Detection of Classifications Of Living Organisms Using Voice Recognition

***Abstract*—The diversity of living organisms within the Animal Kingdom presents challenges for researchers and biologists in terms of species identification and classification.Due to limits in human classification and recognition techniques, many undiscovered species continue to exist. This research paper solves this issue by suggesting a voice recognition-based system to describe and categorize distinct animal species.The primary goal is to develop a tool that helps zoologists and biologists recognise unidentified species and comprehend their behavioral patterns.The system uses speech recognition techniques to distinguish between the vocalizations of various species, providing easy access to data about particular creatures like birds, frogs, and more.This approach improves categorization accuracy and frequency by extracting special sound cues and using deep learning models. The potential impact of this research lies in providing a user-friendly solution for detecting and studying different species within the Animal Kingdom.**

***Index Terms*—species identification, voice recognition, categorization, behavioral patterns**

I. Introduction

The Animal Kingdom is home to a wide variety of species and each with its own unique traits and behaviors and many of which have yet to be fully understood. The task of discovering, identifying and categorizing unknown species is of utmost importance in the field of biology and in efforts to preserve our planet's biodiversity. Conventional methods of species identification often prove insufficient in accurately classifying different organisms solely based on their physical characteristics or appearances. To overcome these limitations this study endeavors to harness the power of voice recognition technology in order to create a robust system capable of detecting and categorizing living creatures based on their distinct vocalizations. In light of the pressing need for a fresh perspective the research aims to devise a system capable of precisely capturing and examining the unique vocal patterns exhibited by various animal species. By harnessing the immense potential of profound learning models and data-centric methodologies, this revolutionary system endeavors to bridge the divide between uncharted species and the comprehension researchers possess regarding their conduct. Not only does this groundbreaking system facilitate the identification of species but it also provides valuable insights into the behavioral characteristics exhibited by these magnificent creatures. Previous successful studies that employed speech recognition to distinguish between species, including birds, frogs, and other animals, were the inspiration for our research. The proposed system builds on the knowledge gained from these studies incorporating the latest methodologies to improve accuracy and usefulness. By developing a system that can effectively extract and analyze vocal features researchers and biologists will have a valuable tool for classifying species, studying behavior and conducting ecological research. This paper presents a unique approach to address the difficulties of identifying species which ultimately advancing our understanding of the diverse Animal Kingdom and enhancing biological research efforts.

II. Problem Detection

When it comes to identifying and classifying the diverse species that make up the vast Animal Kingdom researchers and biologists face a significant challenge. Despite their best efforts innumerable species remain unknown because of the shortcomings of traditional techniques that rely on human vision. This academic study proposes a novel voice recognition-based method to address the urgent problem of species identification. This ground-breaking technique intends to address the critical issue of effectively and precisely recognizing and classifying undiscovered species based on their distinctive vocalizations especially those that have thus far evaded detection. It does this by accurately defining and categorizing diverse animal species, especially those that have escaped detection.

The classification of various species using traditional methods, which typically rely on visual or physical qualities, can be challenging, especially when several species have similar properties. Due to this restriction there are unnamed species which hinder scientific research, ecological understanding, and conservation efforts. The main problem is the lack of a reliable and automated system that can recognize different species by listening to their vocalizations, overcoming the challenges posed by genetic or physical differences.

This issue is caused by our limited ability to recognize complex human vocalizations which is made worse by the need for more comprehensive databases on animal vocalizations. In order to solve this, the researchers developed a voice recognition system that makes use of cutting-edge deep learning models and data-driven approaches in order to precisely record and analyze the specific speech patterns displayed by various animal species.

The researchers also acknowledge that while there have been advancements in the field of voice recognition for specific animals such as avian and amphibian species, there is still room for improvement in terms of accuracy, ability to adapt to new situations and applicability to a variety of habitats. As a result, the need for a novel approach that combines profound learning, characteristic extraction, noise mitigation and taxonomy categorization emerges.

This study paper essentially explores the fundamental problem of accurately identifying and classifying novel species within the enormous Animal Kingdom through the use of speech recognition technology. The goal is to develop a robust framework that can act as a link between as-yet-unidentified species and our understanding of their behavior, helping researchers, zoologists and biologists in their efforts to categorize and advance ecological research and conservation initiatives.

III. Research Objective

Animal Kingdom is the vast place of existence of enormous unknown species of different types of living organisms. Most of the people are unaware of that species and their characteristics.So this creates a problem for humans to detect any unknown species for any research or experimental purposes.As a result sometimes they could not adapt that animal for not knowing their behavioral patterns.In addition biologist or zoologist sometimes fail to classify the unknown species of animals despite of having research. So this research aims to build such a system which would show the characteristics of any unknown species of any specific animal with the power of voice recognition. This system will solve the issue of the human and gives an output of the characteristics and names of that specific species.This will create an easy access for zoologist or biologist to work freely with classification of animals.

- Differentiating different voices according to various species of animal kingdom
- Making easy access for people to collect information about a specific animal for different purposes such as birds,frogs etc
- Can be used easily for extraction of characteristics of any animal of the kingdom.
- Enhancing the performance of the models with more accuracy and frequency
- Extracting the unique sound of different species easily perfectly to detect any unknown species of animal kingdom

IV. Literature Review

Classification of living organisms using voice recognition can help the zoologists or biologists to do any type of experimental work using animals. So many researchers have already worked out on the voice recognition of different animals to analyze their different species including amphibians, reptiles, birds and many more.So some of the glimpse of previous works performed to analyze the different voices of animals to detect their behavioral attitudes on our atmosphere.

One of the research papers discusses a study conducted in Puerto Rico which used deep learning, labeling techniques, and a customized loss function to identify bird and amphibian sounds. The authors improved model accuracy by using both labeled and unlabeled data. Their approach can be applied to other sound classification problems and emphasizes transfer learning and pseudo-labeling for precise classifications. The paper presents a framework for bioacoustics problems and emphasizes the importance of transfer learning and pseudo-labeling in scarce labeled data [1].

Moreover, another article presents a computerized system that can identify different frog species based on their sound. The system analyzes sound samples by breaking them down into smaller parts and extracting three specific features. The purpose of this system is to make it easier for people to access and find information about frogs online. Experimental results show that the system is effective in recognizing frog species with both the kNN and SVM classifiers achieving high accuracy rates. However, there is still room for improvement in accurately classifying frog species with unique sounds, and the researchers plan to develop new methods to address this challenge. Future work: Thus this paper intends to develop new techniques in the feature to extract the features of the sounds for some specific frog species, whose classification accuracy rate falls below average such as Microhyla butleri and Microhyla ornate [2].

A study analyzed the ability of individuals with and without dyslexia to recognize voices in different languages. Participants were matched based on factors and trained to identify speakers' voices. The experiment was conducted in a quiet room with participants completing the test at their own pace. The researchers focused on native English speakers with and without dyslexia using recordings of sentences spoken by both languages. [3].

The article "Bird Sound Recognition Using CNN" investigates the classification of bird sounds using a Convolutional Neural Network (CNN). It tests various configurations and hyperparameters using a dataset of Xeno canto bird songs, providing insights into bird species, characteristics, and color schemes. Transfer learning is used to adapt pre-trained neural networks, and spectrograms are used for compatibility. Alternative CNN methodologies include unsupervised learning, decision tree-based feature selection, recurrent CNNs and Hidden Markov models. The process involves corpus creation, automated data collection, sound preprocessing, pre-trained network utilization, and network training. The trained model categorizes user-recorded bird sounds and returns results after averaging. Future work aims to reduce noise in preprocessing by filtering extreme frequencies and experimenting with different gamma values [4] Another experiment was mainly done on detection of different species of birds by classifying the sounds of birds with high accuracy using the deep learning models. This study presents a novel machine learning model for accurately detecting bird noises using deep learning models. The model uses a multileveled ternary pattern (TP) feature generator, combining high-level, low-level, and mid-level characteristics. The discrete wavelet transform (DWT) is used to create levels. An iterative variant of the ReliefF feature selector (IRF) is used to improve performance.The model's outputs are influenced by classifiers like SVM, LD, and k-nearest neighbors.The multilayer TP and IRF-based bird sound classification approach achieved an impressive 96.67% accuracy on 18 distinct bird sound classes. This method offers a

simple classification model while accurately categorizing bird noises.This research study aims to create a mobile application for cloud-based bird species classification and detection using a multilevel TP and IRF method based on a deep learning model. The method offers advanced capabilities for accurately identifying bird species. The experimental setup used gathered bird sounds as input for the proposed method. The method was implemented using MATLAB 2018a on a personal computer with 32 GB of memory, an i7 7th generation microprocessor, and a 1 TB HDD. The research focuses on generating low, medium, and high-level features and using an effective feature selector for highly accurate bird sound classification. Future works may include presenting a new TP-based deep model for sound classification, addressing parameter setting in recurrent neural networks, and classifying other animals' sounds [5].

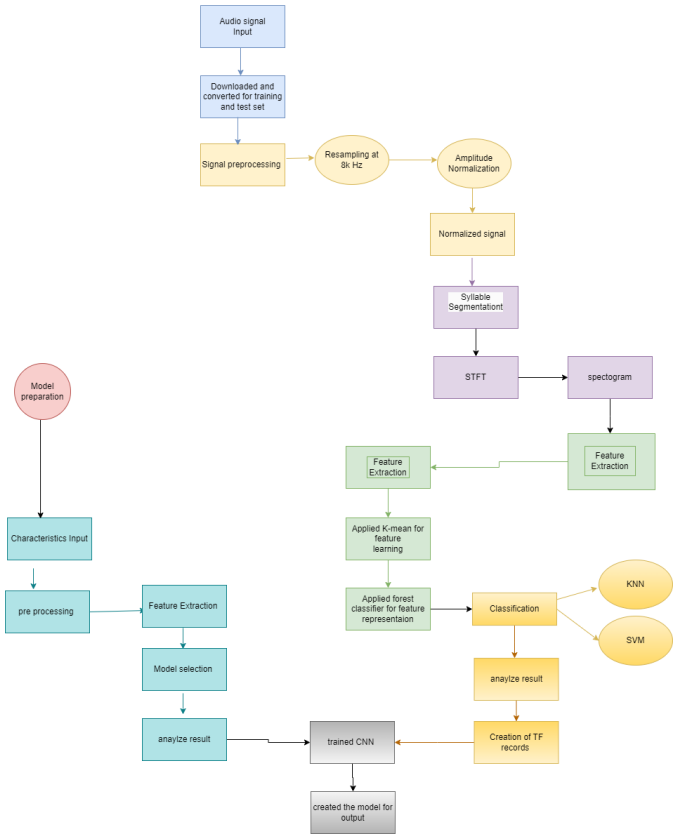
Furthermore, another research study mainly focuses on finding the solution of issues faced by the non-human animals in voice perception to normalize the similar voice differences. This research addresses how non-human animals perceive and recognize human voices. It explores the idea of "speaker normalization," where animals can identify voices independently of who is speaking. The study shows that animals can recognize individuals based on vocal traits like pitch and formants, aiding social interactions. It concludes that animals have perceptual skills for vowel discrimination, speaker voice recognition, and speaker normalization shedding light on how different species communicate and interact [6].

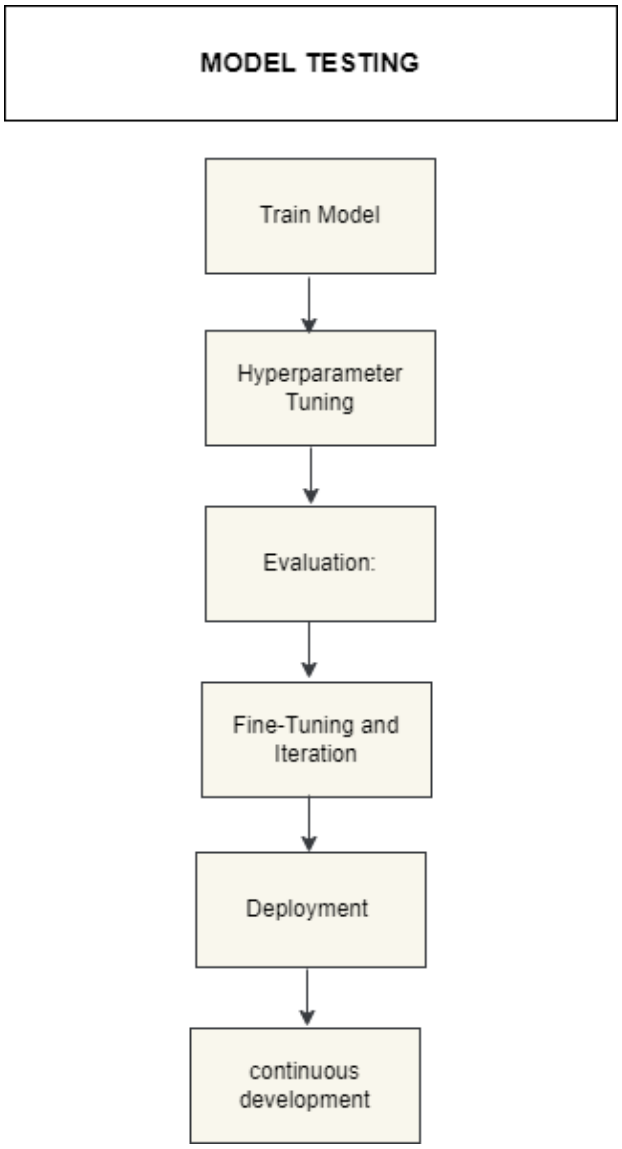
Another wonderful article discusses the use of bioacoustic sensors also known as autonomous recording units (ARUs) for recording wildlife sounds in a scalable and non-invasive manner. The article proposes several machine listening techniques to improve the sound event detection (SED) models across different acoustic environments. To enhance the SED models two noise adaptation techniques are introduced. The article highlights the importance of automated solutions for avian flight call detection in large-scale bioacoustic monitoring. By utilizing machine listening techniques and bioacoustic sensor networks it becomes possible to track bird migration patterns, which can have implications for wildlife conservation, urban planning, and risk assessment.The article discusses enhancing sound event detection (SED) models for avian flight call detection in varied acoustic environments. Twelve models are examined, highlighting the effectiveness of combining PCEN, GDA, and context adaptation for improved SED in bioacoustic sensor networks. These advances have implications for monitoring migratory bird populations on a large scale aiding in their conservation [7]

The researcher conducted another experiment to evaluate machine learning methods for detecting bird sounds in acoustic recordings. The results showed that deep learning methods can achieve high recognition rates, even with mismatched training data. However, there is still room for improvement in generalizing to new conditions. The paper emphasizes the importance of monitoring bird populations for conservation, research, and ecosystem management. The challenge results show that machine learning methods, particularly deep learning, can achieve high recognition rates on remote monitoring acoustic data, even in weather noise, low signal-to-noise ratios and various bird call types. This advancement in machine learning's practical utility in remote monitoring projects is significant, but the study also highlights the gap between performance in matched and mismatched conditions. Overall, the research paper demonstrates the potential of machine learning in improving bird detection and monitoring through acoustic data, highlighting progress in reducing error rates and the need for generalization challenges in real-world deployment scenarios [8].

Another research paper introduces the Automated Remote Biodiversity Monitoring Network (ARBIMON), a system designed to automate data acquisition, management, and species identification based on audio recordings.The paper showcases the ARBIMON system, which collects detailed and long-term acoustical data on vocal activity patterns of birds, frogs, insects, and mammals from Puerto Rico and Costa Rica. The system offers fine temporal resolution and coverage, and automates species identification using machine learning algorithms. It reduces observer bias, provides permanent records, real-time monitoring, and responds quickly to significant events. ARBIMON is a valuable tool for improving biodiversity monitoring, contributing to understanding large-scale ecological processes, and aiding conservation and management effor [9].

Final model workplan





V. Methodology

The study of life's diversity has always been central to biology. Traditional taxonomy used observations, anatomy, and genetics. Now, AI and biology are merging, using voice recognition to identify species by their unique sounds. This breakthrough decodes animal vocalizations, aiding new research in categorization. Different techniques are explored to identify known or unknown animals through their voices.

A. Data Collection

B. Model Training

Firstly, in the deep learning field transfer learning and pseudo learning has been implemented to collect voices of animals for classification of specific species of animals precisely.r learning is a machine learning technique where knowledge gained from training on one task is used to improve the performance on a different but related task. It involves adapting a pre-trained model's knowledge to new similar tasks often leading to faster and more accurate learning. When a machine learning model looks to be learning and improving, it may really be learning from noise or unimportant patterns in the data. This is known as pseudo learning. As a result, the model may overfit and have difficulty generalizing to new data since it is not actually learning anything useful. For classifying the Frog species effectively KNN and SVM classifiers have been used.The sounds of birds has been detected using CNN system with the involvement of different configuration and hyperparameters. Convolutional Neural Networks often known as CNNs, are a particular kind of artificial neural networks created specifically for processing and evaluating visual input, such as pictures and movies. For tasks like image identification, object detection, and picture classification it uses convolutional layers to automatically learn and extract pertinent characteristics from the input data. Spectrograms has been applied to obtain the compatibility for a representational transformation of images of two dimensions and audio of one dimension. The spectrogram interprets the STFT output as a picture using a color map which may be input into an image-based pre-trained network. The multileveled ternary pattern (TP) feature generator, a multileveled and handmade feature is used in the model. The performance of the model is further enhanced by an iterative variation of the ReliefF feature selection (IRF). IRF uses several classifiers such as support vector machine (SVM), linear discriminant (LD) and k-nearest neighbor (kNN), to automatically select the most advantageous characteristics. The outputs of these classifiers have an impact on the results of the final classification. For noise adaptation in SED models two types of techniques have been developed mentioning bioacoustic sensors networks to track the migration of bird species. This technique will help in wildlife conservation urban planning, and risk assessment. Automated Remote Biodiversity Monitoring Network (ARBIMON) helps to automate data acquisition, management, and species identification based on audio recordings. It plays a great role in conservation of biodiversity.

VI. Experimental Analysis

This model actually shows us some experimental analysis depending on running under some datasets and training on classifiers.Two classifiers, kNN and SVM, were compared in the experiments.The overall average recognition rates for both classifiers are similar. They both achieved a high accuracy rate of recognition.Among two datasets birdbox one of them. The BirdVox-full-night collection includes recordings from a single ten-hour autumn migration night as captured by six separate sensors. These sensors each have one reasonably priced omnidirectional microphone and are situated in rural locations close to Ithaca, New York, in the United States. The resultant bioacoustic sensor network spans around 1000 km2 of land in total. BirdVox-full-night contains 62 hours of monaural audio data from its 6 recordings. The same

"leave-one-sensor-out" assessment approach as described in the preceding section is used to divide the data into the training set, validation set, and test set. All models are therefore evaluated on recording scenarios that are not part of the training or validation subsets.In this analysis basically we applied both pseudo learning and transfer learning to demonstrate the workflow of our system.With the help of transfer learning, deep convolutional neural network models may be trained with comparably minimal training data. This is particularly useful when it is challenging to gather a lot of labeled data. Adding or removing background noise, cropping the original mel spectrograms creating more mel spectrograms with time shift from the raw audios or more sophisticated techniques like creating synthetic data using Generative Adversarial Networks (GANs) may all be useful ways to further enhance the model performance. Additionally, in transfer learning, pre-training a model using a sound-only dataset provides an option to employ pre-trained models that are based on ImageNet.The paper demonstrates how deep CNNs can categorize sounds from various bird and amphibian species in a tropical location with over 95% sensitivity and specificity, helped by transfer learning and pseudo labeling. By combining confirmed detections from diverse detectors with template-based data, this unique technique improves bioacoustics categorization.

VII. Future Work

The main goal of this research is to improve the precision of word classification. By employing a fresh and distinctive strategy, the researchers are focused on overcoming challenges brought on by particular types of disturbance, particularly those connected to frog species. This method not only applies to frogs but also holds promise for enhancing the classification of diverse sound-related problems.

One of the primary objectives of this research is to develop innovative methods for extracting sound characteristics from particular frog species that currently exhibit lower than desired average classification accuracy rates. The aim is to enhance the accuracy of classification algorithms through the implementation of these novel techniques ultimately contributing to a greater understanding of frog behavior and communication.

Now, the limitations of the research paper is the firstly is relies on sound detection algorithms and datasets. As we know many of the unknown species might not have sound i.e the power of vocalization. This research is unable to identify those sort of species. That is of course a logical limitation of this research. Secondly, if there are more than one animal sound, the algorithm will be unable to specify or diminish between the several sounds, so we will be trying to develop this in the future work so that the algorithm can identify different sounds when there are several.

Furthermore, there are some tribes which uses animal sounds to communicate with each other such sounds can be counted as animal sounds. We will try to distinguish these sounds with actual animal sounds.

In addition to focusing on the details of classification, the primary objective of this study is to provide a deeper understanding of the way animals perceive and interact through sound. By analyzing sound patterns and classifications, the researchers aspire to gain insights into the cognitive mechanisms that drive social interaction and communication among different species.

While theoretical knowledge often forms the basis for scientific progress, this research underscores the importance of acknowledging the challenges and intricacies associated with implementing such knowledge in real-world situations. The overarching objective is not only to minimize errors in classification but also to fully grasp and tackle the obstacles that arise when endeavoring to put theoretical concepts into practice. By doing so, we aim to achieve a deeper understanding of the complexities involved and effectively address the issues faced in practical contexts.

To conclude this study is dedicated to enhancing the accuracy of word classification through the development of innovative approaches for extracting word features, reducing noise and constructing classification models. The main objective is to deepen our understanding of animal communication, cognition, and social interactions by specifically investigating frog and bird species. Nonetheless, the study recognizes the challenges involved in translating research outcomes into practical applications.

VIII. Dataset

The research paper harnesses the potential of three distinct audio datasets to propel avian bioacoustics research and the automation of bird sound analysis.

The BirdVox-DCASE-20k dataset encompasses 20,000 ten-second audio clips gathered by autonomous recording units near Ithaca, NY, during fall 2015. Primarily aimed at avian vocalization studies and bioacoustic classification model development the dataset showcases an approximately 50 percent prevalence of bird sounds. This dataset is a derivative of BirdVox-full-night which caters to the "Bird Audio Detection" challenge facilitated by DCASE. Equipped with binary "has-bird" labels in its metadata it encourages proper attribution to the original publication, adhering to the CC BY 4.0 license. The creation of this dataset was made feasible by grants, a Google Faculty Award, and contributions from anonymous donors [10].

The NIPS4Bplus dataset introduces a groundbreaking avenue for automating bird species detection and classification using audio recordings. Characterized by recordings spanning regions in France and Spain, it integrates species-indicative tags. The dataset uniquely incorporates manually appended temporal annotations, offering an unprecedented understanding of species occurrence patterns. This augmentation extends the utility of the existing NIPS4B dataset for enhanced avian bioacoustics research [11].

The BirdVox-full-night dataset comprises six extensive ten-hour audio recordings captured in the vicinity of Ithaca, NY, during fall 2015. This dataset provides an extensive repository of avian flight calls amounting to around 35,402 instances. Although species-specific labeling is absent distinctions based

on center frequencies remain feasible. This dataset serves as an indispensable resource for the advancement of bioacoustic classification models. With FLAC audio files, JAMS annotation files, and GPS coordinates, it adheres to the CC BY 4.0 license while emphasizing its "as is" nature [12].

Together these datasets underscore the pivotal role of automation in avian bioacoustics research and offer substantial resources to facilitate advancements in bird sound analysis.

IX. Conclusion

The paper being presented unveils a groundbreaking and comprehensive methodology that holds immense promise for addressing a wide range of bioacoustics challenges. A noteworthy achievement of this research lies in its ability to successfully distinguish between different frog species with exceptional precision. By harnessing the power of both kNN and SVM classifiers, the study attains remarkably high levels of accuracy in species identification. This investigation tackles two crucial problems: the development of diverse feature sets and the implementation of a feature selector to construct an exceptionally reliable model for classifying bird sounds. These efforts are aimed at overcoming the inherent difficulties associated with creating and selecting relevant features in this field.

One remarkable aspect of the research paper is the captivating array of skills exhibited by animals when it comes to acoustic perception. This encompasses their remarkable capacity to differentiate between different vowel sounds, recognize different voices, and adapt to an assortment of speakers. To enhance the precision of sound event detection models in the realm of bioacoustic monitoring, the researchers propose innovative techniques that involve thorough investigation, specifically targeting the identification of avian flight calls amid the vast expanse of this field.

The study's main discoveries revolve around showcasing the transformative power of acoustic data in the realm of bird detection and monitoring, thanks to the utilization of machine learning techniques specifically deep learning. This advanced system has the capacity to greatly enhance our comprehension of large-scale ecological processes and offer vital assistance for conservation and ecological management efforts. The research paper highlights the crucial importance of state-of-the-art technology in reshaping our understanding of the environment and propelling essential conservation endeavors forward.

References

[1] M. Zhong, J. LeBien, M. Campos-Cerqueira, R. Dodhia, J. L. Ferres, J. P. Velev, and T. M. Aide, "Multispecies bioacoustic classification using transfer learning of deep convolutional neural networks with pseudo-labeling," *Applied Acoustics*, vol. 166, p. 107375, 2020.

[2] C.-J. Huang, Y.-J. Yang, D.-X. Yang, and Y.-J. Chen, "Frog classification using machine learning techniques," *Expert Systems with Applications*, vol. 36, no. 2, pp. 3737–3743, 2009.

[3] T. K. Perrachione, S. N. Del Tufo, and J. D. Gabrieli, "Human voice recognition depends on language ability," *Science*, vol. 333, no. 6042, pp. 595–595, 2011.

[4] A. Incze, H.-B. Jancsó, Z. Szilágyi, A. Farkas, and C. Sulyok, "Bird sound recognition using a convolutional neural network," in *2018 IEEE 16th international symposium on intelligent systems and informatics (SISY)*. IEEE, 2018, pp. 000 295–000 300.

[5] T. Tuncer, E. Akbal, and S. Dogan, "Multileveled ternary pattern and iterative relieff based bird sound classification," *Applied Acoustics*, vol. 176, p. 107866, 2021.

[6] B. Kriengwatana, P. Escudero, and C. Ten Cate, "Revisiting vocal perception in non-human animals: a review of vowel discrimination, speaker voice recognition, and speaker normalization," *Frontiers in Psychology*, vol. 5, p. 1543, 2015.

[7] V. Lostanlen, J. Salamon, A. Farnsworth, S. Kelling, and J. P. Bello, "Robust sound event detection in bioacoustic sensor networks," *PloS one*, vol. 14, no. 10, p. e0214168, 2019.

[8] D. Stowell, M. D. Wood, H. Pamuła, Y. Stylianou, and H. Glotin, "Automatic acoustic detection of birds through deep learning: the first bird audio detection challenge," *Methods in Ecology and Evolution*, vol. 10, no. 3, pp. 368–380, 2019.

[9] T. M. Aide, C. Corrada-Bravo, M. Campos-Cerqueira, C. Milan, G. Vega, and R. Alvarez, "Real-time bioacoustics monitoring and automated species identification," *PeerJ*, vol. 1, p. e103, 2013.

[10] V. Lostanlen, J. Salamon, A. Farnsworth, S. Kelling, and J. P. Bello, "Birdvox-dcase-20k: a dataset for bird audio detection in 10-second clips," Mar 2018. [Online]. Available: https://zenodo.org/record/1208080

[11] V. Morfi, Y. Bas, H. Pamuła, H. Glotin, and D. Stowell, "Nips4bplus: a richly annotated birdsong audio dataset," *PeerJ Computer Science*, vol. 5, p. e223, Oct 2019. [Online]. Available: https://arxiv.org/pdf/1811.02275v2.pdf

[12] V. Lostanlen, J. Salamon, A. Farnsworth, S. Kelling, and J. P. Bello, "Birdvox-full-night: a dataset for avian flight call detection in continuous recordings," *Zenodo*, Oct 2017. [Online]. Available: https://zenodo.org/record/1205569