# Visual and Acoustic Identification of Bird Species

**Dhiraj Patil[1], Rutwi Bodhe[2], Rupali Pawar[3], Tanisha Doshi, Vidhya Vasekar**

*Department of Computer Engineering, Sinhgad Institute of Technology and Science, Narhe, Pune 411041.*

---

**Abstract -** *This paper combines both approaches for bird species identification by extracting visual features from bird images and acoustic features from bird calls. Some bird species are rarely found in certain regions, and it's difficult to track them if done the prediction is difficult. In order to withstand this issue, we've come across a significant and easier way to recognize these bird species based on their features. We've used BirdCLEF 2022 dataset for the audio segment and the BIRDS 400 dataset for the image segment for the training and testing parts. Since among most of the approaches, we have studied CNN as vanquishing, therefore we've used CNN for both visual as well as acoustic identification. CNN is the strong assemblage of ML which has proven efficient in image processing. Our project has become attractive because of the techniques and recent advances within the domain of deep learning. With novel pre-processing and data augmentation methods, we train a convolutional neural network on the largest public obtainable dataset. By establishing a dataset and using the rule of similarity comparison algorithms, our system can provide the best results. By using our system, everyone will simply be able to determine the species of the particular bird which they provide image/audio or both as input.*

***Key Words***: **Bird Species, CNN, ML**

## 1. INTRODUCTION

In total there are 9000 bird species in our world. As the “extinction capital of the world,” Hawai'i has lost 68% of its bird species, the results of which might harm entire food chains[7]. Bird species identification arouses interest in different groups of admirers and experts whether through the beauty of birds and their sound or by their ecological importance. When collaborating with the latest large availability of automated recording units it becomes prominent why remote, systematic and non-intrusive, acoustic biodiversity surveys are getting popular in the past decade. Acoustically active biota groups and their specific information can be obtained by acoustic monitoring, and an index of biodiversity can be generated based on how complex the calls recorded within a region are.

Population monitoring is used by researchers to understand how native birds react to changes in the environment and conservation efforts. Several real-world applications can rely on birds such as monitoring of environmental pollution, assessing the quality of the environment and estimating sustainability indicators. But many of the remaining birds across the islands are isolated in difficult-to-access, high-elevation habitats[8]. With physical monitoring difficult, scientists have turned to automatic image and sound recordings. Known as bio-acoustic monitoring or bird watching, this approach could provide passive, low labour, and cost-effective strategy for studying endangered bird populations. Approaches like these have interesting correct classification rates which range between 78 - 95%, which depends upon the number of bird species that have been tested.

Prediction of bird species, to which category they belong by using image or audio data is known as bird species identification. The recognition of bird species can be possible through audio or image. The use of automated methods for bird identification is an effective way to assess the quantity and diversity of birds that appear in a region. Identification is a challenging problem both for humans as well as for computational algorithms that focus to do this task automatically.

Looking complexity of the problem, a scenario which is visually and acoustically limited, a very high number of classes, high visually and acoustically similar bird species, background noise and a high variety of the acquisition conditions. Novel methods are required to provide more reliable and accurate results than those achieved till now by both visual and acoustic approaches. The problem can be solved by training a mechanism which can make predictions which can be similar to test samples which are prior available.

Providing a solution to this problem, this system has been proposed that uses both .jpg & .wav files to predict bird species from data users put as input.

### 1.1 PROBLEM DEFINITION:

Identifying a bird can be a challenge, even for experienced birders. If you're new in using field guides, it can be difficult to figure out how & were to even start searching in the 100’s of pages of bird species. By some features like size, shape and colour birds can be classified. We can classify species of birds using CNN.

### 1.2 OBJECTIVES

- To increase the correct classification rate & decrease rejection rates, even while the number of bird species gradually increases using CNN.

- To use a transfer learning approach for better identification of bird species.
- To develop a web GUI for the proposed system.

## 2. RELATED WORK

Aditya Khamparia et. al [1], investigate three types of time-frequency representations (TFRs): Mel-spectrogram, harmonic-component based spectrogram, and percussive component-based spectrogram. A different deep learning architecture, SubSpectralNet, is utilized to classify bird sounds. Experimental results on classifying 43 bird species show that fusing selected deep learning models can effectively increase classification performance.

The author proposes an attention-based spatially recursive network that learns to attend to discriminative parts of an object with spatial manipulations. Owing to the capability of integrating features in an interactive manner, bi-linear pooling is adopted to integrate the detector and feature extractor from two independent CNNs. To further capture the spatial relationship amid critical object parts, a spatial recursive encoding with spatial LSTM units is proposed to produce spatially expressive representations of fine-grained objects[2].

Yo-Ping Huang et. al [3], developed a mobile app platform that uses cloud-based deep learning for image processing to identify bird species from digital images uploaded by a user on a smartphone. The learned parameters of bird features were used to identify images uploaded by mobile users. The proposed CNN model with skip connections reached higher accuracy of 99.00% compared with the 93.98% from a CNN and 89.00% from the SVM for the training pictures. With reference to the test dataset, the average sensitivity, and accuracy were 93.79%, 96.11%, and 95.37%, respectively.

The author of this paper, observes the popularity of methods using deep neural networks, and the still widely used Mel Frequency-based representations, with only a few approaches standing out as radically different. It also introduces the calculation of confidence intervals based on a jackknife re-sampling procedure to perform a statistical analysis of the challenge results. The analysis indicates that while the 95% confidence intervals for several systems overlap, there are significant variations in performance between the top systems and therefore the baseline for all tasks. The results presented in this paper indeed show that networks trained with such a small amount of data are learning only the more common classes and have inconsistent behaviour regarding the less common ones[4].

Philip Eichinski et. al [5], provide the survey on Clustering and visualization of long audio clips for fast exploring avian surveys. This paper presents an approach for a surveyor to rapidly scan long time recordings of environmental audio by automatically filtering parts with low activity and repetitions of the same call types. The approach presented in this paper could give ecologists a new approach to handling large datasets.

Chang-Hsing Lee et. al [15], provide a method for the automatic classification of birds into different species. A feature set, TDMFCC (Two-dimensional Mel-frequency cepstral coefficient) and DTDMFCC (dynamic-TDMFCC) are employed as the vocalization features for the classification of each individual syllable segmented from continuous birdsong recordings. The test and train syllables in the experiment are split from various recordings. When both TDMFCC and DTDMFCC are combined together, a classification accuracy of 84.06% can be obtained for the classification of 28 bird species.

## 3. SYSTEM ARCHITECTURE

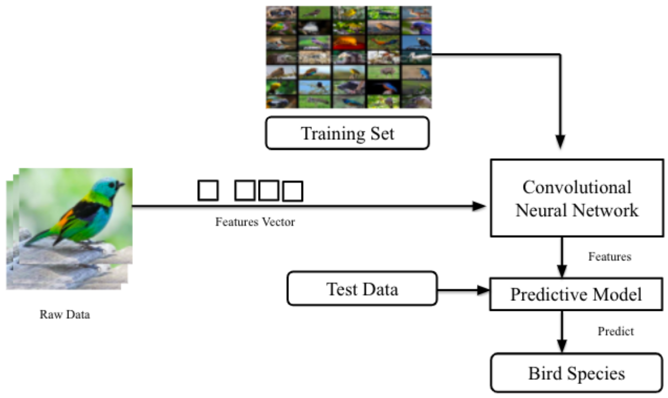


**Fig -1**: Identification through bird image

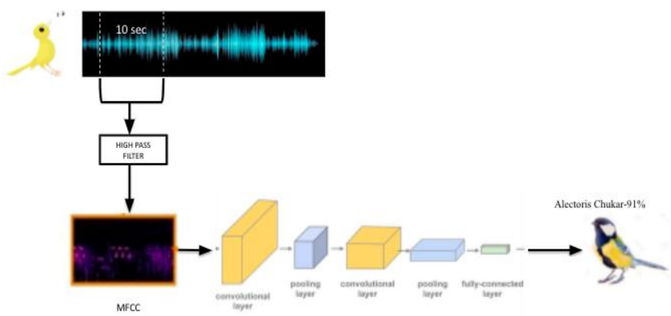


**Fig -2**: Identification through bird calls

## 4. PROPOSED SYSTEM

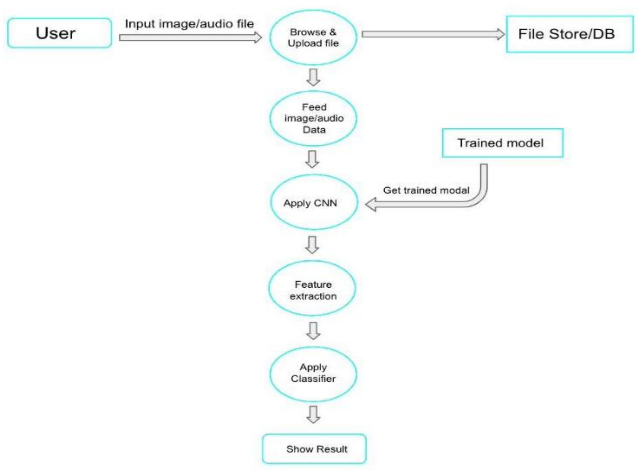


**Fig -3**: Flow of System

The above figure(Fig-3) represents the actual flow of how our proposed system works. To develop such a system a pre-trained model is required to classify an image or audio. The pre-trained model consists of train results and test results. The accuracy of the training dataset is 90%. The testing dataset consists of nearly 500 images with an accuracy of 80%. Further, the dataset is validated with an accuracy of 75% to increase the performance of the system.

When a user uploads an input file on a website, the image or audio is temporarily stored in a database. This input data is then fed to the system and given to CNN where CNN is coupled with the pre-trained model. A CNN has various convolutional layers. Various alignments/features such as head, body, colour, beak, shape, and the entire image and MFCCs of bird are considered for classification to yield maximum accuracy. Each alignment is given through a deep convocational network to extract features from multiple layers of the network. Then CNN is used to classify that image.

## 5. RESULTS

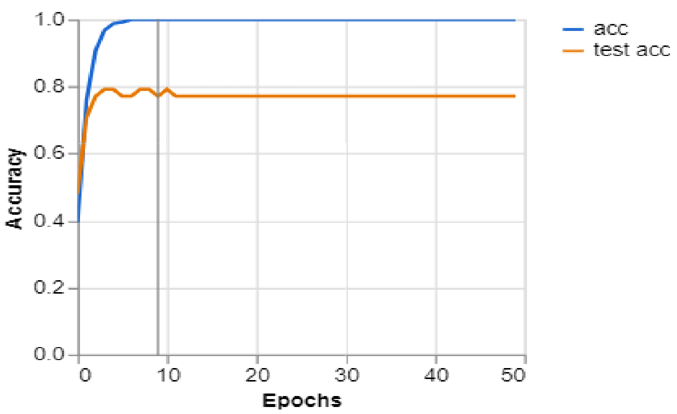


**Fig - 4**: Model Accuracy

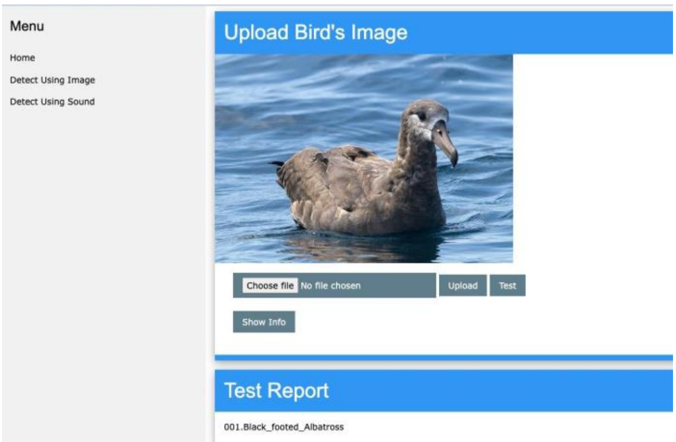


**Fig - 5**: Web-app Home Page

## 6. CONCLUSION

The main idea behind developing the identification website is to build awareness regarding bird watching, birds and identification. It also caters to the need of simplifying the bird identification process and thus making bird watching easier. The technology used in the experimental setup is Convolutional Neural Network (CNN). The result produced by our project, t has provided 80% accuracy for the prediction of bird species.

## 7. FUTURE WORK

- Create an android/iOS app instead of the website which will be more convenient for the user.
- The system can be implemented using the cloud which can store a large amount of data for comparison and provide high computing power for processing (in the case of Neural Networks).
- Real-time processing of bird images/ audio if done would be of great importance.

## REFERENCES

[1] Aditya Khamparia, Babita Pandey, Prayag Tiwari,, Ashish Khanna, Deepak Gupta, Babita Pandey, & Nhu Gia Nguyen, "Sound Classification Using Convolutional Neural Network and Tensor Deep Stacking Network," IEEE Access, volume 7, pp. 7717-7727, 2019.

[2] Junbin Gao, Xue Li, Lin Wu and Yang Wang, "Deep Attention-Based Spatially Recursive Networks for Fine-Grained Visual Recognition" IEEE Access, volume 47, pp. 1-12, 2018.

[3] Haobijam Basanta and Yo-Ping Huang, "Bird Image Retrieval and Recognition Using a Deep Learning Platform," IEEE Access, volume. 7, pp. 66980–66989, 2019.

[4] Tuomas Virtanen, Annamaria Mesaros, Toni Heittola, Aleksandr Diment, Benjamin Elizalde, Emmanuel Vincent, Bhiksha Raj, Fellow, "Sound Event Detection in the DCASE 2017 Challenge" IEEE Access, volume 27, pp. 992-1006, 2019.

[5] Paul Roe and Philip Eichinski, "Clustering and visualization of long-duration audio recordings for rapid exploration avian surveys," presented at the IEEE 13th International Conference on eScience, Brisbane, Australia, 2017.

[6] Ching-Chien Chuang, Chang-Hsing Lee, and Chin-Chuan Han, "Automatic Classification of Bird Species From Their Sounds Using Two-Dimensional Cepstral Coefficients," IEEE Trans. Audio, Speech, Language Process., volume. 16, no. 8, pp.1541–1550, Nov. 2008.

[7] https://www.kaggle.com/competitions/birdclef-2022/overview

[8] https://www.kaggle.com/c/birdsong-recognition

## BIOGRAPHIES

**DHIRAJ PATIL** senior student(Computer Engg.) atSavitribai Phule Pune University. His research interests include environmental sound classification, data analytics, and deep learning.

**RUTWI BODHE** senior student(Computer Engg.) at Savitribai Phule Pune University. Her research interests include data mining and pattern recognition.

**RUPALI PAWER** senior student(Computer Engg.) at Savitribai Phule Pune University. Her research interests include signal processing and image segmentation.

**TANISHA DOSHI** senior student(Computer Engg.) at Savitribai Phule Pune University. Her research interests include computer vision and artificial intelligence .