IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

A STATISTICAL APPROACH USING HARNESSING AUDIO FOR AVIAN IDENTIFICATION OF BIRD SPECIES IDENTIFICATION TECHNIQUES

KOKANE PRAJAKTA RAMDAS[1], ASST.PROF. V. S. KARWANDE[2]

ME Student, Department of Computer Science & Engineering, EESGOI, India[1]
HOD, Assistant Professor, Department of Computer Science and Engineering, EESGOI, India. [2]

Abstract: This project aims to advance avian identification through the utilization of audio signals, employing a comprehensive approach that integrates with deep learning techniques. The study focuses on developing three distinct models: Ridge Classifier, Support Vector Machine (SVM) Classifier, and Artificial Neural Network (ANN). By converting audio signals into numerical data using standard Python libraries, we aim to create robust models capable of accurately identifying bird species based on their distinct audio signatures. The Ridge and SVM classifiers, representing classical ML paradigms, will leverage advanced feature extraction methodologies. Simultaneously, the ANN model, a deep learning approach, will harness the power of neural networks to capture intricate patterns within the audio data. This interdisciplinary fusion of ML and deep learning techniques is poised to significantly enhance avian species identification, contributing to biodiversity monitoring and conservation efforts.

Keywords: *Bird Spices Classification; Audio Signal Processing; SVM, Ridge Classifier; Artificial Neural Network.*

I INTRODUCTION

The Birds, as integral components of ecosystems, play a crucial role in maintaining biodiversity and ecological balance. Effective conservation of avian populations requires accurate monitoring, traditionally achieved through labor-intensive methods like visual surveys and mist netting. However, these conventional approaches are time-consuming and may not be feasible for large-scale, prolonged observation. Embracing technological advancements, particularly in the realm of audio signal processing, offers a transformative solution for bird species identification.

This project report delves into the contemporary landscape of automated bird species identification, scrutinizing the latest advancements in the application of audio signal processing techniques. The core of our investigation revolves around the advancement of deep learning methodologies, which holds promise in achieving accurate identification of a wide array of bird species.

The challenges inherent in bird species identification, including the extensive variety of avian species, the dynamic nature of bird songs, and the omnipresence of background noise, are meticulously addressed in this paper. Despite these hurdles, recent research indicates that automated bird species identification can achieve remarkable accuracy, with top-performing models surpassing 90% accuracy on certain datasets. This breakthrough opens avenues for revolutionizing bird monitoring strategies.

The significance of automated bird species identification lies in its potential to revolutionize the monitoring of avian populations across expansive territories and protracted timeframes. Moreover, its ability to identify elusive birds, such as nocturnal species and those residing in dense vegetation, adds a dimension of observation that is unattainable through traditional visual methods. The amalgamation of technology and ecology, as explored in this report, positions automated bird species identification as a promising tool for bolstering conservation efforts.

As we navigate through this discourse, it becomes evident that the continual development of technology is poised to yield even more precise and efficient bird identification systems. The trajectory of progress in this field signals a transformative era in bird monitoring, promising advancements that align seamlessly with the goals of ecological preservation and biodiversity conservation. In the subsequent sections, we unravel the intricacies of various audio signal processing techniques, dissect their applications, and envision a future where the fusion of technology and ecology redefines the landscape of bird species identification and conservation.

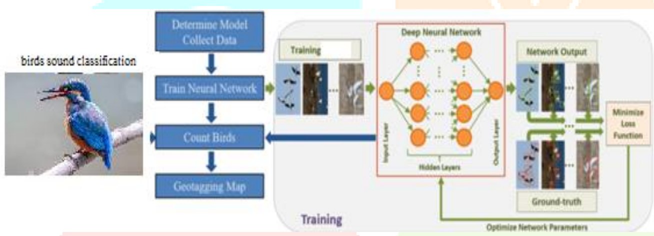

Fig 1: The complete workflow of the automatic bird monitoring system

II LITERATURE SURVEY

In this research Presenting a novel approach to computational bioacoustics, this study [1], Acconcjaioco et al. introduce one-shot learning. The research focuses on addressing the challenge of bird species identification in dynamic environments where the class dictionary is incomplete. The proposed framework employs a Siamese Neural Network on logMel spectrograms, demonstrating its ability to effectively detect and incorporate novel class's on-the-fly. Through evaluations on diverse bird datasets, the study showcases the state-of-the-art performance of the framework in both stationary and non-stationary conditions, utilizing a standard audio representation. The key to the method's success lies in its simultaneous consideration of similarities and dissimilarities to known classes. Additionally, the study explores the application of the method to analogous problems, investigates optimal data quantities for non-stationary performance, and delves into strategies for optimal class selection during training.

In their paper [2], Alqahtani et al. conduct a comprehensive review of deep time-series clustering (DTSC), examining challenges and advancements in analyzing time-series data with deep learning. Focusing on movement behavior clustering through modified deep clustering architectures, the authors address issues like varying time scales, noise, and dimensionality, showcasing how deep learning effectively addresses these challenges. The study, illustrated with the Imperial Cormorant bird dataset, yields promising results in enhancing clustering performance. The conclusion discusses challenges, compares architectures, reviews state-of-the-art methods.

In their paper [3], Baevski et al. present a self-supervised learning framework for speech representations, surpassing semi-supervised methods. Employing masking in the latent space and a contrastive task on quantized representations, the model achieves impressive 1.8/3.3 WER on Librispeech with all labeled data, outperforming the state of the art with 100 times less labeled data. Notably, using just 10 minutes of labeled data results in a WER of 4.8/8.2. The approach excels in noisy environments, with potential enhancements using a seq2seq architecture and word piece vocabulary. This work establishes the effectiveness of pre-training on unlabeled data, demonstrating promise in scenarios.

In their research [4], Bravo Sanchez et al. investigate the application of SincNet, efficient neural network architecture, for bioacoustic avian call classification directly from raw sound waveforms. In contrast to traditional methods relying on extracted features, SincNet learns from the waveform itself, achieving over 65% accuracy on the NIPS4Bplus bird sound dataset with limited data. While comparable to traditional methods after hyperparameter tuning, SincNet's efficiency stands out. Automatically selecting relevant elements from the raw waveform eliminates the need for manual feature extraction, making it a promising tool for bioacoustic analysis.

The study by Chen et al. [5] introduces WavLM, a large-scale self-supervised pre-trained model designed for comprehensive speech processing tasks. To address the universal representation learning challenge across diverse speech facets, WavLM incorporates masked speech prediction and denoising during pre-training. Employing a Transformer structure with gated relative position bias and an expanded dataset from 60k to 94k hours, WavLM achieves state-of-the-art results on the SUPERB benchmark. Notably, it demonstrates significant enhancements in speaker verification, speech separation, and speaker diarization. Extending the HuBERT framework, WavLM emerges as a versatile backbone network for both automatic speech recognition (ASR) and non-ASR tasks, showcasing its potential as a next-generation solution.

In the study by Chhaya et al. [6], the concept of "Community Bioacoustics" is explored, delving into the use of acoustic signals within animal communities to unravel ecological dynamics. The paper illuminates how diverse signaling outcomes are shaped by ecological processes, with a focus on signal space and partitioning. By utilizing acoustic data, the research provides non-intrusive insights into the organization and turnover of biological communities. Emphasizing its value in ecological monitoring and conservation across various landscapes and taxa, the review introduces the potential of passive acoustic monitoring. It proposes an analytical framework that enables the comprehensive understanding of ecosystems, spanning from insects to large whales, through the lens of acoustic community structure.

In the investigation conducted by Choi et al. [7], the focus lies on the territorial hoot calls of male tawny owls in South Korea, aiming to evaluate their potential as a monitoring tool. The study unveils distinctive individual-specific calls, with a particular emphasis on the discriminative capabilities of the third note. Notably, acoustic distances between males within the same territory were smallest for nesting sites, suggesting a pattern of site fidelity. These calls, easily captured during breeding seasons, provide traceable identities through visual and statistical analyses. The research proposes the utilization of tawny owl territorial calls as a valuable acoustic monitoring tool for assessing individual site fidelity. It advocates for long-term studies and the establishment of a sonogram catalog to bolster conservation efforts for this endangered species in South Korea.

In the research conducted by Deng et al. [8], the investigation delves into the stability of vocal characteristics in individual male Common Cuckoos throughout a breeding season. The study, encompassing 1032 syllables from 30 males in Northeast Asia, reveals a high consistency in same-day calls, with a correct identification rate of 93.6%. However, over a 19-day period, this rate declines sharply to 40.5%, indicating reduced stability in vocal characteristics. While call consistency is repeatable within successive bouts, it diminishes with additional bouts or days. The findings highlight concerns regarding the reliability of individual male cuckoo calls for monitoring and underscore the necessity to account for within-season variability in vocal characteristics for accurate non-invasive identification.

In the research conducted by Enari et al. [9], an evaluation is made regarding the efficiency of passive acoustic monitoring (PAM) and camera traps in detecting sika deer and Japanese macaques across seven sites in eastern Japan. While fully automated PAM exhibited false positives, the introduction of semi-automated PAM, incorporating manual validation, significantly enhanced detection rates. PAM demonstrated superior performance over camera traps in both detectability and recall rates, with detection areas 100 to 7,000 times wider. The strength of PAM lies in its capacity to offer socio-behavioral data through inter individual vocal communications, thereby enriching insights into population dynamics and group compositions. The study underscores PAM's potential as a reliable and versatile method for large-scale wildlife monitoring.

In their paper [10], Florentin et al. present a novel approach to the automated detection and identification of European woodpeckers' drumming and calls. Leveraging deep convolutional neural networks (CNNs) and the Acoustic Complexity Index, the study transforms sounds into spectrogram images, achieving high accuracy in drum detection and call identification. While CNNs excel in call identification, a simpler method outperforms in drum identification. The research underscores the significance of large training sets and suggests the direct construction of deep networks for sound processing. Tested on continuous field recordings, these methods provide unprecedented insights into woodpecker behavior, evolution, and conservation. This highlights the practicality of deep networks in ornithological research, particularly for monitoring European woodpeckers.

III. SYSTEMS ARCHITECTURE

The system architecture for Bird Species Identification using an Artificial Neural Network (ANN) Classifier comprises several integral components. Initially, audio recordings of bird species are collected and subjected to preprocessing, involving tasks like data balance and feature extraction, typically in the form of balanced spectrograms.

The extracted features serve as input to the ANN classifier, the cornerstone of the system. This computational model, inspired by neural structures in the human brain, learns intricate patterns during a dedicated training phase. Here, labeled training data is fed to the ANN, enabling it to adjust internal parameters for accurate species classification. Subsequently, the system's performance is evaluated through separate validation and test sets, gauging its generalization capabilities. Post-processing techniques may be applied for further result refinement, while a user interface can offer an interactive platform. The system then presents identified bird species, along with relevant information, providing a valuable tool for various applications, from research to conservation efforts. Continuous monitoring and refinement mechanisms ensure the system's accuracy and effectiveness over time.

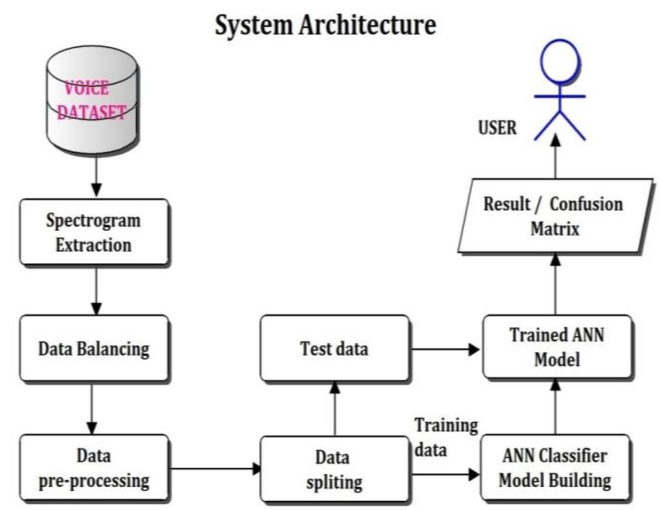

Figure No 3.1: System Architecture

Voice dataset collection: A dataset of bird calls and songs is collected. This dataset can be collected from a variety of sources, such as online repositories, field recordings, and citizen science projects.

Data preprocessing: The audio data is preprocessed to remove noise and other artifacts. This may involve filtering, normalization, and segmentation.

Spectrogram extraction: A spectrogram is extracted from each audio segment. A spectrogram is a visual representation of the audio signal, and it can be used to identify key features of the bird call or song.

Data balancing: The dataset may be unbalanced, with some bird species being overrepresented and others being underrepresented. To address this, the dataset can be balanced by oversampling the underrepresented species or under sampling the overrepresented species.

Test data splitting: The dataset is split into two sets: a training set and a test set. The training set is used to train the deep learning model, and the test set is used to evaluate the performance of the model on unseen data.

Feature extraction: Features are extracted from the spectrograms. These features can be based on a variety of acoustic properties, such as pitch, duration, and spectral shape.

ANN model building: An artificial neural network (ANN) model is trained to classify the bird calls or songs into different species.

ANN classifier: The trained ANN classifier is used to identify bird species from new audio recordings.

The system architecture diagram shows the following components:

Voice dataset: This component represents the collection of audio recordings of bird calls and songs.

Spectrogram extraction: This component represents the process of extracting spectrograms from the audio recordings.

Data balancing: This component represents the process of balancing the dataset to ensure that all bird species are equally represented.

Test data splitting: This component represents the process of splitting the dataset into training and test sets.

Data preprocessing: This component represents the process of preprocessing the audio data to remove noise and other artifacts.

Feature extraction: This component represents the process of extracting features from the spectrograms.

ANN model building: This component represents the process of training the ANN model to classify the bird calls or songs into different species.

ANN classifier: This component represents the trained ANN classifier that is used to identify bird species from new audio recordings.

Result/confusion matrix: This component represents the results of the ANN classifier, as well as a confusion matrix, which shows how well the classifier is able to distinguish between different bird species.

IV EXPERIMENTAL RESULTS

The Accuracy plot graph as in Fig illustrates the training and validation accuracy of an Artificial Neural Network (ANN) model designed for bird species identification using audio recordings. The x-axis denotes the number of epochs, indicating iterations over the training data, while the y-axis represents the model's accuracy in identifying bird species. Two lines are plotted: the blue line depicts training accuracy, measuring performance on the training data, while the orange line represents validation accuracy, gauging performance on unseen data. Ideally, both accuracies should rise steadily with epochs. However, after approximately 50 epochs, the training accuracy reaches about 95%, while validation accuracy lags slightly at around 90%, suggesting potential overfitting where the model learns training data specifics rather than generalizable patterns. Key considerations include dataset specificity, recognizing overfitting signs like declining validation accuracy despite rising training accuracy, and employing techniques like regularization to mitigate overfitting. Adjusting the number of epochs based on dataset and model architecture is crucial for optimal performance.

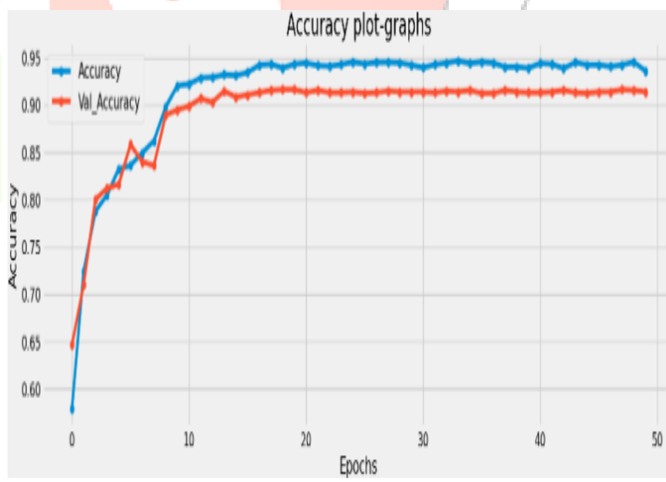

Figure No 4.1: Accuracy plot graph of Artificial Neural Network

The loss plot graph as provided in Fig. illustrates the loss plot for an Artificial Neural Network (ANN) algorithm utilized in bird species identification. In machine learning, loss measures the disparity between model predictions and actual values, with lower loss indicating better performance. The x-axis denotes epochs, representing iterations over the training data, while the y-axis shows loss, with two lines plotted: blue for training loss and orange for validation loss. Ideally, both should decrease with epochs. In this graph, both training and validation loss decline with increasing epochs, suggesting the model effectively learns data patterns and improves performance. Important considerations include determining the optimal number of epochs based on dataset specifics and model architecture, detecting overfitting indicated by increasing validation loss despite decreasing training loss, and implementing techniques like regularization to mitigate overfitting.

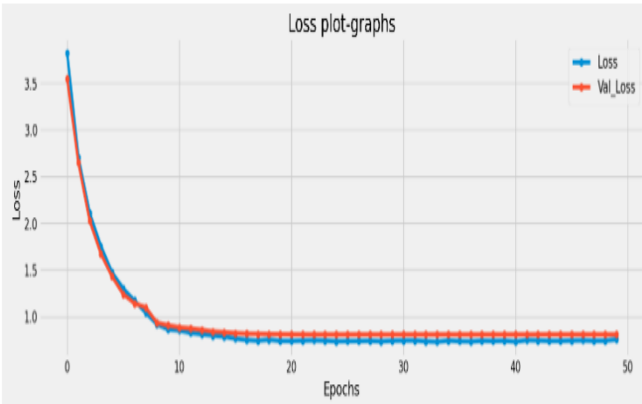

Figure No 4.2: Loss plot graph of Artificial Neural Network

V CONCLUSION

This project underscores the critical role of audio signals in revolutionizing avian species identification for effective biodiversity monitoring and conservation. The comprehensive review traverses a spectrum of techniques, from traditional acoustic analyses to cutting-edge deep learning algorithms, with a particular emphasis on the advancements facilitated by deep learning models. The significance of audio as a potent tool in ornithology and wildlife monitoring is evident, yet challenges such as background noise and variability in vocalizations necessitate ongoing refinement. This interdisciplinary exploration at the intersection of ornithology, acoustics, and technology contributes to a holistic understanding of avian identification. By synthesizing diverse information, the study empowers researchers, conservationists, and technologists in their collective mission to enhance avian species identification, thus fortifying our capabilities in the preservation of avian biodiversity.

REFERENCES

1.Samruddhi Bhor,Rutuja Ganage,Omkar Domb,Hrushikesh Pathade and Shilpa Khedkar ,"Automated Bird Species Identification using Audio

Signal Processing and Neural Network",ICEARS,2022.

2. Chandu B, Akash Munikoti, Karthik S Murthy, Ganesh Murthy V and Chaitra Nagaraj,"Automated Bird Species Identification using Audio

Signal Processing and Neural Networks",AISP,2020.

3. Hritik Tomar, Kartik, Mayank Bhama and Dr Rajesh kumar yadav,"Automated Bird Species Identification Using Acoustic Features and Neural Network Classification",

Eur. Chem. Bull.,Special Issue 4, 2023.

4. Gitanjali Pote, Ashwini Sase, Puja Barwal, Pooja Nandre and Prof. Neelam Joshi,"Automatic Bird Recognition using Signal Processing and Neural

Networks",IJRPR,Vol 4, no 5,May 2023.

5. Tanay Sonavane, Kireet Verma and Hrithik Wadile,"Automated Bird Species Identification Using Deep Learning With Image and Audio",

IJMR,2022.

6. Jiri Stastny, Michal Munk and Lubos Juranek,"Automatic bird species recognition based

on birds vocalization",EURASIP,2018.

7. Mario Lasseck,"Audio-based Bird Species Identification with Deep

Convolutional Neural Networks",CEURWS,2017.

8. Ricky Mohanty a, Bandi Kumar Mallik b and Sandeep Singh Solanki, "Recognition of bird species based on spike model

using bird dataset",Elsevier,2020.

9. Raja Shekar Kadurka,"Automated Bird Detection in Audio Recordings by a Signal Processing Perspective",

Int. J.Adv.Sig.Img.Sci, Vol.7, No.2, 2021.

10.Loris Nanni,Rafael L. Aguiar, Yandre M.G. CostaSheryl Brahnam, Carlos N. Silla Jr., Ricky L.Brattin and Zhao Zhao,"

Bird and whale species identification using sound images",IET Computer Vision,2017.

11.Bhuvaneswari Swaminathan, M. Jagadeesh and Subramaniyaswamy Vairavasundaram,"Multi-label classification for acoustic bird species detection using transfer

learning approach",Elsevier,2023.

12. Dan Stowell,Michael D. Wood, Hanna Pamuła,Yannis Stylianou and Herve Glotin,"Automatic acoustic detection of birds through deep learning:

The first Bird Audio Detection challenge",wileyonlinelibrary,2019.

13. Seppo Fagerlund ,"Bird Species Recognition Using Support Vector Machines",

EURASIP Journal on Advances in Signal Processing Volume, 2007.

14.Prof. Pooja Wale, Abhishek Mankar, Pratik Padale, Sanket Gawade and Prasanna Ghogare,"A Survey On Bird Species Identification Using

Audio Signal Processing And Neural Network",IJARCCE,2022.

15.Shaokai Zhang , Yuan Gao, Jianmin Cai, Hangxiao Yang, Qijun Zhao and Fan Pan,"A Novel Bird Sound Recognition Method Based on

Multifeature Fusion and a Transformer Encoder",MDPI, 2023.