Exploring Audio Features for Enhanced Bird Identification Using Neural Networks

[1] Ashwin Bhatt
Department of Master of Computer Application, CMR University,
Bangalore, India Ashwinbhatt11@gmail.com

[2] Sudharsan S
Assistant Professor, School of Science and Computer Studies
CMR University (Lake-side Campus)
Bangalore

Abstract

The identification of bird species using sound signals is crucial for wildlife conservation and environmental monitoring. In this study, we present an automatic bird species recognition system that leverages Mel-frequency cepstral coefficients (MFCCs) as features and employs an Artificial Neural Network (ANN) for classification. By utilizing a dataset of bird songs, the system pre-processes audio data to remove noise and extracts important acoustic features. The proposed ANN model demonstrates strong performance in species classification, handling real-world bioacoustics challenges such as noisy environments. This paper discusses the system architecture, including audio pre-processing, feature extraction, classification, and user interface implementation, along with the results achieved through our experiments. The system's ability to generalize well in practical settings suggests its utility in large-scale bioacoustics monitoring.

Keywords: *Audio Signal Processing, MFCC, Neural Networks, Bird Species Classification, Artificial Intelligence, Bioacoustics.*

I. INTRODUCTION

Bird populations face critical challenges due to environmental threats, including habitat loss, climate change, and human-induced disturbances. These challenges underscore the need for efficient monitoring methods to track bird species populations. Birds are often regarded as bioindicators, as their presence, behavior, and population trends offer valuable insights into the health of ecosystems [1]. As ecosystems undergo changes due to environmental stressors, declines in bird populations reflect broader ecological issues [2][3]. Thus, effective monitoring of bird species becomes an essential component of conservation efforts, especially for species that are endangered or face rapidly changing habitats.

Traditional methods of bird species identification rely heavily on visual or acoustic observations by trained ornithologists. While this method can yield high accuracy, it is labor-intensive, time-consuming, and dependent on human expertise [4]. Furthermore, these methods are impractical for large-scale, continuous monitoring, especially in remote or difficult-to-access areas. This limitation has driven the development of automated approaches to bird species identification using audio recordings and machine learning [5][6].

Among the machine learning techniques used for bioacoustics classification, artificial neural networks (ANNs) and convolutional neural networks (CNNs) have emerged as powerful tools for identifying bird species based on audio signals [7][8]. These methods allow the processing and classification of complex audio data, making them particularly suited for bioacoustics analysis. In this paper, we propose an automated bird species recognition system using Mel-frequency cepstral coefficients (MFCCs) as features and an ANN model for classification. The system is designed to analyze bird songs, predict species, and provide additional data about the bird, such as habitat, size, and region. By integrating an intuitive user interface (UI), the system can be easily operated by researchers and non-experts alike, making it accessible for large-scale monitoring.

II. SIGNIFICANE OF THE PROBLEM

Birds communicate a variety of ways through vocalizations. Birds use their sounds to send various warnings and threats about impending danger, to identify individual birds or insects in a group, and to mark and define a territory. The specialty of the call indicates that these are fairly immediate and more effective vocal expressions.

The difficulty of distinguishing different species of birds from their recorded songs can be defined as the problem of identifying birds using an automated system with the sound of birds. Bird songs, experts say, are more musical and better at identifying species than the bird calls used here. To achieve accurate identification, the entire audio signal must undergo pre-processing to isolate the most relevant segments and extract meaningful features. This process allows the system to focus on the distinct characteristics of each bird's song, enabling more precise species recognition. Therefore, building an automated solution that can effectively process and classify bird vocalizations presents both a technical challenge and an opportunity for advancing bioacoustics monitoring.

III. LITERATURE SURVEY

Several studies have explored the application of machine learning algorithms for bird species identification using audio signals.

Vivekanandam B. [1] investigated the use of Lightweight Convolutional Neural Networks (LW-CNN) in crowd counting. Their work highlighted the efficiency of LW-CNN models in reducing computational complexity while maintaining high accuracy in object counting tasks. This model structure could be applied to bird species classification for managing large datasets with limited computational resources.

In the study by researchers [2], a Convolutional Neural Network (CNN) with ResNet architecture was deployed to classify bird species. They separated the bird song recordings into two distinct classes—signal and noise—to ensure that the model trained only on relevant data. To improve generalization, noise segments were also utilized to boost the training set. This method helped in enhancing the model's ability to identify bird species in noisy environments.

Researchers in [3] proposed a system that applied CNNs to predict bird species based on features such as genus, species, subspecies, and other bird characteristics. The audio files were pre-processed using techniques like silence removal and data reconstruction before being converted into WAV format. Spectrograms generated from these files were used as inputs for a MobileNet CNN. The output was then compared with a pre-trained neural network, and bird species predictions were made.

Another significant contribution came from the authors of [4], where Hidden Markov Models (HMMs) were employed to identify bird species from field recordings. In their approach, the auditory scene was divided into spectro-temporal segments, each represented by a frequency track. These segments were used to train unsupervised HMMs, with each set representing a different bird species. This system enabled automatic bird species detection by comparing the probability ratios of the target and background models.

The work of [5] focused on acquiring bird sounds from various environments in India, combining data from natural sources and standard datasets like Xeno-canto. Mel-frequency cepstral coefficients (MFCCs) were extracted from the recordings and used as input features for a Support Vector Machine (SVM) classifier. The system achieved an overall accuracy of 64%, with the highest classification accuracy being 89% for specific bird species.

Authors in [6] explored the use of a GMM-UBM (Gaussian Mixture Model-Universal Background Model) paradigm for bird call recognition. MFCC features were extracted from various bird vocalizations, and note models were trained using data from different sources. The model aimed to classify bird species based on the acoustic properties of their songs, achieving notable success in species recognition.

An alternative approach was suggested by [7], where the Extreme Learning Machine (ELM) algorithm was used to overcome the computational inefficiencies of traditional Feed-Forward Neural Networks. Variations of ELM, such as Voting-based ELM and Symmetric ELM, were tested, achieving a classification accuracy of 94.10%. This method offered a faster computation time while maintaining high accuracy in identifying bird species.

The dataset used by the researchers in [8] contained 400 bird sound recordings of four species—cuckoo, sparrow, crow, and laughing dove. Each species had 100 recordings, and the audio clips were standardized to sampling frequencies of 44.1kHz or 48kHz. To prevent overfitting and increase diversity, the dataset was supplemented with human voice clips and environmental noise. This combination ensured that the model trained effectively even with background interference.

Researchers in [9] focused on the segmentation of bird vocal signals in a database containing recordings of 70-75 bird species from South America's Southern Atlantic Coast. They compared manual and automatic segmentation techniques, finding that automatic segmentation improved the classification accuracy for 43 out of 75 bird species. This approach provided a more efficient means of handling large-scale datasets.

The authors of [10] utilized a deep learning framework to simultaneously segment and classify bird syllables using a CNN-based encoder-decoder architecture. Their system was designed for real-time performance on in-situ recordings from forests, demonstrating significant improvements over previous techniques. The combination of visual segmentation and CNNs provided an innovative solution for bird species classification.

In the research by [11], five audio inputs were collected and saved as WAV files before being analyzed using MATLAB. Techniques such as MSE-based methods, correlation frequency shifting, and Wiener Filter theory were employed to analyze the frequency patterns of bird sounds. The system compared the results of different analysis methods, producing reliable results but with a significant time constraint due to the complexity of running multiple algorithms simultaneously.

In a study by [12], the authors employed parameter-free clustering techniques based on compression distances to assess bird species separability. By computing the distances between audio data groups and applying various clustering algorithms, they demonstrated that species separation could be achieved without the need for extensive data analysis beforehand. This approach provided a novel method for clustering bird species based on their vocalizations.

Research [13] employed the techniques of "Segmentation and Estimation of Frequency Tracks" and "HMM-based Modeling of Frequency Tracks" to develop a system for identifying bird species from audio recordings. They tested the system using bird sounds recorded in one, two, and three-second segments, achieving recognition accuracies of 92%, 88.8%, and 83.3%, respectively, for single species classification.

In the work of [14], an automated bird identification system was designed to track the presence of birds around airports by analyzing their vocalizations. The system extracted features from recorded bird sounds and used machine learning algorithms to classify new bird songs based on the pre-trained model. The system demonstrated effectiveness in both long and short audio fragments, helping mitigate issues related to bird activity near critical infrastructure like airports.

In article [15] author proposed a transfer learning approach using the Wav2vec model for bird species classification, focusing on multi-label classification from real-time audio recordings. They fine-tuned the Wav2vec model, which enhanced the model's ability to learn bird species' pitch and other sound characteristics, achieving an F1-score of 0.89 using the Xeno-Canto dataset. This study emphasized the importance of transfer learning, especially when there is limited labeled data available, and demonstrated that Wav2vec can capture long-range relationships in audio data more effectively.

Lastly in [16] the author utilized Convolutional Neural Networks (CNNs) to classify bird species based on audio signals. Their study acknowledged the challenges in audio classification, such as handling background noise and variations in bird calls. They used Short-Time Fourier Transform (STFT) and mel spectrograms for feature extraction, converting the audio recordings into visual representations for the CNN model to process. While they achieved high training accuracy, the model faced challenges with validation accuracy due to overfitting

IV. PROPOSED METHODOLOGY

The main purpose of the conversation is to predict the species of birds based on their voice/audio. The suggested framework contains five main phases, as illustrated in Figure 1:

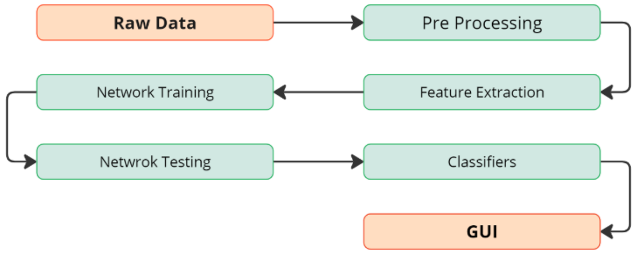

Figure 1. Steps or phases of the model.

1. Data is collected from a set of respected data.

2. Data preprocessing techniques such as clipping and denoising are used in the dataset.

3. The data is processed using the artificial neural network (ANN) approach.

4. The output model will appear on the screen.

This pipeline ensures that the system can handle the various challenges posed by real-world bioacoustics data.

A. Data sets

The first step of the implementation is to collect data from a dataset obtained from XENO-canto/Kaggle. Audio recordings of birds in MP3 format are included in this resource. This dataset contains audio recordings of birds in MP3 format. XENO-canto/Kaggle are open-source websites dedicated to datasets where users upload their own data. As part of our study, a dataset of bird sounds is needed. Bird gender, species, subspecies, locality, type, color, size and sound quality are labeled in this dataset (from A to E, where A is the best quality). Since many characteristics are defined in the data set, their combination is used to define a class (such as genus and species etc.) and classify the birds based on them.

B. Data/Sound pre-processing

Pre-processing is an essential step to ensure that the audio data is clean and ready for analysis. Raw audio recordings often contain noise from environmental sources such as wind, rain, or human activity, which can interfere with species identification. To address this, the raw recordings were first converted to a uniform WAV format to ensure consistency across different sources. Noise reduction techniques were then applied to filter out unwanted sounds. In particular, we used silence removal algorithms to exclude periods of non-vocalization, ensuring that only the relevant bird calls were analyzed.

"Audio normalization and noise filtering were essential to prepare the dataset for accurate classification," as they allowed the model to focus on the relevant acoustic signals while ignoring extraneous noise. After noise reduction, the audio data was segmented into smaller frames, each representing a window of the sound signal. This segmentation allowed for a more detailed analysis of the bird calls, particularly when differentiating between species with similar vocalizations.

C. Feature Extraction (MFCC)

Mel-frequency cepstral coefficients (MFCCs) were used as the primary features for bird species classification. MFCCs are widely recognized for their ability to capture the essential spectral characteristics of sound, making them particularly effective for tasks such as speech and bioacoustic analysis. Each audio frame was transformed into a 40-dimensional MFCC vector, which serves as the input to the neural network.

"MFCCs provide a compact representation of the acoustic properties of bird songs," allowing the system to capture subtle differences in species vocalizations. This feature extraction method ensures that even bird species with similar calls can be differentiated based on their unique acoustic profiles. The use of MFCCs also allows the system to handle noisy data more effectively, as the transformation focuses on the most relevant features of the audio signal.

D. Model Architecture (Artificial Neural Network)

The classification model was based on an Artificial Neural Network (ANN) architecture. The ANN consisted of several fully connected layers, each with ReLU (Rectified Linear Unit) activation functions, followed by a SoftMax output layer for multi-class classification. The softmax layer ensures that the output of the model is a probability distribution over all possible bird species, with the species having the highest probability being selected as the predicted class.

The design of the ANN was chosen to balance computational efficiency with high classification accuracy. "By using an ANN, we are able to achieve strong performance even in noisy environments, while keeping the model lightweight enough to be deployed in field research settings"[10][13]. The model was trained using the categorical cross-entropy loss function, which is well-suited for multi-class classification tasks. The Adam optimizer was employed to accelerate convergence and improve the efficiency of the training process.

E. Training and Testing

The dataset was divided into training and testing subsets, with 90% of the data used for training and 10% reserved for testing. This split ensures that the model has sufficient data for training while still leaving enough data to evaluate its generalization ability. A training regime of 100 epochs with a batch size of 32 was employed to ensure that the model achieved optimal performance without overfitting. By employing early stopping, we were able to prevent overfitting, ensuring that the model remains robust when applied to new data.

Figure 3. Training and Validation Accuracy

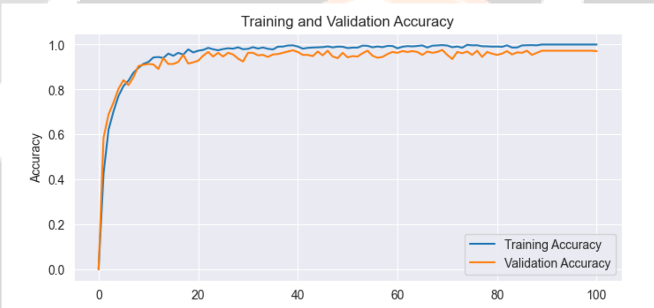

Figure 2. Precision-Recall Curve

The training process involved adjusting the weights of the neural network based on the error between the predicted and actual species labels.

Over time, the model learned to minimize this error, as reflected in the consistent improvement of training and validation accuracy across epochs. The accuracy curves showed a rapid increase during the initial epochs, with both training and validation accuracy stabilizing at a high level near 1.0. This indicates that the model was able to generalize well to unseen data, confirming its capability to learn from the training data effectively without significant overfitting.

The precision-recall curves further validated the model's performance, demonstrating that the system maintained high precision and recall for all bird species classes. The curves were consistently close to the top-right corner for all species, indicating that the model not only accurately predicted bird species (high precision) but also successfully identified most of the actual instances (high recall). This confirms the model's effectiveness in handling complex classification tasks even in the presence of noise or overlapping bird songs.

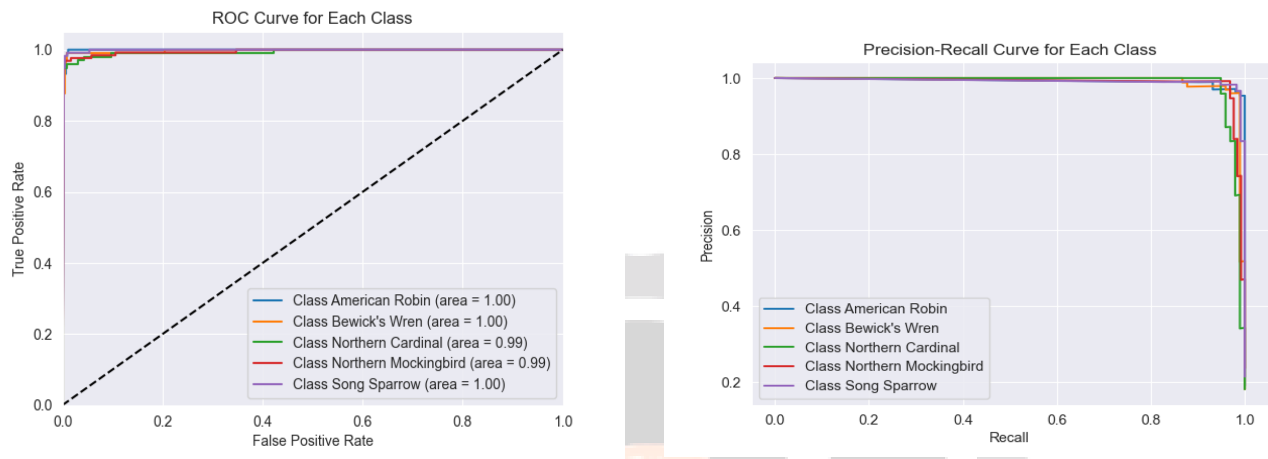

Figure 5. ROC Curve

American Robin	0.95	0.99	0.97	103
Bewick's Wren	0.96	0.97	0.96	98
Northern Cardinal	1.00	0.94	0.97	98
Northern Mockingbird	0.98	0.97	0.97	128
Song Sparrow	0.97	0.98	0.97	116
accuracy			0.97	543
macro avg	0.97	0.97	0.97	543
weighted avg	0.97	0.97	0.97	543

Figure 4. Precision Recall and F1 Score

Once trained, the model was tested on the reserved dataset to evaluate its performance in terms of accuracy, precision, recall, F1-score, and ROC-AUC (Receiver Operating Characteristic – Area Under the Curve). The high scores in these metrics, along with the impressive precision-recall curves, confirm that the model is robust and capable of accurately classifying bird species in real-world bioacoustic environments.

User Interface Implementation

To ensure the system is accessible to a broad audience, including conservationists, researchers, and citizen scientists, a simple yet intuitive user interface (UI) was developed. The UI allows users to interact with the system without requiring specialized technical expertise, making the tool suitable for both experts and non-experts.

The key functionalities of the UI are as follows:

1. **Audio File Upload**:

Users can upload audio recordings of bird songs directly through the interface. The system supports WAV files, and users can easily browse and select the recordings they wish to analyze. This step is simplified to ensure that the system can be operated by users with minimal technical knowledge.

2. **Species Prediction**:

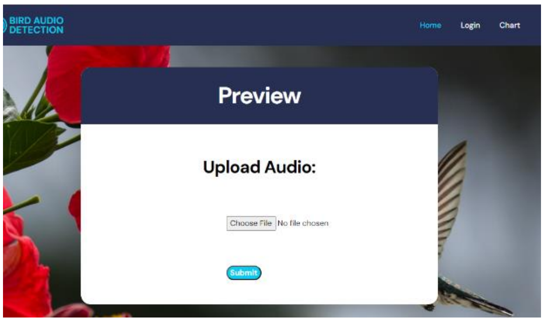

Figure 6. User Interface

After analyzing the uploaded audio file, the system predicts the bird species and displays the result. In addition to the species name, the UI presents an image of the predicted bird, along with supplementary information such as the bird's size, typical habitat, and geographical region.

V. RESULT

The ANN model achieved impressive classification results during testing, with an overall accuracy of 97%. Key performance metrics such as precision, recall, and F1-scores were high across all bird species (as shown in Figure 6), indicating that the model effectively handles variability in bird songs. The Training and Validation Accuracy, as shown in Figure 2, reveals only a small number of misclassifications, demonstrating the robustness of the model in distinguishing between species even in noisy environments.

The ROC (Receiver Operating Characteristic, Figure 4) curves and precision-recall curves further confirmed that the model performs well across a variety of conditions. The ANN model outperformed traditional classifiers like SVM, particularly in handling complex and noisy audio inputs. "In comparison to traditional classifiers like SVM, the ANN model demonstrated superior performance in handling diverse bird species and environmental noise"[5][8].

VI. ADVANTAGES

Systematic recordings of outside noises at the moment are feasible way to automatic audio recorders, that have lately spread-out new possibilities for environmental conservation and restoration. Due to the truth that many chook species have rather excessive vocal frequency, audio recordings have emerged as one of the simplest strategies to do have a look at on them. Biological sounds of birds can provide exact and standardized statistics at the dynamics and distribution of flora and fauna habitats. Audio studies and surveys are an amazing device to investigate the species' density, abundance, and occupancy due to the fact many chook species produce wonderful and steady sounds. Furthermore, picturized tracking is intricate for plenty small and touchy birds, enigmatic species, and species dwelling in environments that environmentalists locate tough to reach.

VII. LIMITATIONS

Audio signals provide much more information about a bird, as it can be further classified in songs, calls and sounds, so noise in these different types of audios can be a problem. Having this kind of extra set of properties and classifications, also makes identification of birds little difficult.

VIII. FEASIBILITY STUDY

The use of Convolutional Neural Networks (CNNs) and other AI models offers significant advancements in automated bird sound recognition, with various practical applications:

1. **Ornithology and Biological Research:**

Automated bird sound recognition aids ornithologists and researchers by providing continuous monitoring of bird populations, reducing the time and cost involved in manual identification. This technology supports government agencies and conservationists in biodiversity studies.

2. **Birdwatching and Tourism:**

   With birdwatching being a popular activity, such systems have commercial potential, enhancing experiences in nature parks or as portable tools for enthusiasts.

3. **Integration with Hardware:**

   AI models can be implemented on devices like Raspberry Pi, making them suitable for wildlife sanctuaries and environmental parks, allowing for real-time bird monitoring in remote areas.

4. **Mobile Applications:**

   Developing an Android app enables users to identify bird species using their smartphones, making bird recognition accessible to both amateurs and professionals.

5. **Data Collection:**

   Collected data can be stored locally or in the cloud, contributing valuable insights for studying bird migration, population, and diversity.

These applications highlight the potential of automated systems to significantly impact ornithological research, conservation, and public engagement with bird species. The use of CNNs and ANNs in bird species identification has practical applications in conservation, tourism, and environmental studies. Future work may involve expanding the system to include more species.

IX. Discussion

The results of this study demonstrate the effectiveness of using MFCC features in conjunction with an ANN for bird species classification. "ANNs are particularly well-suited to tasks like bird species classification because they can handle the inherent variability in natural environments," making them ideal for bioacoustics monitoring. While CNNs have shown promise in other classification tasks, our results show that ANNs offer a computationally efficient solution that can still achieve high accuracy, making them suitable for practical applications in field research.

One of the strengths of this system is its ability to handle noisy environments, where traditional classifiers often fail. By employing effective pre-processing techniques and utilizing MFCCs as input features, we ensured that the model could differentiate between species even when background noise is present. This capability is crucial for real-world applications, where audio recordings are often taken in uncontrolled environments.

X. CONCLUSION

In this study, we presented an automated bird species identification system that utilizes MFCC features and an Artificial Neural Network. The system achieved high classification accuracy, even in noisy environments, demonstrating its potential for large-scale bioacoustics monitoring. With further improvements, including the integration of unsupervised learning techniques and the extension to more bird species, this system could become a valuable tool for conservationists and ornithologists. "By automating bird species identification, we can significantly enhance our ability to monitor ecosystems and ensure the protection of biodiversity on a global scale"[8][14].

XI. FUTURE SCOPE

This research can be expanded in several ways to improve its practical application and performance. Some possible directions include:

1. Incorporating More Species: The current system can be trained on a larger variety of bird species to improve its applicability in different regions. Expanding the dataset with more bird vocalizations will enhance the model's generalization abilities.

2. Real-time Classification: Implementing real-time bird species identification would allow the system to be used in field research and monitoring, enabling immediate species detection during fieldwork.

3. Mobile Application Development: Developing a mobile app for this system would make it easier for conservationists, researchers, and bird enthusiasts to use the model on the go. Mobile support can increase the system's reach and make it more accessible.

4. Unsupervised Learning: The system could benefit from unsupervised learning methods to classify species without requiring pre-labeled datasets. This would make it possible to identify unknown species or recognize species from incomplete data.

5. Improving Accuracy in Noisy Environments: Further research can focus on improving the system's ability to classify bird species accurately in environments with higher levels of noise or overlapping sounds.

6. Integration with Other Bioacoustics Tools: The system could be integrated with other wildlife monitoring tools, such as camera traps and environmental sensors, for comprehensive ecosystem monitoring.

7. Climate and Migration Studies: The system could also be adapted to study the effects of climate change on bird migration patterns by tracking species over time.

REFERENCES

[1] Vivekanandam, B. "Speedy Image Crowd Counting by Light Weight Convolutional Neural Network." *Journal of Innovative Image Processing* 3, no. 3 (2021): 208-222.

[2] Aarti Madhavi, R. P. (2018). Deep Learning Based Audio Classifier for Bird Species. IJSDR

[3] Incze, A., Jancso, H.-B., Szilagyi, Z., Farkas, A., & Sulyok, C. (2018). Bird Sound Recognition Using Convolutional Neural Network. IEEE 16th International Symposium on Intelligent Systems and Informatics, 000295–000300.

[4] Jancovic, P., & Kokuer, M. (2015). Acoustic recognition of multiple bird species based on penalised maximum likelihood. IEEE Signal Processing Letters, 1-1.

[5] Rai, P., Golchha, V., Srivastava, A., Vyas, G., & Mishra, S. (2016). An automatic classification of bird species using audio feature extraction and support vector machines. International Conference on Inventive Computation Technologies, 1-5.

[6] Graciarena, M., Delplanche, M., Shriberg, E., & Stolcke, A. (2011). Bird species recognition combining acoustic and sequence modeling. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 341-344.

[7] Manoharan, J. Samuel. "Study of Variants of Extreme Learning Machine (ELM) Brands and its Performance Measure on Classification Algorithm." Journal of Soft Computing Paradigm (JSCP) 3, no. 02 (2021): 83-95.

[8] Chandu B, A. M. (2020). Automated Bird Species Identification using Audio. 2020 International Conference on Artificial Intelligence and Signal Processing (AISP).

[9] Evangelista, T. L., Priolli, T. M., Jr, C. N., Angelico, B. A., & Kaestner, C. A. (2014). Automatic Segmentation of Audio Signals for Bird Species Identification. IEEE International Symposium on Multimedia,223–228.

[10] Narasimhan, R., Fern, X. Z., & Raich, R. (2017). Simultaneous Segmentation and Classification of Bird Song Using Cnn. IEEE Conference, 146-150.

[11] C., A. P. (2015). Automatic Recognition of Birds Through Audio Spectral Analysis. Fifth International Conference on Advances in Computing and Communication, 395–398.

[12] Guillermo Sarasa, A. G. (2017). An Approach of Algorithmic Clustering Based on String Compression to Identify Bird Songs Species in Xeno-canto Database. 3rd International Conference on Frontiers of Signal Processing.

[13] Jancovic, P., Kokuer, M., & Russell, M. (2014). Automatic Detection of Bird Species from Audio Field Recordings using HMM-based Modelling of Frequency Tracks. EUSIPCO, 8252 8256.

[14] Lopes, M. T., Gioppo, L. L., Higushi, T. T., Kaestner, C. A., Silla Jr., C. N., & Koerich, A. L. (2011). Automatic Bird Species Identification for Large Number of Species. IEEE International Symposium on Multimedia, 117-122

[15] Swaminathan, B., Jagadeesh, M., & Vairavasundaram, S. (2024). Multi-label classification for acoustic bird species detection using transfer learning approach. *Ecological Informatics*, *80*, 102471.

[16] Wang, J., & Goldsztein, G. (2023). Audio Classification of Bird Species Using Convolutional Neural Networks. *Journal of Student Research*, *12*(1).