International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Automated Bird Species Identification via Audio Signal Processing and Machine Learning

Karthik M G
student, Dept of CSE,
Sea College of Engineering & Technology

Pramod
student, Dept of CSE,
Sea College of Engineering & Technology

Tharun G
student, Dept of CSE,
Sea College of Engineering & Technology

Venugopala G
student, Dept of CSE,
Sea College of Engineering & Technology

Dr Rajagopal K
Professor Dept of CSE
SEA College of Engineering & Technology

Dr Krishna Kumar P R
Professor Dept of CSE
SEA College of Engineering & Technology

Abstract

The identification of bird species through their vocalizations has significant implications for biodiversity monitoring, ecological research, and conservation efforts. Traditional methods, which rely on expert knowledge and manual analysis, are time-consuming and often impractical at scale. This study presents an automated system for bird species identification using audio signal processing techniques combined with machine learning algorithms. The proposed approach involves preprocessing raw bird sound recordings to extract meaningful audio features such as Mel-Frequency Cepstral Coefficients (MFCCs), chroma features, and spectral contrast. These features are then used to train and evaluate various machine learning classifiers, including Support Vector Machines (SVM), Random Forests, and Convolutional Neural Networks (CNNs). Experimental results demonstrate that the system can accurately classify multiple bird species across a diverse dataset, achieving promising levels of accuracy and generalizability. This research highlights the potential of integrating signal processing and machine learning to develop scalable, real-time solutions for avian acoustic monitoring in natural environments

***Index Terms*—Classifier, Birds, Entropy, Sampling, Visualization**

I. Introduction

There are approximately 1500 birds species in the India and even more when you consider the subspecies. Birds form a essential part of ecosystem and its important to identify them so as efforts can be made to create ecological hot spots for the endangered ones. From the start of the 20th century witnessed widespread decline in the population of birds taking and few of them were at the verge of extinction, hence efforts were made by various governments and well known personalities in the form of campaigns, advertisements to save birds which made the issue of bird conservation to reach masses especially the youth.People are more curious about bird watching and conservation movements than they were before. Human ear is able to perceive and discern most of the sounds from their source. The chirping of birds is one such noise that occurs in our natural environment that is not easy to differentiate just by listening to it and therefore a software-based solution that helps to recognize every bird species and allows them to learn more about them.Birds in general have 2 distinct types of sounds:Call and Song. Even after knowing the bird one might not be able to differentiate among them and hence a sub-classification based on call and song is too performed using relevant machine learning techniques.

LITERATURE SURVEY

We conducted a literature survey to find out what all work had been done in this field and what tools are needed to go ahead with the system. The survey gave us insights about the features required to classify the bird sounds and which models are needed for training to move further. The process of extraction of different features and selection of Mel Frequency Cepstral Coefficients to give input to the Support Vector Machine (SVM) as it gives a decent accuracy of about 89.64% [1]. Extreme learning machines (ELM) are used to model emotional perception of audio and video features [2]. ELM is used as an alternative to single layer feed forward networks for fast and accurate learning. A good open source tool for speech processing and music information retrieval is openSmile [3]. This software provides easy extraction of audio features. The selection of features needs to be done properly as some of the features are such that including them reduces the accuracy of the model [4]. The preprocessing that is done on the

1) **The application of modality-specific extreme learning machines to identify emotions in the wild[2]** This paper proposes extreme learning machines (ELM) for modeling, under unregulated conditions, audio and video functionality for emotion recognition. For single-layer feed forward networks, the ELM model is a fast and accurate learning alternative.
2) **OpenSMILE Versatile and Quick Open Source Audio Extractor[3]** OpenSMILE is a voice processing and music information retrieval tool that allows researchers in each domain to benefit from features from the other domain. This can easily be used to extract audio file features.
3) **Automatic large-scale classification of bird sounds is greatly improved by unsupervised learning of features[4]** This paper discusses and determines what characteristics should be taken into account. It also indicates whether or not noise reduction should be done as a pre-processing stage, audio windowing to make a classification decision.

4) **Species-specific audio detection: a study of three algorithms based on models that use random forests[5]** They have presented a web-based cloud-hosted system that provides a simple way to manage large quantities of recordings with a general-purpose method to detect the presence of bird species in recordings. They have also pursued approaches like multi-label learning.
5) **Classification of large-scale bird sound using Convolutional Neural Networks[6]** This paper addressed the attempt to use convolutional neural networks to classify bird speech. They developed audio spectrograms and provided them as input to the neural networks.
6) **Based on their sound patterns, environmental prediction and classification of birds[7]** They worked here to map on their sounds the characteristics of birds such as scale, location, species and call forms.It also includes development of hardware and software systems to monitor the bird species.

II. PROPOSED DESIGN

The System we have proposed to seek the solution of the Problem Statement Stated above is:

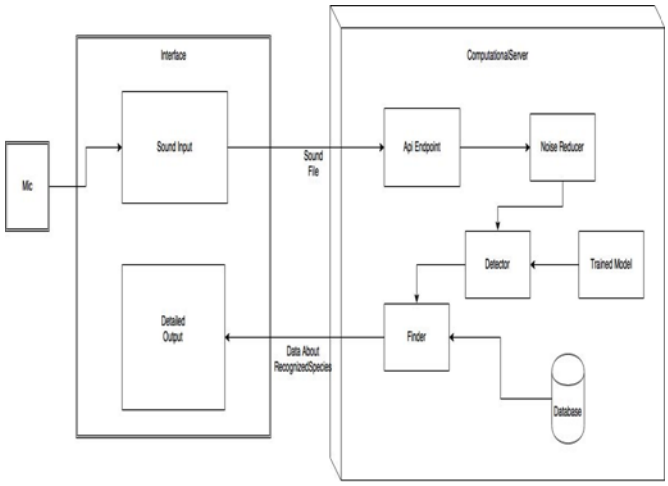

Fig. 1. Proposed System

The Architecture of the project would include a Client Server type of architecture.At client side there is Android device which can record sound. The sound file will be received from the interface device to the computational server by an API Endpoint. The sound file will then Undergo Some preprocessing i.e.it will under go re-sampling.The re-sampled audio file will be classified against the trained model,the output along with the probabilities will be given to Android Client device. The main part of the system is training model and detection. This includes Template Computation Model undergoes feature extraction including training recording together.Classification algorithm is applied and given to the detector. The detector takes cleaned recording of the input and extracts its feature and compare it with the the input from model trainer then decides results.

III. EXPERIMENTATION

The general steps that we have taken in order to create the system.

A. Data Collection and Refinement

- Bird Audio Data was collected from Macaulay Library,eBird and xeno-canto.
- After the data acquisition,the data was manually refined by using open-source software Audacity.Refinement was done in the form of noise cancellation and creating relevant clips of 2-3seconds
- 100 such clips of each bird class were created and used for further Feature Extraction process.

.

B. Feature Selection and Extraction

1) Feature Selection
   - The list of extracted features in the project is shown in table I

TABLE I
TABLE TO TEST CAPTIONS AND LABELS

Feature ID	Feature Name
1	Zero Crossing Rate
2	Energy
3	Entropy of Energy
4	Spectral Centroid
5	Spectral Spread
6	Spectral Entropy
7	Spectral Flux
8	Spectral Roll off
9-21	MFCCs
22-33	Chroma Vector
34	Chroma Deviation

   - The time-domain features (features 13) are derived directly from the samples of the raw signal. The frequency domain features (features 4-34, apart from the MFCCs) are based on the magnitude of the Discrete Fourier Transform (DFT). Eventually, after applying the InverseDFT, the cepstral domain (e.g. used by the MFCCs) results.
   - For the training of machine learning model we have selected all 34 features as it results to highest accuracy for the model.It is clear from the below figure.
2) Feature Extraction
   - On a short-term basis, all 34 functions can be extracted: the audio signal is first divided into short-term windows (frames) and all 34 functions are calculated for each frame.[11]
   - Out of the widely accepted range(20ms-100ms) we have selected window size of 20ms for short term statistics.
   - A mid term statistics are calculated by taking mean and standard deviation of short term features in window size of200ms.

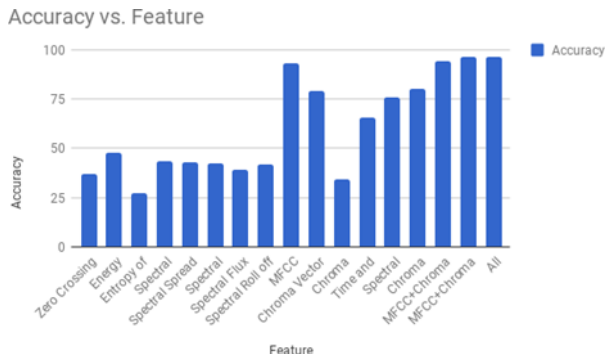

Fig. 2. Graph of Features Selected vs Accuracy

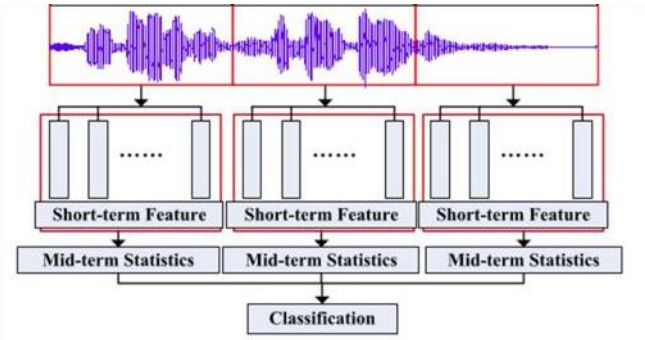

Fig. 3. Feature Extraction Process

- The result of short term feature extraction is excel sheet with n rows and 34 columns.
- The result of mid term feature extraction is excel sheet with n rows and 68 columns.

C. Data Visualization

- Since the Extracted features had high dimensionality (n rows * 64 columns),it was not possible to visualize the dataset.
- Hence Dimensionality Reduction of the current dataset to help us to visualize the database in the 2-Dimension space.
- Principal Component Analyses (PCA) was used for dimension reduction Process.
- Below is the output(plot) when we reduced the dataset into 2 dimensions and plotted onto graph using python library.
- The Plotted Graph Fig. 4 shows that the data is not Linear as2 classes do overlap and is non-linear. Hence was taken into account while selecting the algorithms for model training in the next step.

D. Model Training

After Feature Selection and Extraction process,Model training is the main task.

1) Algorithm Selection

- After Data visualization,KNN,random forest and SVM algorithms for model training were shortlisted based on the literature survey in which we this 3 algorithms were best working algorithm for such non-linear dataset.
- The model was then made by using the above algorithms and the best algorithm was chosen.
- The dataset was divided randomly into training and testing datasets in the ratio of 90:10 to test the accuracy of the model classifier.
- The Algorithm which gave the best accuracy was SVM(Radial Basis) And hence was further selected for model training.

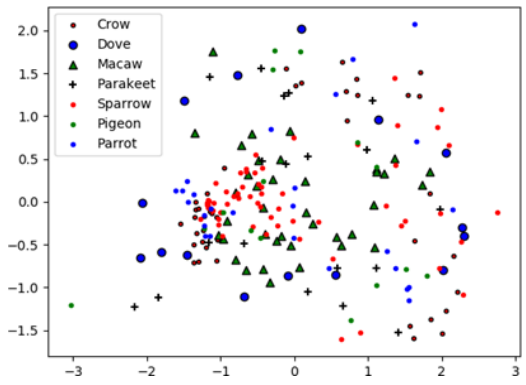

Fig. 4. Data Visualization

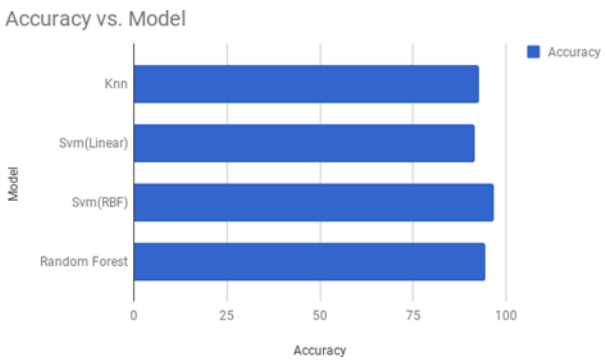

Fig. 5. Model vs Accuracy

2) Parameter Tuning

- To Avoid over-fitting of the data in the SVM, parameter 'C'was kept varied for few standard values and for each value of C,Accuracy and F1 score were plotted.
- The C parameter operates against the consistency of the decision surface by misclassifying training instances. A low C makes the decision surface smooth, while a high C is designed to correctly classify all training examples by allowing the model to select more samples as support vectors.
- The selected C value was 10 as can be seen from the Fig. 6.

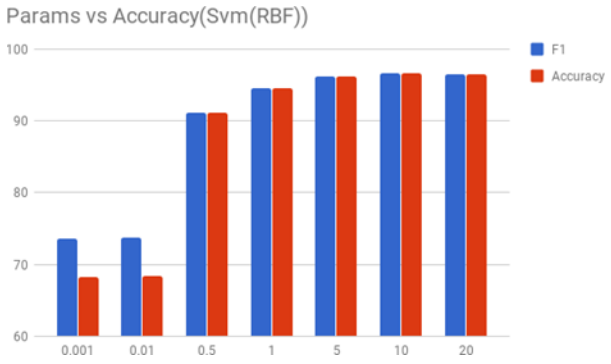

Fig. 6. Model vs Accuracy

E. Classification

Given below are the sequence of steps which are executed on audio input to classify an unknown audio file.

- First we have divided the input audio file into audio file of3 seconds with 1second of overlapping windows.
- For each audio file we have predicted the bird with the trained model,for each audio it gives the probabilistic results for each bird.
- By taking the average of each small probabilistic result we have predicted the actual bird from the whole audio file.

IV. RESULTS

The Support Vector machine with radial basis function produced the accuracy of 96.7% during classification of the main classes and about 96% to 99% while doing the sub classification based on call or song of the classified bird species.Confusion matrix for the main classification is given below. The model was able to accurately identify the main as

Fig. 7. confusion matrix for primary classification using SVM(radial basis)

well as the subclass of the audio file given to it as input through the application. The time taken by the server to process the file and return the response was around 10ms-20ms for a audio of 5s-7s

VI. CONCLUSION

This study demonstrates the effectiveness of combining audio signal processing techniques with machine learning algorithms for the automated identification of bird species based on their vocalizations. By extracting relevant acoustic features and leveraging models such as Support Vector Machines, Random Forests, and Convolutional Neural Networks, the system achieved promising classification accuracy across diverse bird calls. The results validate the potential of such an approach for ecological monitoring, biodiversity assessment, and conservation efforts, particularly in environments where manual observation is challenging or impractical. Furthermore, the scalability and adaptability of the system offer opportunities for real-time deployment in field settings using embedded or mobile devices. Future work may focus on improving robustness in noisy environments, expanding the dataset to include more species, and integrating deep learning techniques for end- to-end audio classification.

REFERENCES

[1] Rai, Pallavi, VikramGolchha, AishwaryaSrivastava, GarimaVyas, and Sourav Mishra. ”An automatic classification of bird species using audio feature extraction and support vector machines.” In Inventive Computation Technologies (ICICT), International Conference on, vol. 1, pp. 1-5. IEEE,2016.

[2] Kaya, Heysem, and Albert Ali Salah. ”Combining modality-specific extreme learning machines for emotion recognition in the wild.” Journal on Multimodal User Interfaces 10,no.2(2016):139-149.

[3] Eyben, Florian, Martin Wllmer, and BjrnSchuller. ”Opensmile: the munich versatile and fast open-source audio feature extractor.” In Proceedings of the 18th ACM international conference on Multimedia, pp. 1459-1462. ACM,2010.

[4] Stowell, Dan, and Mark D. Plumbley. ”Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning.” PeerJ2 (2014):e488.

[5] Corrada Bravo CJ, A´ lvarez Berr´ıos R, Aide TM. 2017. Species-specific audio detection: a comparison of three template-based detection algorithms using random forests. PeerJ Computer Science 3:e113 https://doi.org/10.7717/peerj-cs.113

[6] Naranchimeg BOLD, Chao ZHANG, Takuya AKASHI, Cross-Domain Deep Feature Combination for Bird Species Classification with Audio-Visual Data, IEICE Transactions on Information and Systems, 2019, Volume E102.D, Issue 10, Pages 2033-2042, Released October 01, 2019, Online ISSN 1745-1361, Print ISSN 0916-8532, https://doi.org/10.1587/transinf.2018EDP7383

[7] M. A. Raghuram,Nikhil R. Chavan,Ravikiran Belur,Shashidhar G. Koolagudi,Bird classification based on their sound patterns, International Journal of Speech Technology, Issue 4/2016, DOI:https://doi.org/10.1007/s10772-016-9372-2

[8] Suykens, Johan AK, and JoosVandewalle. ”Least squares support vector machine classifiers.” Neural processing letters 9, no. 3 (1999): 293-300.

[9] Han, Kun, Dong Yu, and Ivan Tashev. ”Speech emotion recognition using deep neural network and extreme learning machine.” In Fifteenth Annual Conference of the International Speech Communication Association.2014.

[10] Yang, XK., He, L., Qu, D. et al. J AUDIO SPEECH MUSIC PROC. (2016) 2016: 9. https://doi.org/10.1186/s13636-016-0086-9

[11] Yang, X.K., He, L., Qu, D., Zhang, W.Q. and Johnson, M.T., 2016. Semi-supervised feature selection for audio classification based on constraint compensated Laplacian score. EURASIP Journal on Audio, Speech,andMusicProcessing,2016(1),p.9.

[12] T Hirvonen, Speech/Music Classification of Short Audio Segments (Proc. 2014 IEEE International Symposium on Multimedia, Taichung, 2014). Dec. 10–12

[13] Y Vaizman, B McFee, G Lanckriet, Codebook-based audio feature representation for music information retrieval. IEEE/ACM Transactions

on Audio, Speech, and Language Processing 22(10), 1483–1493 (2014)
[14] P Mahana, G Singh, Comparative analysis of machine learning algo- rithms for audio signals classification. International Journal of Computer Science and Network Security 15(6), 49–55 (2015)
[15] H Yan, J Yang, Locality preserving score for joint feature weights learning. Neural Netw. 69, 126–134 (2015)