and deep learning methodologies, the proposed system offers a comprehensive approach to analyzing and interpreting bird vocalizations. The methodology outlined in this study provides a structured framework for collecting, preprocessing, and classifying bird sounds, with a primary focus on leveraging the capabilities of CNNs to accurately identify bird species based on their vocalizations. By incorporating techniques such as Mel Frequency Cepstral Coefficients (MFCCs) and convolutional layers, the system effectively captures both temporal and spectral features of bird vocalizations, enabling robust and accurate classification across diverse environmental conditions. The proposed system addresses several key challenges in avian bioacoustics, including background noise, simultaneous vocalizations, and variations in recording conditions. Advanced signal processing algorithms and data augmentation techniques are integrated into the system to enhance its robustness and generalization capability, ensuring reliable performance even in complex acoustic environments. Furthermore, the versatility and scalability of the proposed system make it well-suited for deployment in various real-world applications, including wildlife monitoring, ecological research, and conservation initiatives. By providing automated and accurate classification of bird vocalizations, the system facilitates the collection of valuable data on avian behavior, population dynamics, and ecosystem health. The successful implementation of the proposed system has the potential to revolutionize the way researchers and conservationists study and monitor bird populations. By enabling rapid and cost-effective analysis of large-scale audio datasets, the system accelerates the pace of scientific discovery and supports evidence-based conservation strategies. Looking ahead, further research and development efforts are warranted to continue improving the performance and scalability of the proposed system. This includes exploring novel techniques for feature extraction, model optimization, and data augmentation, as well as enhancing the system's ability to handle diverse and challenging acoustic environments.

REFERENCES

[1] Aska Ezadeen Mehyadin, Dathar Abas Hasan, Adnan Mohsin Abdulazeez, Jean Najeeb, " Bird Sound Classification based on Machine Learning Algorithms" on Asian Journal of Research in Computer Science [June 2021]

[2] Silvester Dian Handy Permana , Gusti Saputra, Budi Arifitama, Yaddarabullah, Wahyu Caesarendra, Robbi Rahim, "Classification of bird sounds as an early warning method of forest fires using Convolutional Neural Network (CNN) algorithm" on Journal of King Saud University – Computer and Information Sciences [5 May 2021]

[3] Anand R1 Shanthi T1 Dinesh C 1 Karthikeyan S 1 Gowtham M1 Veni S2, "AI based Birds Sound Classification Using Convolutional Neural Networks" on IOP Conference Series: Earth and Environmental Science [2021]

[4] Xin Guo, Qing Zhing Liu, "A Comparison study identify bird species based on bird song signals" on ITM Web of Conferences [2017]

[5] Arja Selin, Jari Turunen, Juha T. Tanttu, "Wavelets recognition of Bird Sounds" on   EURASIP Journal on Advances in Signal Processing [2006]

[6] Gnostothea-Veroniki (Veronica) Morf, "Automatic detection and classification of bird sounds in low-resource wildlife audio datasets" on Queen Mary University of London School of Electronic Engineering and Computer Science [2019]

[7] Shelby Lawson, Mark E. Hauber, "Special Collection on Bird Song" on The Auk: Ornithological Advances and The Condor: Ornithological Applications [2016]

[8] Jiang Liu, Yan Zhang , Danjv Lv, Jing Lu, Shanshan Xie, Jiali Zi, Yue Yin & Haifeng X, "Birdsong classification based on ensemble multi-scale convolutional neural network" [23 May 2022]

[9] Fan Yang, Ying Jiang, And Yue Xu, "Design of Bird Sound Recognition Model Based on Lightweight" on IEEE Access [11 August 2022]

[10] Sue Anne Zollinger, Tobias Riede, Roderick A. Suthers, "Two-voice complexity from a single side of the syrinx in northern mocking bird" on Mimus polyglottos vocalizations [15 June 2008]