Animal Speech Recognition using Deep Learning techniques

Aiswarya Shajil Kumar
School of Computing
Amrita Vishwa Vidyapeetham
Amritapuri,India
amenu4aie21107@am.students.amrita.edu

K Bharadwaj
School of Computing
Amrita Vishwa Vidyapeetham
Amritapuri,India
amenu4aie21135@am.students.amrita.edu

Navami S K
School of Computing
Amrita Vishwa Vidyapeetham
Amritapuri,India
amenu4aie21146@am.students.amrita.edu

Parvathy G Pillai
School of Computing
Amrita Vishwa Vidyapeetham
Amritapuri,India
amenu4aie21150@am.students.amrita.edu

Megha Mohan
School of computing
Amrita Vishwa Vidyapeetham
Amritapuri,India
amenu4aie21176@am.students.amrita.edu

***Abstract*—** Animal speech recognition is an emerging field that bridges bioacoustics and artificial intelligence, aiming to interpret and categorize animal sounds for various applications such as wildlife monitoring, conservation efforts, and understanding animal behavior. This project explores the recognition of animal sounds using a machine learning approach, specifically focusing on a neural network model. We utilized a dataset comprising recordings of different animals, including cats, dogs, ducks, horses, and others. The project involves preprocessing these audio files, extracting relevant features, and training a neural network to classify the sounds. Feature extraction was carried out using the Root Mean Square (RMS) energy from audio signals, a technique that captures the variations in the amplitude of sound waves, providing a robust representation of the audio characteristics. We implemented the feature extraction using the Librosa library and visualized the results with spectrograms to analyze the power distribution across frequencies. The extracted features were then fed into a Multi-Layer Perceptron (MLP) classifier, a type of neural network suited for this classification task. The classifier was trained on a dataset of labeled animal sounds and subsequently tested on new audio samples to evaluate its performance. The results demonstrated the model's ability to accurately differentiate between various animal sounds, showcasing the potential of machine learning in animal speech recognition. This project not only highlights the technical aspects of developing an animal sound recognition system but also underscores the broader implications for wildlife conservation and animal behavior studies, providing a foundation for future research in this interdisciplinary domain.

1.Literature Review

The development of animal voice recognition systems has seen substantial progress, with various research studies exploring different methodologies and technologies to improve accuracy and applicability. This literature review examines five significant research papers that have contributed to this field.

The first study by Che Yong Yeo, S.A.R. Al-Haddad, and Chee Kyun Ng titled "Animal Voice Recognition for Identification (ID) Detection System" presents a robust system for identifying animals based on their vocal patterns. The system combines Zero-Cross Rate (ZCR), Mel-Frequency Cepstral Coefficients (MFCC), and Dynamic Time Warping (DTW) algorithms. ZCR is used for endpoint detection by eliminating silence from the input voice, MFCC extracts compact features from the voice, and DTW classifies the voice pattern by finding the optimal path between the input and reference voices. The experimental results showcase the system's effectiveness in accurately recognizing animal voices, proving its potential for applications in security and veterinary contexts.

Another relevant study, "Sound-spectrogram based automatic bird species recognition using MLP classifier," focuses on recognizing bird species based on their sounds. This system utilizes a virtual instrument tool to acquire and preprocess sound samples, generating a feature matrix from short-term Fourier transform spectrograms. An MLP classifier is trained, tested, and optimized using a feedforward-backpropagation algorithm. The study reports a high recognition accuracy of 96.1%, indicating the system's efficiency and flexibility. The model's scalability allows for future retraining with larger datasets, enhancing its accuracy and reliability.

The research paper titled "Evaluation of MPEG-7-Based Audio Descriptors for Animal Voice Recognition over Wireless Acoustic Sensor Networks" introduces an innovative environmental monitoring system using Wireless Acoustic Sensor Networks (WASNs). This study focuses on the application of MPEG-7 standard audio descriptors for recognizing animal voices, particularly targeting anuran species in Spanish natural parks. The system architecture consists of low-power acoustic nodes and a base station, aiming to reduce data transmission and storage requirements. The real-world tests demonstrate high classification performance, validating the system's efficacy in scalable and automated wildlife monitoring.

In "Revisiting vocal perception in non-human animals," the authors review how non-human animals perceive and discriminate vowels, recognize speaker voices, and normalize speaker differences. This paper delves into whether these abilities in animals are comparable to those in humans and explores the evolutionary aspects of speech perception. The findings suggest some evidence of speaker normalization in animals, but current data are insufficient to conclude that vowel perception asymmetries and voice recognition are comparable to human abilities. The paper emphasizes the need for further research to better understand these mechanisms in both humans and animals.

Lastly, the research paper "Animal Voice Recognition for Identification (ID) Detection System" by Che Yong Yeo et al. revisits the development of an animal ID detection system using voice pattern recognition algorithms. The system employs Zero-Cross Rate (ZCR) for endpoint detection, Mel-Frequency Cepstral Coefficients (MFCC) for feature extraction, and Dynamic Time Warping (DTW) for voice pattern classification. The goal is to accurately identify animals based on their vocalizations, which is beneficial for applications in security and veterinary contexts. The experimental results demonstrate the system's effectiveness in recognizing specific animal voices from a database.

2. METHODOLOGY

The methodology of this project involves several key steps: data collection, feature extraction, model training, and evaluation. Initially, audio recordings of different animals were collected. These recordings included sounds from cats, dogs, ducks, horses, and other animals, providing a diverse dataset for training the recognition model.The first step in our methodology was to collect audio samples of various animal sounds. These sounds were sourced from publicly available datasets and curated recordings. Each audio file was trimmed to a duration of two seconds to maintain consistency and facilitate efficient processing. Preprocessing involved normalizing the audio signals and ensuring they were in a suitable format for feature extraction.Feature extraction is a critical step in transforming raw audio data into a format that can be used for machine learning. We used the Librosa library to extract the Root Mean Square (RMS) energy from the audio signals. RMS energy measures the magnitude of the audio signal, providing a representation of the sound's intensity over time. This feature was chosen for its robustness in capturing the dynamic variations in animal sounds. We also visualized the audio data using spectrograms, which represent the distribution of energy across different frequencies over time. These visualizations helped in understanding the distinct characteristics of each animal sound, providing insights into their spectral properties. The extracted features were then used to train a Multi-Layer Perceptron (MLP) classifier. The MLP classifier is a type of neural network that is well-suited for classification tasks. Our model consisted of an input layer, one or more hidden layers, and an output layer. The input layer received the RMS energy features, while the output layer provided the predicted animal class.We split our dataset into training and testing sets to evaluate the model's performance. The training set was used to train the MLP classifier, while the testing set was used to assess its accuracy. The model was trained using supervised learning, where the input features were mapped to their corresponding animal labels.The final step in our methodology was to evaluate the trained model on new audio samples. We used metrics such as accuracy, precision, recall, and F1-score to measure the model's performance. These metrics provided a comprehensive evaluation of the classifier's ability to correctly identify different animal sounds.

3. RESULT

The results of our animal speech recognition project demonstrated the effectiveness of using a machine learning approach to classify animal sounds. The trained MLP classifier achieved a high level of accuracy in distinguishing between different animal sounds, with specific metrics indicating the model's robustness and reliability.During the training phase, the MLP classifier showed rapid convergence, with the loss function decreasing significantly over successive epochs. This indicated that the model was learning effectively from the training data. The accuracy on the training set was consistently high, suggesting that the model was able to capture the distinguishing features of the different animal sounds.The performance of the classifier was evaluated on a separate testing set. The accuracy of the model on the testing set was also high, indicating that the model generalizes well to new, unseen data. The precision and recall for each animal class were analyzed to understand the model's performance in detail.For example, the classifier achieved high precision and recall for cat and dog sounds, indicating that it was able to correctly identify these sounds with few false positives or false negatives. The performance was slightly lower for sounds with less distinct features, such as those from horses and ducks, but still within acceptable ranges.The use of spectrograms provided a visual confirmation of the model's effectiveness. The log power spectrograms of the audio samples showed clear patterns that corresponded to different animal sounds. These visualizations helped in understanding the spectral properties of the sounds and the basis for the classifier's decisions.A confusion matrix was used to further analyze the model's performance. The matrix showed the number of correct and incorrect predictions for each animal class, providing insights into which sounds were most challenging for the model to classify. The majority of errors occurred between sounds with similar spectral properties, highlighting areas for potential improvement.

4. CONCLUSION

This project successfully demonstrated the application of machine learning techniques to the task of animal speech recognition. By using RMS energy features and an MLP classifier, we were able to achieve high accuracy in classifying a diverse set of animal sounds. The results underscore the potential of machine learning in bioacoustics, offering a robust method for analyzing and categorizing animal vocalizations.The implications of this work are significant for fields such as wildlife monitoring, conservation, and animal behavior research. Accurate recognition of animal sounds can aid in tracking species populations, identifying stress signals, and studying communication patterns. Furthermore, this project provides a foundation for future research in animal speech recognition, with potential improvements including the use of more advanced neural network architectures and additional audio features. In conclusion, our project has demonstrated the feasibility and effectiveness of using machine learning for classifying animal sounds. The high accuracy and robustness of the MLP classifier highlight the potential for further advancements in this field. By building on this foundation, future research can enhance our understanding of animal communication and contribute to the conservation and study of wildlife.

5. REFERENCES

[1] https://ieeexplore.ieee.org/abstract/document/5946409?casa_token=iuFiKgbx2X0AAAAA:DRxEDf-1vRY-aCZoA_lttuZuzGEqaTUw3EOFgiEazrdJLcMxznk1JBMdeaJ3CKcIPtYn_FcY4pG2Mqw

[2] https://www.sciencedirect.com/science/article/pii/S0003682X21001705

[3] https://www.mdpi.com/1424-8220/16/5/717

[4] https://www.sciencedirect.com/science/article/abs/pii/S0010482521003632

[5] https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2014.01543/full

[6] https://ieeexplore.ieee.org/abstract/document/5759872?casa_token=upqIH_Ea7PkAAAAA:Q8_ImDSZbJs_B5kRE1ERMYalHldS0fTnn2Z7GBTcZaIMtjTlkZGdg5eA4dhf51lYnFv09Qv8BM9_cv4

6. CODE

https://github.com/angeepique/Speech-Processing_Group9.git