Project Report (Stage I)

on

An Automatic Classification Of Bird Species Using Audio Feature Extraction

submitted in partial fulfillment of the requirement
for the award of the Degree of

Bachelor of Engineering
in
Computer Engineering

by

Siddhey Sankhe (2014130047)
Zain Ahmed Sayed (2014130048)
Harsh Vora (2014130062)

under the guidance of

Prof. Anand A. Godbole

Department of Computer Engineering
Bharatiya Vidya Bhavan's
Sardar Patel Institute of Technology
Munshi Nagar, Andheri-West, Mumbai-400058
University of Mumbai
November 2017

Approval Certificate

This is to certify that the Project entitled "**An Automatic Classificaion Of Bird Species Using Audio Feature Extraction**" by Siddhey Sankhe, Zain Ahmed Sayed and Harsh Vora is approved for the partial fulfillment of Stage I for Final Year Project towards obtaining the Bachelor of Engineering Degree in Computer Engineering from University of Mumbai.

Project Supervisor	**External Domain Expert**
(signature)	**(signature)**
Name:	**Name:**
Date:	**Date:**

Head of Department	**Principal**

Contents

List of Figures

1 Abstract

We are developing a Android application which is cloud supported that allow users to record bird sounds and identify them. The system also provides tools to convert these annotations into datasets that can be used to train a computer to detect the presence or absence of a species. Dataset recorded are preprocessed to remove unwanted noise and divide useful sounds in frames so that it can act as an input to classifier.The system will be tested through different algorithms and algorithm that will give best results will be chosen for implementation.System may use audio features like MFCC,Mel-Spectra etc.

2 Introduction

2.1 Motivation

There are approximately 1500 birds species in the India and even more when you consider the subspecies. Birds form a major part of ecosystem and its important to identify them so as efforts can be made to create ecological hot spots for the endangered ones. Human ear is capable of perceiving most of the sounds and differentiating them from their source. The chirping of birds is one such audio existing in our natural surroundings which is not easy to differentiate just by listening to it and therefore a software based solution which will help to identify any species of birds and help them know more about them.

2.2 Scope

The system developed will help bird species to be identified by taking input only of the sound of that bird. The system will basically focus on the bird species and cater to identify the bird species only and scope limited to Indian Bird species only.In this project , we will focus on identification of maximum four bird species.Once the prototype is developed it can be extended to identify more species.

2.3 Assumptions

The following assumptions are to be taken into consideration:-

- The user will only send the audio clip which contain bird sound recordings to enable the learning model work properly.
- The user has an android phone with a microphone, minimum specifications and a network connection to send the data to the computational server.

3 Literature Survey

3.1 Papers Read To Start With

1. **An Automatic classification of bird species using audio feature extraction and support vector machines [1]**

   - Automatic identification of bird species based on the chirping sounds of birds was experimented using feature extraction method and classification based on support vector machines (SVMs).
   - In [5] the researchers have implemented identification system for songs of six species common to Manitoba, Canada were tested these songs were defined in terms of their temporal and as well as spectral parameters.
   - This paper proposes a system for labelling of bird species by extracting features from bird sound using MFCC and describes the SVM based algorithm used for classification.
   - In this paper, researchers have proposed a Mel Frequency Cepstral Coefficients(MFCCs) for feature extraction which is the most used features to describe the spectrum of an audio recording in very compact yet informative manner.
   - The paper proposes classification using Support Vector Machine(SVM) , SVMs are a set of machine learning methods used for classification. They are a type of general linear classification and help to minimize classification error. Support vector machine is a kernel-based methodology used as a powerful tool in automatic audio classification and recognition systems.
   - Their accuracy and superior generalization properties offer advantage over other types of classifiers.
   - The proposed technique was used for the recognition of the following four classes of bird species which were amongst the most commonly found birds in India. The technique used mel cepstral coefficients as extracted features for the classification of test samples. Classification was done using multi SVM technique.The overall accuracy being 64% within a maximum accuracy of 89.74%.

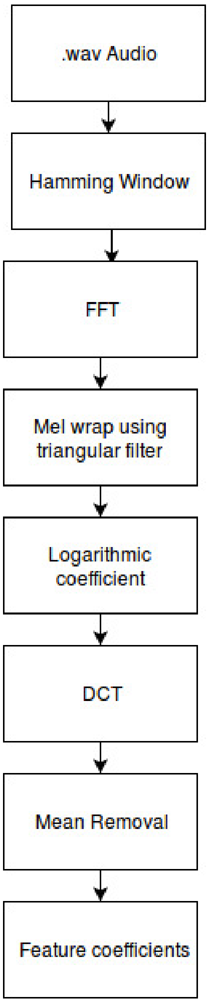


Figure 1: Feature extraction method proposed by the paper .

2. **Combining modality-specific extreme learning machines for emotion recognition in the wild** [2]

    - This paper proposes extreme learning machines (ELM) for modeling audio and video features for emotion recognition under uncontrolled conditions.
    - The ELM paradigm is a fast and accurate learning alternative for single layer Feed forward networks.
    - The extreme learning machine (ELM) classifier was first introduced in [6] as a fast alternative training method for single layer Feed forward networks (SLFNs).
    - The basic ELM paradigm has matured over the years to provide a unified framework for regression and classification, related to generalized SLFN class including least square SVM (LSSVM) [7] [8].

- Despite the speed and accuracy of ELMs,they were only recently employed in affective computing exhibiting outstanding performance with typically under sampled, high dimensional datasets [9] [?].

3. **openSMILE The Munich Versatile and Fast Open-Source Audio Feature Extractor [3]**

   - A tool for speech processing and Music Information Retrieval, enabling researchers in either domain to benefit from features from the other domain. [3]
   - It introduced openSMILE, an efficient, on-line (and also batch scriptable), open-source, cross platform, and extensible feature extractor implemented in C++. A well structured API and example components make integration of new feature extraction and I/O components easy. openSMILE is compatible with research tool-kits, such as HTK, WEKA, and LibSVM by supporting their data-formats.
   - Although openSMILE is very new, it is already successfully used by researchers around the world. The openEAR project builds on openSMILE features for doing emotion recognition. openSMILE was the official feature extractor for the INTERSPEECH 2009 Emotion Challenge and the ongoing INTERPSEECH 2010 Paralinguistic Challenge. It has also been used for problems as exotic as classification of speaker height from voice characteristics. Development of openSMILE is still active and even more features such as TEAGER energy, TOBI pitch descriptors, and psychoacoustic measures such as Sharpness and Roughness are considered for integration.
   - openSMILE will soon support MPEG-7 LLD XML output. In the near future we aim at linking to openCV10, to be able to fuse visual and acoustic features. Due to openSMILEs modular architecture and the public source code, rapid addition of new and diverse features by the community is encouraged.
   - Future work will focus on improved multi threading support and cooperation with related projects to ensure coverage of a broad variety of typically employed features in one piece of fast, lightweight, flexible open-source software.

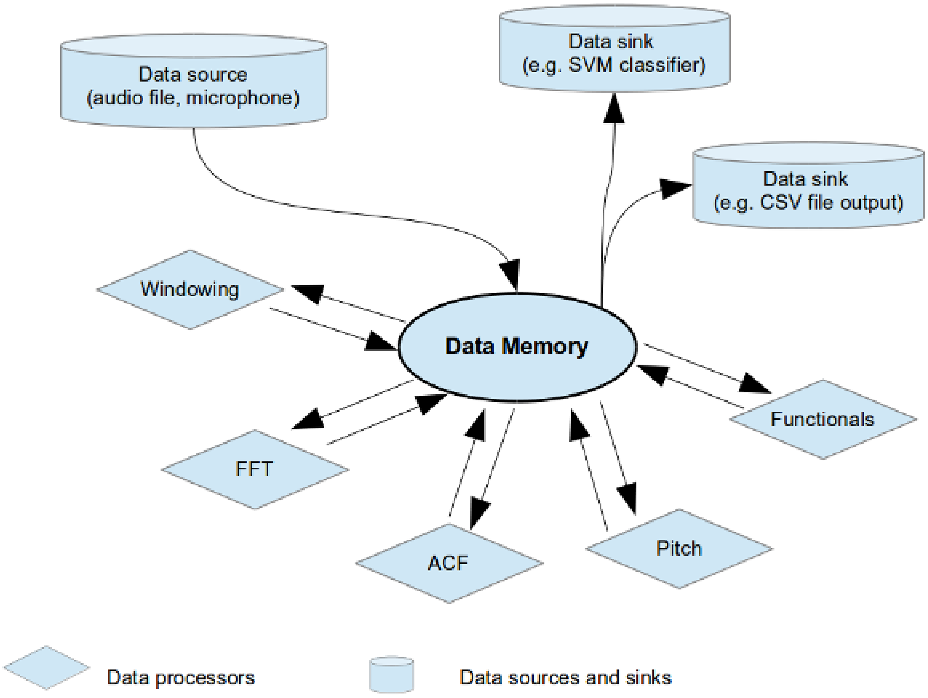


Figure 2: openSMILE Architecture explained by the paper.

3.2 Deciding Algorithms for Classification and Feature Selection

1. **Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning** [4]

   - Automatic species classification of birds from their sound is a computational tool of increasing importance in ecology, conservation monitoring and vocal communication studies
   - The Paper evaluates and finds out: [1] Choice of features (MFCCs, Mel spectra, or learned features) and their summarisation over time (mean and standard deviation, maximum, or modulation coefficients); Whether or not to apply noise reduction to audio spectra as a pre-processing step; Decision windowing: whether to treat the full-length audio as a single unit for training/testing purposes, or whether to divide it into shorter-duration windows (1, 5 or 60 s); How to produce an overall decision when using decision windowing (via the mean or the maximum of the probabilities); Classifier configuration: the same random forest classifier tested in single-label, multilabel or binary-relevance setting.
   - Paper proposes feature learning method with **Spherical k-means**.
   - Spherical k-means is related to the simple and well-known k-means clustering algorithm, except that instead of searching for cluster centroids which minimise the Euclidean distance to the data points, we search for unit vectors (directions) to min-

imise their angular distance from the data points. This is achieved by modifying the iterative update procedure for the k-means algorithm: for an input data point, rather than finding the nearest centroid by Euclidean distance and then moving the centroid towards that data point, the nearest centroid is found by cosine distance,

$$\text{cosine distance} = 1 - \cos(\theta) = 1 - \frac{A \cdot B}{\|A\| \|B\|},$$

where A and B are vectors to be compared, is the angle between them, and is the Euclidean vector normal.

- The basis vectors are used in the paper to represent input data in the new given space which is calculated by the dot product,

$$x^{'}(n, j) = \sum_{i=1}^{M} b_j(i) x(n, i),$$

where x represents the input data indexed by time frame n and feature index i (with M the number of input features, e.g., the number of spectral bins), bj is one of the learnt basis vectors (indexed by j[1, k]), and x is the new feature representation.

- They found out that standard implementation of k-means clustering requires an iterative batch process which considers all data points in every step which is not feasible for high data volumes

2. **Species-specific audio detection: a comparison of three template-based detection algorithms using random forests**

   - They have shown how the algorithm used in the ARBIMON II web-based cloud-hosted system was selected. They compared the performance in terms of the ability to detect and the efficiency in terms of time execution of three variants of a template-based detection algorithm. The result was a method that uses the power of a widely use method to determine the similarity between two images, but to accelerate the detection process, the analysis was only done in regions where there was a strong-match determined by the OpenCVs matchTemplate procedure. The results show that this method performed better both in terms of ability to detect as well as in terms of execution time.

   - A fast and accurate general-purpose algorithm for detecting presence or absence of a species complements the other tools of the ARBIMON system, such as options for creating playlists based on many different parameters including user-created tags (see Table 5). For example, the system currently has 1,749,551 1-minute recordings uploaded by 453 users and 659 species specific models have been created and run over 3,780,552 min of recordings of which 723,054 are distinct recordings. While this research was a proof of concept, they provided the tools and encouraged users to increase the size of the training data set as which would improve the performance of the algorithm.

- In addition, they pursued other approaches, such as multi-label learning.
- Here they 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 their presence in recordings.

3. **Large-Scale Bird Sound Classification using Convolutional Neural Networks**

    - They provided insights into their attempt of large-scale bird sound classification using various convolutional neural networks. After they conducted numerous experiments to identify the best techniques of dataset augmentation, training methods and network architectures, their best submission to the 2017 BirdCLEF challenge achieved a score of 0,605 MAP ranking second of all submissions. The results show that there is still a lot of room for improvements especially for the soundscape domain, which likely is the most important real-world application.
    - Additionally, they provided a GitHub repository for the free use of their code base and with that, hope to offer a baseline for future BirdCLEF tasks.
    - Their work flow consisted of four main steps. First, they extracted spectrograms from all audio recordings. Secondly, they extended their training set through extensive dataset augmentation. Next, they tried to find the best CNN architecture with respect to number of classes, sample count and data diversity. Finally, they trained their models using consumer hardware and Open Source toolkits and frameworks.

4. **Environmental Prediction and Bird classification based on their sound patterns**

    - The proposed methodology may be used to conduct survey of birds. The proposed methods may be used to automatically classify birds using different audio processing and machine learning techniques on the basis of their chirping patterns.
    - An effort has been made in this work to map characteristics of birds such as size, habitat, species and types of call, on to their sounds.
    - This paper also supports part of a broader project that includes development of software and hardware systems to monitor the bird species that appear in different geographical locations which helps ornithologists to monitor environmental conditions with respect to specific bird species.
    - The proposed classification methods are:

  - **Naive Bayes** The Naive Bayes classifier is one among the many different classifiers that are based on the Bayes Theorem and is useful particularly when the input feature space is of high dimensionality.
  - **Support vector machines** Support Vector Machines is a popular classification technique which tries to determine a large-margin hyperplane that can act as decision boundary. In the experiments, we make use of a polynomial kernel of degree two,which performs an implicit mapping from the input feature vector to high-dimension feature space for identifying a clearer margin.
  - **Random forest** An ensemble method of classification, the Random Forest classifier constructs multiple decision trees and returns the mode of the classes (for classification tasks) and mean of prediction (for regression tasks). It makes use of the tree bagging technique with the selection of random subset of the feature space for building a decision tree. This results in a large forest composed of shallow trees because of which the individual trees are less likely to over-fit for large training data-sets. In the experiments, we set the number of decision trees to ten and use all of the features for each tree.
  - **Neural networks** Artificial neural networks (ANN) are highly interconnected networks of simple processing elements or units (neurons). These neurons are organized into layers, namely input, hidden and output layers which converts an input vector into output.

4 Research Gap Identified

The literature survey discusses several real world problems encountered in the Classification of bird species taking the audio input.Study identified the following gaps which can be further reaserched:

- One bird can have different types of calls and the papers have not covered their classification.

- The papers focus on the International birds. There is no paper to classify the local birds.

- Though Extreme Learning Machine (ELM) provides speed and accuracy there is no standardised way to employ ELM with typically under sampled ,high dimensional datasets.

5 Problem statement

There are a number of bird species in our ecosystem.Every bird has a distinguished sound (chirping).And there is a difference between the different types of calls that a bird makes.To develop and implement a system to identify the distinguished bird sounds from an audio and effectively classify them into various birds species and their current state.

6 Objectives

The Objectives defined are as follows:-

- The main task is to design a system that, given a short audio recording, returns a decision for the presence/absence of bird sound (bird sound of any kind). There would be a weighted or probability based output which would give us information of the species available in the surroundings.

- Distinguish the different sounds produced by the same species of birds and their classification based on the scenarios in the surrounding.

- To identify accuracy of different algorithms such as KNN, Random Forest, Multilayer Perceptron, Bayes in classification of birds species.

- The application could also work as a crowdsourcing application to add the sound of the new species that are not found by the application.

7 Outcomes

What will our project serve in the end:

- A comprehensive model which can distinguish the bird species by their sound and also can categorize them.

- A system which can analyze the sound patterns of the birds and classify them as different types of sound produced by the particular species of bird.

- An Android app for recognizing various bird sounds. The android application is a good option because this would increase the portability and allowing users to easily record sounds and use the application when mobile.

8 Proposed System

8.1 Architecture

The Architecture of the project would include a Client Server type of architecture where there would be a Mic that would be connected to an interface which would allow the recording and sending of the sound to the Computational Server. 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 undergo Noise reduction. After that it will be fed to the Detector that will analyze the sound and get to know which bird species it is. The Detector will Have inputs from the Trained Model which will help it decide. The decision made by the Detector will be given to the Finder which will have access to the bird database for th details of the birds. It will fetch the data from the database and send it to the Interface to display the result on the device.

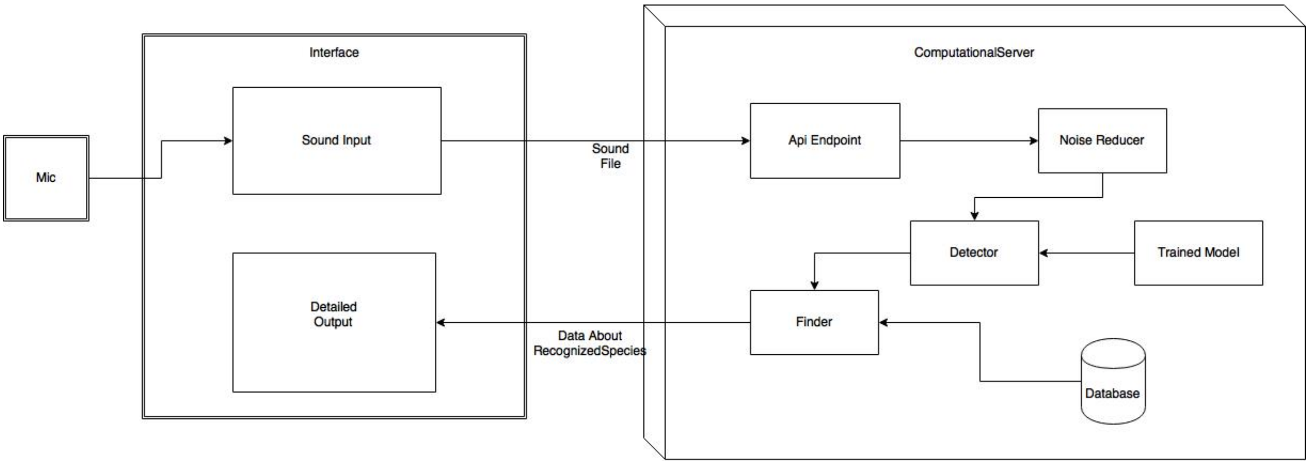


Figure 3: Architecture Of the Project.

8.2 Training Model

The main part of the entire system is the Training model and Detection. This is the heart of the system. This includes Template Computation, Model Training and Detection. The template computation block takes small ROI's and creates template to give it to the Recognition Function of the Model Training Block. This then undergoes feature Extraction and including training recordings together we apply the classification algorithm and give it to the detector. The detector takes the cleaned recording of the input and extracts its features and compares it with the input from the model trainer and then decides on the result. This result is the identity of the bird whose sound was given as input

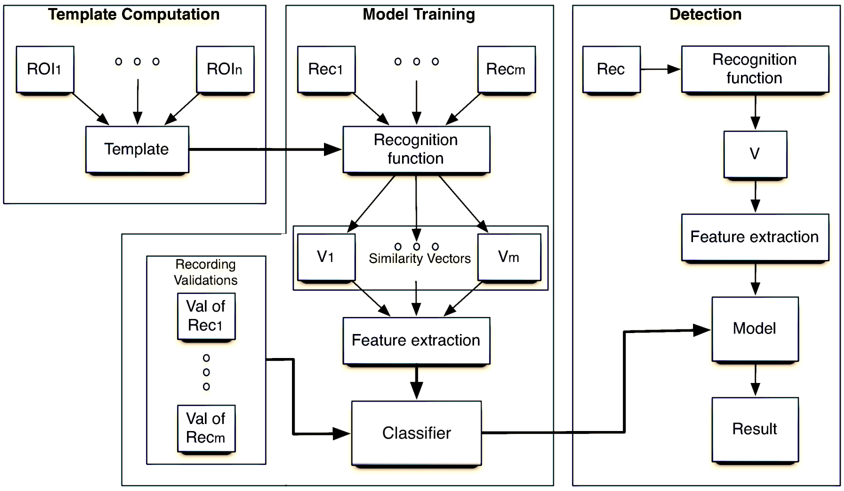


Figure 4: Training Model.

8.3 Making Predictions

Once audio samples have been recorded, we store them in a server. It is here that each recording undergoes preprocessing to get the audio in a suitable format for predictions. The process is as follows:

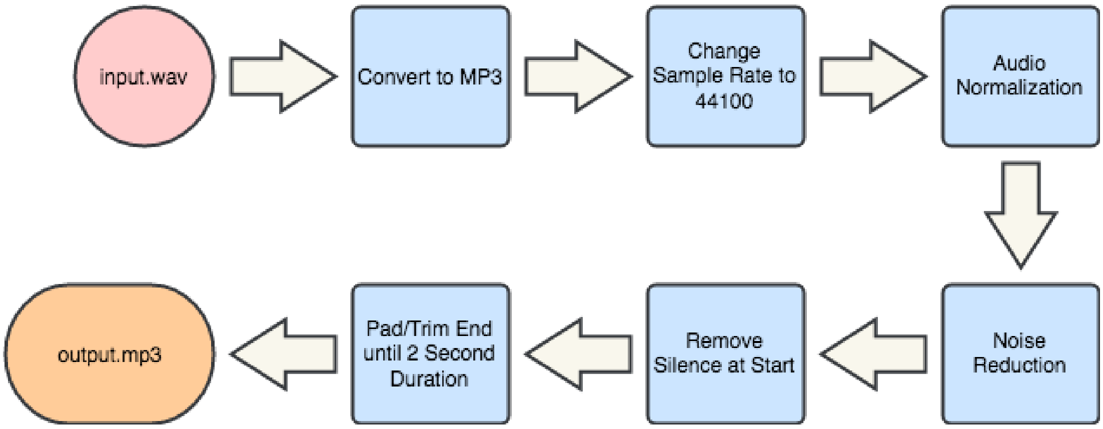


Figure 5: Preprocessing that each audio sample undergoes before it will be passed to predictive model

Once the recording has been preprocessed, it is fed into a program that runs it against the chosen model.

8.4 Functional Requirements

The Functional requirements of our project are as follows:-

- The system will properly reduce the noise for the input provided by recording.
- The model generated should be able to extract appropriate features from audio clip.
- The system will be able to recognize the species of bird with the help of trained model and the features extracted from user input.
- It can identify the different species present in a particular area.

8.5 Non-Functional Requirements

The Non-Functional requirements of our project are as follows:-

- The response time of the detection should be as low as possible.
- When sometimes the network is not available the application should be able to handle the situation by storing the audio and then send it to the server when it regains connection.
- The system should be equally accurate for all the species of bird chosen.
- The system should be portable as system should be used on portable devices like mobile to detect bird species.

8.6 Planning

We have decided upon the time line to complete the project as below:-

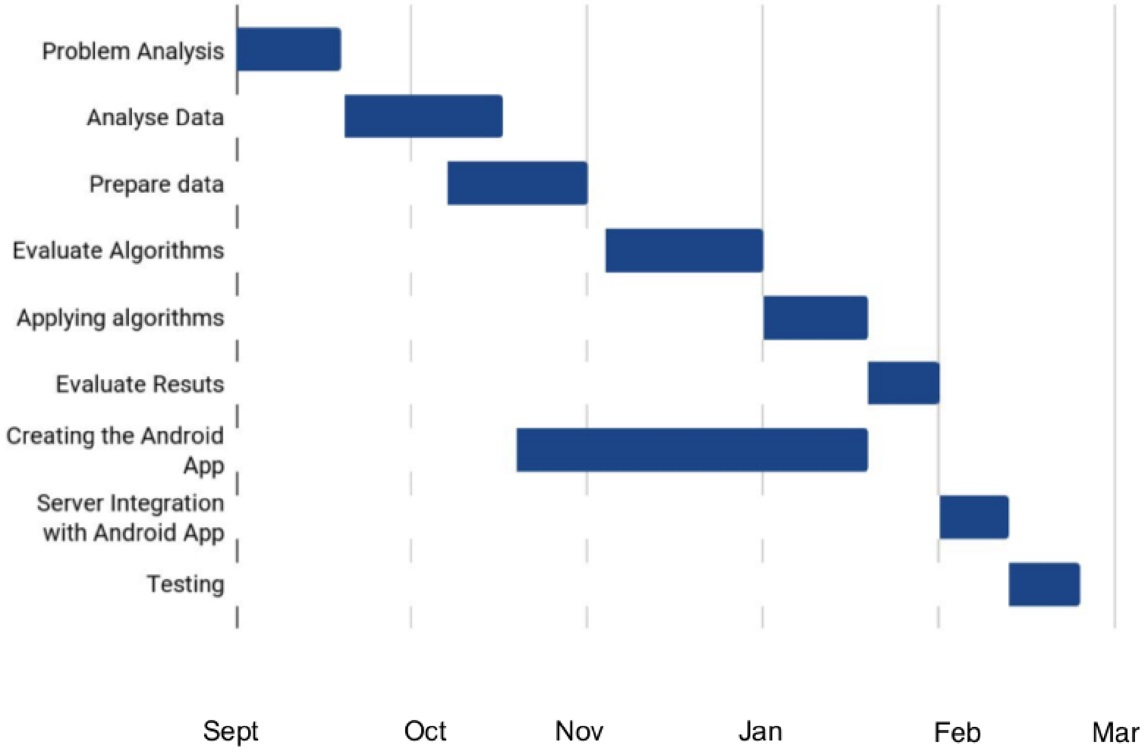


Figure 6: Timeline Chart.

8.7 Use Case Diagram

The Use case diagram is proposed as:

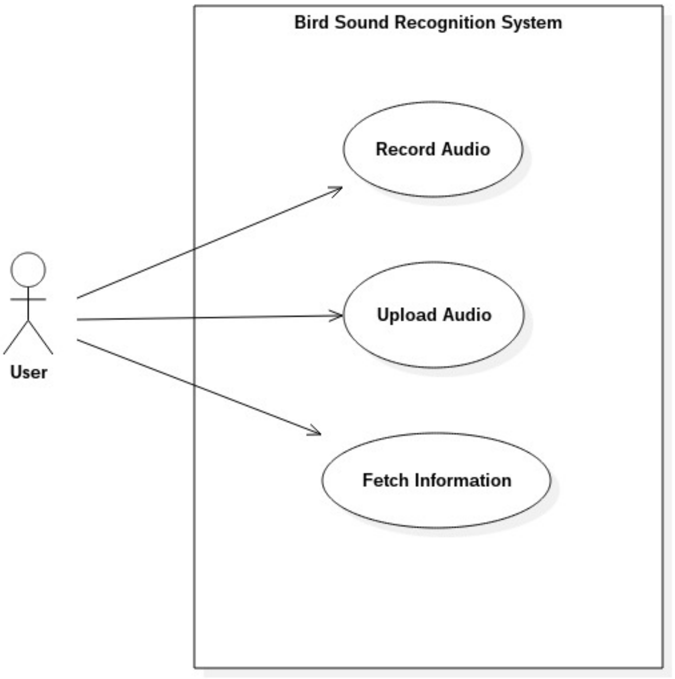


Figure 7: Usecase Diagram.

8.8 Sequence Diagram

Proposed Sequence diagram for our system is as follows:

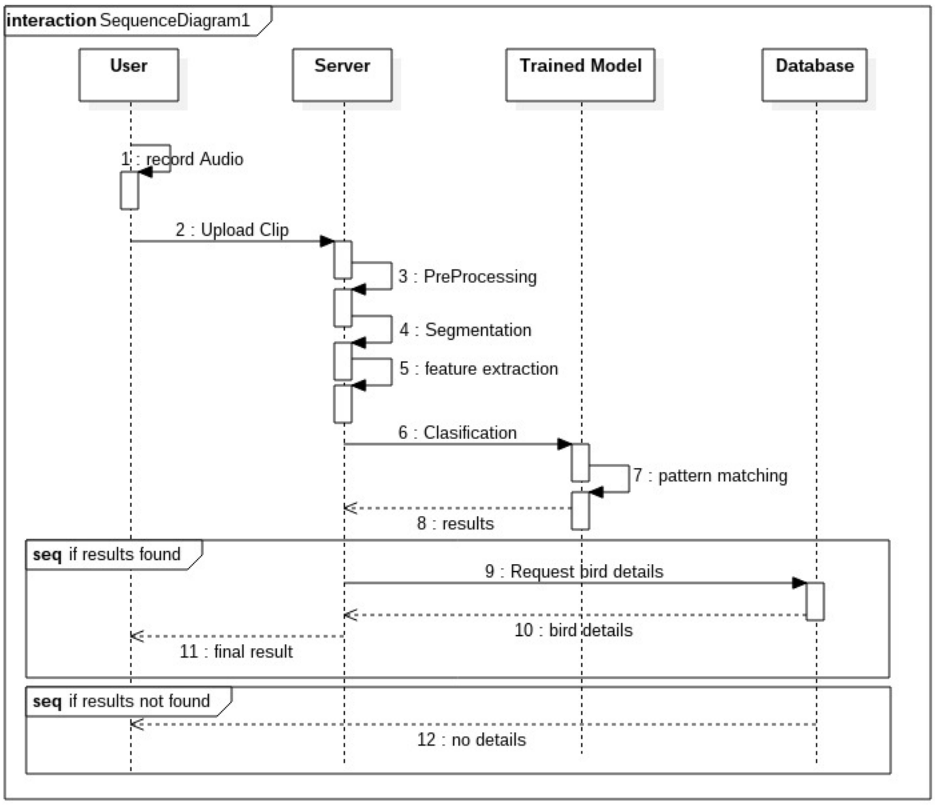


Figure 8: Sequence Diagram.

8.9 Activity Diagram

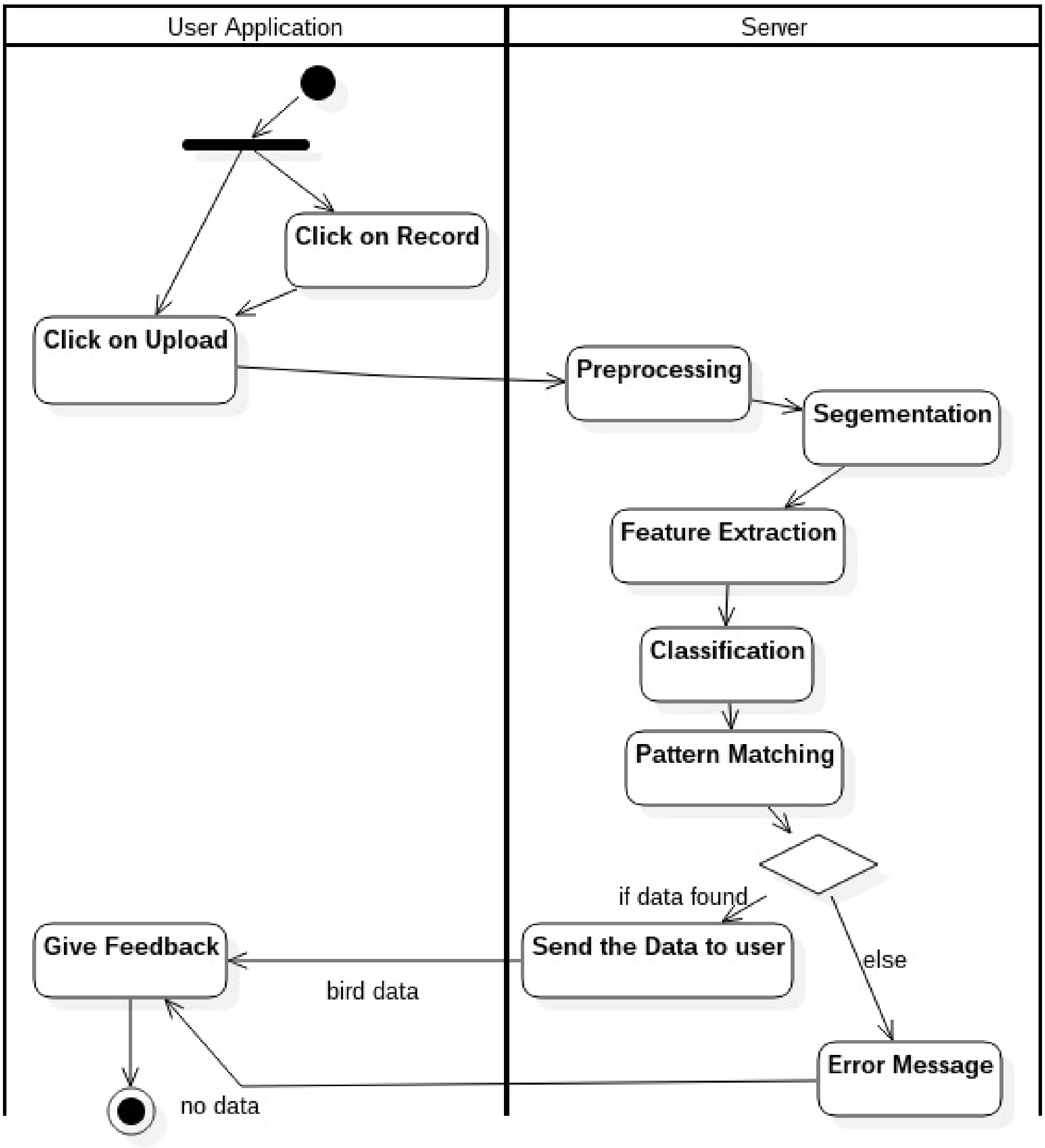
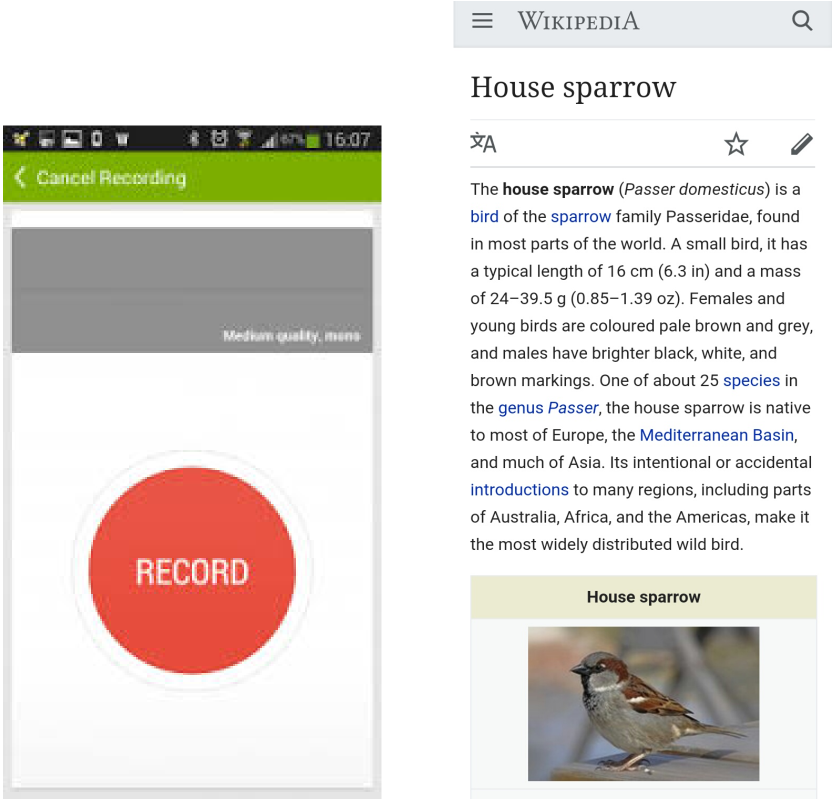


Figure 9: Activity Diagram.

8.10 How the interface will look

The Screens of input and output will look as follows:-



House sparrow

The **house sparrow** (*Passer domesticus*) is a bird of the sparrow family Passeridae, found in most parts of the world. A small bird, it has a typical length of 16 cm (6.3 in) and a mass of 24–39.5 g (0.85–1.39 oz). Females and young birds are coloured pale brown and grey, and males have brighter black, white, and brown markings. One of about 25 species in the genus *Passer*, the house sparrow is native to most of Europe, the Mediterranean Basin, and much of Asia. Its intentional or accidental introductions to many regions, including parts of Australia, Africa, and the Americas, make it the most widely distributed wild bird.

House sparrow

(a) Input Audio (b) Output Details

Figure 10: UI Screens

9 Preprocessing Done As of Now

Noise Reduction using Audacity

Audacity is an open source platform for sound processing,it provides arrays of options for sound preprocessing,in our project we have used noise cancellation feature to remove noise from the dataset.

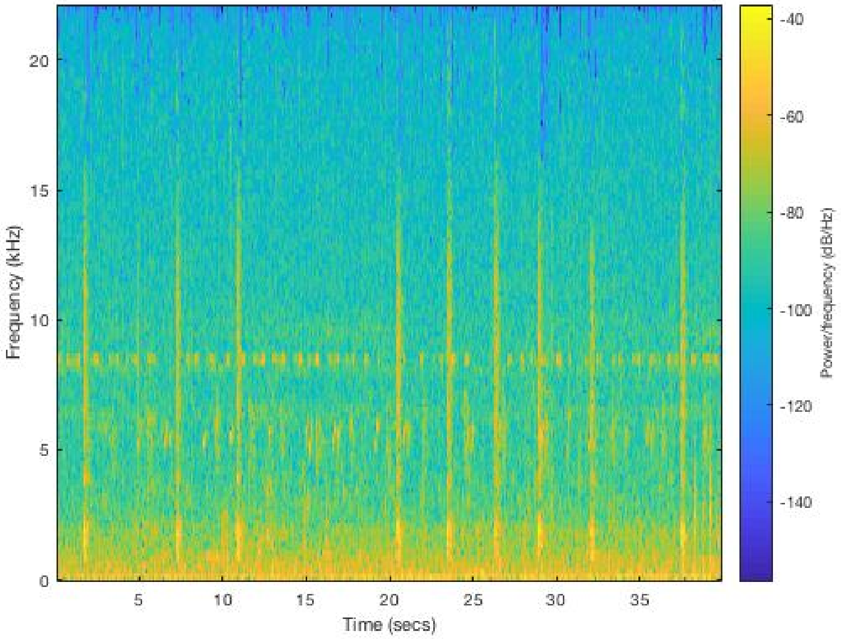


Figure 11: Spectrogram Before Noise Reduction.

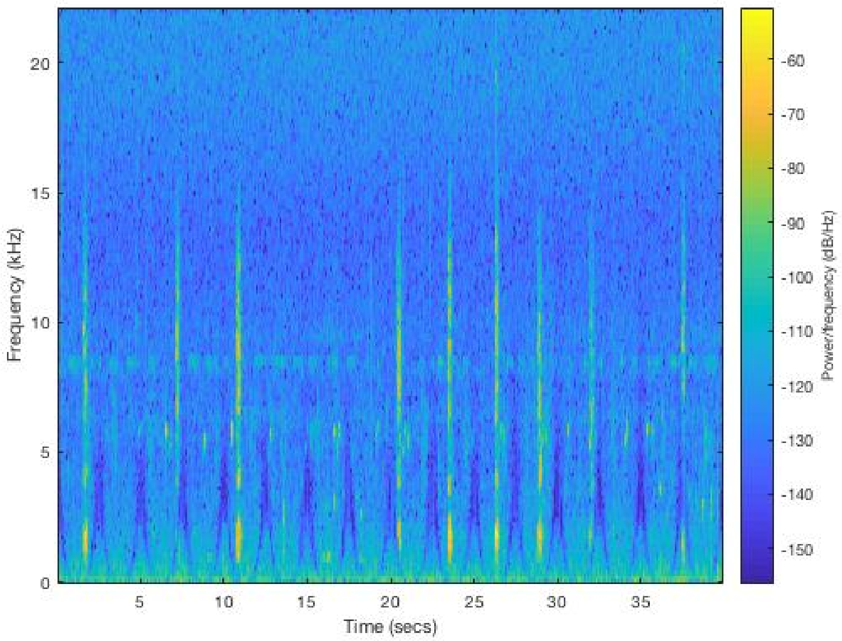


Figure 12: Spectrogram Before Noise Reduction.

On comparing we can see the difference in the spectrograms

10 Summary

In this project we are attempting to develop an Android application for recognition of bird sounds for local birds found in India. So far we have read papers related to sound recognition, bird sound features,Machine learning classifier for sounds and tools which are used for sound processing like OpenSMILE, Matlab, Audacity etc. Datasets are being selected and will be randomly fed to the system of which 80% will be used to train the model and rest 20% will be used for identification and calculations. We have also identified the system flow of our proposed project which will comprise of sound recording, its preprocessing and then feeding it to the trained model to get appropriate results. Now we are researching on the best machine learning models which can be incorporated in our project for bird sound recognition based on their accuracy, few of which are been shortlisted based on research papers we read so far.

References

[1] Rai, Pallavi, Vikram Golchha, Aishwarya Srivastava, Garima Vyas, 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 Bjrn Schuller. "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." *PeerJ* 2 (2014): e488.

[5] McIlraith, Alex L., and Howard C. Card. "Birdsong recognition using backpropagation and multivariate statistics." *IEEE Transactions on Signal Processing* 45, no. 11 (1997): 2740-2748.

[6] Huang, Guang-Bin, Qin-Yu Zhu, and Chee-Kheong Siew. "Extreme learning machine: a new learning scheme of feedforward neural networks." In *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, vol. 2, pp. 985-990. IEEE, 2004.

[7] Huang, Guang-Bin, Hongming Zhou, Xiaojian Ding, and Rui Zhang. "Extreme learning machine for regression and multiclass classification." *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 42, no. 2 (2012): 513-529.

[8] Suykens, Johan AK, and Joos Vandewalle. "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] Raghuram, M. A., Nikhil R. Chavan, Ravikiran Belur, and Shashidhar G. Koolagudi. "Bird classification based on their sound patterns." *International Journal of Speech Technology* 19, no. 4 (2016): 791-804.