Int J Speech Technol (2016) 19:791–804
DOI 10.1007/s10772-016-9372-2

Bird classification based on their sound patterns

M. A. Raghuram[1] · Nikhil R. Chavan[1] · Ravikiran Belur[1] · Shashidhar G. Koolagudi[1]

Received: 28 July 2016 / Accepted: 20 September 2016 / Published online: 30 September 2016

Abstract In this paper we focus on automatic bird classification based on their sound patterns. This is useful in the field of ornithology for studying bird species and their behavior based on their sound. 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 study is also 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.

Keywords Machine learning · Audio processing · Bird call · Bird species · Bird weight · Bird habitat

✉ M. A. Raghuram
maraghuram@gmail.com

Nikhil R. Chavan
nikhilchavan93@gmail.com

Ravikiran Belur
ravikiranbelur@gmail.com

Shashidhar G. Koolagudi
koolagudi@yahoo.com

[1] Department of Computer Science and Engineering, National Institute of Technology Karnataka, Surathkal, Mangalore, India

1 Introduction

Bird sounds are of great importance in ecology and in monitoring the environment. Throughout history, humans have considered birds to be their protectors, their vigilant sentinels, writes the Nobel laureate immunologist Peter Doherty in his book "Their Fate Is Our Fate: How Birds Foretell Threats to Our Health and Our World" in 2012. According to an eco-toxicologist at the University of Saskatchewan in Saskatoon, birds can tell us a lot about what's going on around us that we might not be able to see. In this context, it is easier to monitor birds to understand environmental changes. Therefore, processing of birds sounds is of significant interest among ornithologists. In recent years, new techniques have evolved to apply signal processing and machine learning tools to process bird sounds.

Technical analysis of bird sounds has a long and challenging history. There are over nine thousand bird species in the world and because of this large number of species most people cannot recognize them. Earlier studies used small data sets and classification was done manually with minimal practical needs. Majority of the traditional studies on the analysis of bird's vocalization are based on visual inspection of sound spectrogram. To continuously identify the spectrogram of large set of bird sounds is a strenuous and tedious task and many of the currently available methods are accurate for only a relatively small sets of bird sounds. Therefore, automating the process of identification of birds with minimal manual intervention is of great importance. Another important aspect of previously studied methods is the quality of recordings. Most of these studies are based on recordings with low or negligible noise component and rely on expensive audio equipment for this purpose. In such cases, several techniques have proved to be successful in classifying birds from audio, but cannot

scale to practical scenarios without significant changes in the methodology (Scott 2008; Somervuo et al. 2006; Juang and Chen 2007).

Task specific research on classification and identification of bird sounds has been a challenging task and has only recently attracted the attention of the research community because of its wide variety of applications that are highly relevant in recent ecological scenarios. For example, changes in bird song can be used to understand anthropogenic noise particularly in urban areas (Bermúdez-Cuamatzin et al. 2010; Luther and Baptista 2010; Dowling et al. 2012). The aim of this study is to develop automated techniques for identifying bird based o their sound patterns. Possible applications of these techniques would enable people identify different features of birds like weight, habitat, species etc. and use these results for further ecological or biological studies.

Classification and identification of bird's sounds can be done by comparing some basic features that all birds share. In this study, we collect bird sound recordings from openly available resources on the Internet such as ecology audio libraries (Macaulay, Xeno-canto etc.) or those uploaded by hobby bird enthusiasts. These recordings are representative of the natural environment which includes several types of noise, low quality equipment and arbitrary length of recording. We describe the process of pre-processing bird sound recordings to eliminate the background noise that degrades the signal. Then we attempt to identify several important attributes of the bird from its sounds—namely weight, habitat, type of sound and the specie itself. Finally, we compare several audio features and machine learning algorithms to determine efficient techniques for identifying these bird attributes. While studying audio features, we explore simple alternatives such as pitch, energy, tempo etc. instead of the commonly used audio feature—MFCC, wherever possible.

The rest of the paper is organized as follows. In Sect. 2 the related works are presented. Section 3 describes methodology, while Sect. 4 explains experiments and results. The paper concludes with Sect. 5 by highlighting some future research directions.

2 Literature review

Automatic classification of bird species from bird sound samples has recently attracted the interest of the research community. The two major tasks involved in this process are use of signal processing and machine learning techniques (Taylor 2008). Signal processing is a broad term that involves the use of audio processing techniques to improve signal quality and extract a set of features from the audio signal (Lathi 2004). Machine learning algorithms use these features to develop decision methods that can predict and classify the audio patterns (Mitchell 1997; Frank 2005). There have been many recent studies in the field of automatic bird species identification with the use of machine learning techniques because of their high accuracy (Lopes et al. 2011; Sun et al. 2013; Acevedo et al. 2009). The major drawbacks with these studies is the large number of training samples and high quality of audio recordings required.

Bird sounds can be broadly classified as calls and songs. Bird songs are longer vocalizations which usually include a variety of notes in a sequence while bird calls are short communications which are often the single notes. Prior to analyzing bird calls it is important to segment calls into distinct syllables. A syllable is a short utterance by a bird which may be a call or a song. The algorithms used in segmentation depend on the audio samples that are available. The most commonly used technique for bird sound processing is energy based time-domain approach which is reliable for single bird's samples with low noise (Somervuo et al. 2006; Juang and Chen 2007). In the case of multiple bird's sounds in noisy environments, two dimensional time–frequency based segmentation is used (Mellinger and Bradbury 2009). The most widely used features to describe bird's sounds are Linear predictive coefficients (LPC) and Mel-frequency cepstral coefficients (MFCCs) which are also prevalent in other areas of signal processing (Chen and Maher 2006; Davis and Mermelstein 1980). Recent studies have been carried out in using multi-instance multi-label learning (MIML) techniques which were earlier used on image, text and audio samples where it is less costly to obtain labels at the bag level in contrast to labeling individual calls. Attempts have also been made to solve the problem of detecting bird sounds in complex environments such as the sound samples obtained from automatic recording units (Briggs et al. 2012; Bardeli et al. 2010).

There are references of many papers in recent years which have addressed the fundamental problem of automatic bird species identification (ABSI) using signal processing and machine learning techniques. Majority of these papers can be analyzed based on the feature-set and the machine learning techniques employed. Somervuo et al. (2006) achieved an accuracy of about 71.3 % using Mel-frequency cepstral coefficients (MFCCs). Fagerlund (2007), in another work, obtained an overall accuracy of 98 % with a database of eight bird species using MFCC and low level signal parameters. Support vector machines is used at every node in the global decision tree for classification (Fagerlund 2007). Vilches et al. (2006) employed data mining algorithms such as ID3 and J4.8 which yielded the result of classification of about 98.4 % from 154 bird song recordings. Lopes et al. (2011) showed that the music analysis, retrieval and synthesis for audio signals

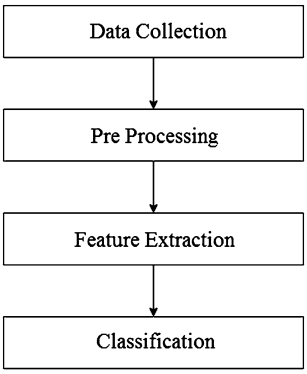


Fig. 1 General stages in automatic bird identification process

(MARSYAS) feature set along with a multi-layer perceptron neural network classifier achieves the best known accuracy of 99.7 % using about 100 recordings of three species of bird. Carlos et al. (2013) address the task of hierarchical bird species identification from audio recordings. Their work evaluates three different approaches of hierarchical classification problem, namely, the flat classification approach, local-model per parent node classifier approach and global-model of hierarchical classification approach. The first and second approach use classical Naive Bayes algorithm and the third approach uses Global Model Naive Bayes (GMNB) algorithm.

Kwan et al. (2004) proposed the development of a high performance bird classification system using Hidden Markov Model and Gaussian Mixture Model. The recognition in Hidden Markov Model uses Principal Component Analysis (PCA) and Vector Quantization (VQ). PCA is used for data dimension reduction. In Gaussian Mixture Model (GMM), the algorithm first extracts feature vectors from sound data and matches them with the parameters of trained GMMs, one GMM trained for each bird class. Difference between the probabilities is compared to a preset threshold to classify bird class. In Tsai et al. (2013) the authors proposed an automatic bird sound identification system built upon a two-stage identification framework. The first stage performs identification of song/ call and the second stage is performed by song or call handler based on the output of first stage. In both stages timbre and pitch features are used to identify the bird species. In using timbre features, audio clips are converted into MFCCs and their first derivatives. While using pitch feature, sound clips are converted into MIDI note sequences and then bi-gram models are used to analyze the dynamic change information of the notes.

3 Methodology

The proposed approach is divided into four stages, as shown in Fig. 1—collection of bird sound data, pre-processing of audio clips to improve the quality of the signal, feature extraction and using machine learning algorithms on extracted features for classifying the sound patterns.

3.1 Data collection

The dataset used in this work is obtained from different openly available web sources such as Xeno-canto, Macaulay Library, Western Soundscape Archive etc where the sound recordings are labelled by users. The sound recordings are with different sampling rates. To conduct the experiments, all the sound data were converted into WAV format at sampling rate of 16,000 Hz. Our dataset finally consists of thirty nine commonly-seen bird species of different sizes based on the quality of audio recordings that could be obtained. These species include White-throated Thrush, American Black Duck, Black Vulture, Jabiru, Wild Turkey, Brown Thrasher, Northern Lapwing, American Robin, Common Myna, Common Sandpiper, Barred Owl, Gyrfalcon, Indian Peafowl, Asian Brown Flycatcher, Great Kiskadee, Himalayan Snowcock, Ferruginous Hawk, Common Snipe, Anna's Hummingbird, Snowy Owl, Dark-sided Flycatcher, Nelson's Sparrow, Wood Stork, Eastern Bluebird, Bald Eagle, Oriental Cuckoo, Anhinga, Black-and-white Warbler, Bananaquit, Fan-tailed Warbler, Roseate Spoonbill, Abert's Towhee, Cactus Wren, Common Starling, Barn Swallow, Kelp Gull, Red-billed Tropicbird and Roseate Tern.

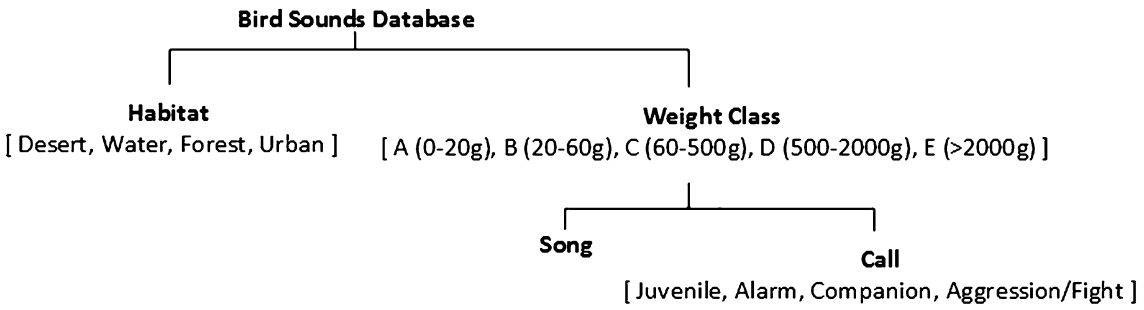
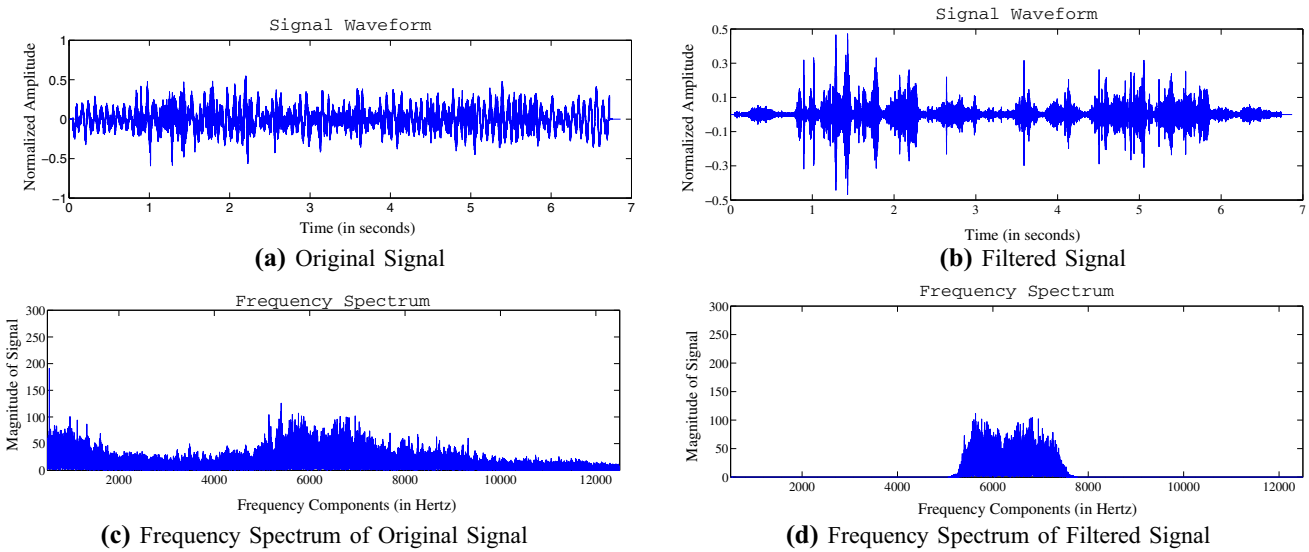


Fig. 2 Organization of bird sound database



(a) Original Signal **(b)** Filtered Signal

(c) Frequency Spectrum of Original Signal **(d)** Frequency Spectrum of Filtered Signal

Fig. 3 Effect of Butterworth filter on Anna's Hummingbird song recording

Database Structure The bird sound database is arranged in a hierarchical manner as shown in Fig. 2. There are thirty-three bird species that have been arranged in this manner. The weights (in grams) of each bird specie was found manually and the birds were then classified into five categories / classes (class A through class E) based on their weight. The entire corpus is classified based on the weight class and also classified based on the habitat at the same time. Once sorted by the weight class, they are classified as song or call, and then further classified based on the type of call.

3.2 Pre-processing of audio clips

Since the audio clips are natural recordings, they have lot of background noise. The original audio files also include undesired environmental sounds. It is imperative to isolate the desired bird call/ song from background noise, or at least reduce the effect of noise. Hence pre-processing of the sound recordings is necessary to have good performance of the system to be developed. Noise reduction techniques (like filters) have to be used to remove from signal some unwanted noise components thereby improving the audio quality of the recordings. A frequency based band-pass filter (Butterworth filter) is used to remove background noise as the frequency range can be specified and band-pass filter has an extremely flat frequency response in the pass band and thus enhances the quality of the audio.

Overlap of sound signals of multiple birds can be observed in the natural recordings. However, in this work we assume that there is only one dominant bird species which produces the sounds in the clip. Any other components adding to its audio signal must be filtered out. To obtain filtered signal that contains the sound of a specific bird, a band of known frequency range for a bird is set. This process of filtering signals based on frequency bands is semi-automatic, in the sense that average pitch value over the recording is computed using the harmonic product spectrum pitch estimation algorithm. Using this information, the Butterworth filter is manually adjusted and spectrogram is used to verify the signal waveforms. The obvious drawback with this method is if some other bird also makes sound in the exact same passband. However, in this case, birds having similar frequency bands, tend to share similar characteristics. Therefore it is more feasible to use multi-label learning methods to infer characteristics of the different species present in the recording simultaneously rather than painfully separate the audio signals of different birds. In this study, we simply ignore such recordings because of the abundance of alternative recordings in our data set. Figure 3 shows the effects of using band-pass Butterworth filter on a Anna's Hummingbird recording in the audio database.

Figure 3a and c show the waveform and frequency spectrum respectively of the song recording of Anna's Hummingbird as present in our database. A bandpass Butterworth filter (from around 5000 to 7500 Hz) as shown in Fig. 3d, without any amplitude amplification is applied. This results in a filtered signal waveform (as shown in Fig. 3b), which is free from the background noise and has better audio quality. The original signal in Fig. 3a may not contain any useful information whereas the signal obtained in Fig. 3b is suitable for feature extraction.

3.3 Feature extraction

There has been a lot of research aimed at understanding the similarities between bird vocalization and human speech (Doupe et al. 1999; Okanoya and Scharff 2010; Beckers 2011). In these studies, it has been shown that there are numerous parallels between the development and mechanistic processes for production of sounds in birds and humans. Therefore, some of the features that are commonly applied to process human speech are also applicable to birds.

Bird species specific feature extraction is one of the main steps before one proceeds with bird classification. The simplest individual sounds that birds produce, referred to as song elements may be observed as notes. A set of one or more elements that occur successively in a regular pattern is referred to as a song syllable. A sequence of one or more syllables that occurs repeatedly is regarded as a song motif or phrase (Lee et al. 2008). Based on this knowledge, it is decided to use prosody related information for characterizing bird sounds. The important correlates of prosody of a natural signal are pitch pattern (intonation), energy profile and duration. MFCCs are also used to capture the characteristics of vocal cavity of specific species of bird.

- *Pitch features* In this work pitch is used as a primary feature to classify birds. The Fast Fourier Transform (FFT) method is applied to translate time domain information to frequency domain information. Then, we use the simple harmonic product spectrum (HPS) algorithm to estimate the average pitch for each frame of the signal. The HPS of each frequency bin(f), as defined in Eq. 1, is the geometric mean of the amplitudes of its harmonics. In our case we consider up to $N = 4$ harmonics of each frequency bin. These frame-wise estimates are then aggregated over the recording to compute statistics like average, standard deviation, maximum and minimum pitch values.

$$HPS(f) = \sqrt[N]{\prod_{i=1}^{N} Amplitude(i * f)} \tag{1}$$

- *Energy based features* Energy is also one of the basic prosodic features. Frame wise energy values are calculated from the time domain information for the signal by taking the sum of square of each sample in the frame. Average, standard deviation, maximum and minimum energy values are obtained from all frames in the signal.
- *Duration related features* Speech units, pause and rate are important duration related attributes used in speech processing. Duration feature is similar to identifying the silence or noise region in the speech signal. It is generally computed frame-wise and then aggregated over the entire signal. In our experiments, we measure duration as the time between two silence regions. This is computed by finding the auto-correlation of low amplitude signals (Eq. 2). In the equation $\phi_f(t)$ represents the auto-correlation of signal f at time t. It is an indirect measure of the length of a single vocalization of bird. In case of bird calls, this length is expected to be shorter while in case of songs, it is expected to be longer.

$$\phi_f(t) = \frac{1}{N} \sum_{m=-N}^{N} f(m) f(m+t) \tag{2}$$

- *Tempo* Tempo is a musical feature that describes the speed of a song as perceived by human ear. A similar metric is the number beats in the audio—beats per minute (BPM). Tempo is computed by selecting the peaks over the auto-correlation of the onset curve. In our experiments we use a simple logarithmic novelty curve on the magnitude spectrogram for onset detection, as shown in Eq. 3 where X is the onset curve, S is the magnitude spectrogram and C is a large constant. In case of bird sounds, tempo is useful for identifying different types of bird calls.

$$X = log(1 + C * S) \tag{3}$$

- *Pulse clarity* Pulse clarity is also a musical feature that is representative of the clarity of beats as perceived by a listener. With respect to bird sounds, pulse clarity along with tempo is used to identify bird calls efficiently. This is because of the different structure/beat patterns in different types of calls. For implementation, we leverage the *mirpulseclarity* functionality provided in MIRtoolbox (Lartillot and Toiviainen 2007). It was developed by using several pulse clarity predictors estimated using different onset curves which were then evaluated by trained musical participants (Lartillot et al. 2008).
- *MFCCs* Mel-frequency cepstrum (MFC) is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency f as show in Eq. 4. Mel-frequency cepstral coefficients (MFCCs) are coefficients that collectively make up an MFC. The shape of the vocal tract manifests itself in the envelope of the short time power spectrum, and the job of MFCCs is to accurately represent this envelope.

$$Mel(f) = 1125 ln(1 + \frac{f}{700}) \tag{4}$$

The feature vector extracted from the sound signal is shown in Fig. 4. Around twenty seven features are

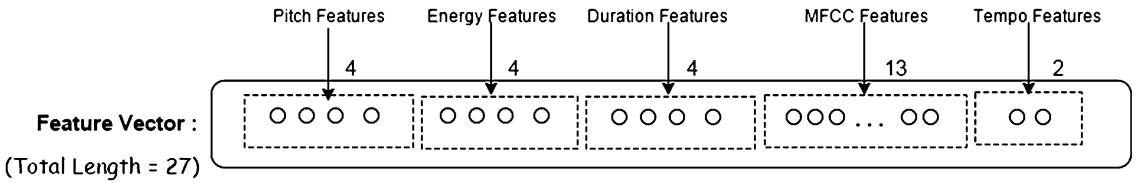
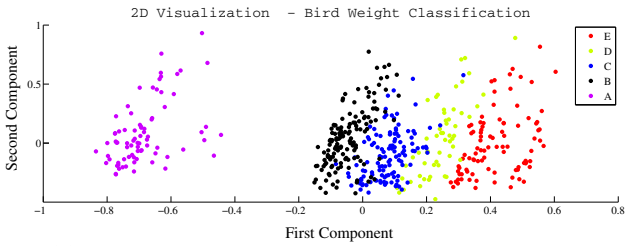
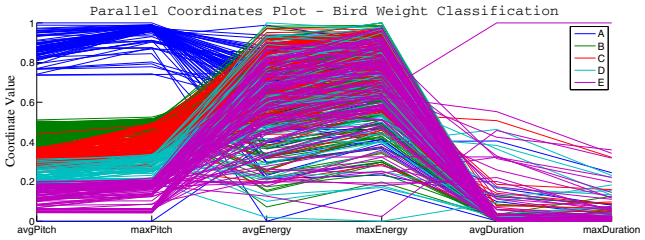


Fig. 4 Feature vector extracted from bird sound signal



(a) 2-D plot of feature vector after principal component analysis



(b) Parallel plot of feature vector

Fig. 5 Visualization of feature vector (after normalization) used in predicting bird weight class

computed, but it is important to note that the entire feature vector is not used for every experiment. For example, in case of predicting the bird habitat only the pitch and energy features are sufficient and give good classification accuracy. Figure 5 gives an understanding of the feature vector used in one of the classification tasks—predicting the weight class of bird.

3.4 Classification methods

In the classification phase, the comparison of popularly used supervised classification methods is carried out. This section briefly describes the working principles of the classifiers that were used for training/testing the data-set.

- **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. Given a set of features $X = f_1, f_2, ...f_d$ extracted from the audio and a set of classification categories $c_1, c_2, ...c_k$, the Naive Bayes classifier assigns that class c_i that has the maximum posterior probability i.e. $c_i = c_j | P(C_j|X)$ is maximum. In the experiments, the Gaussian distribution is used whenever the attributes are continuous. The Bayesian classifiers are useful when the features selected are known to have conditional independence. For example, in case of estimating bird weight, we use pitch and energy features which can be considered to be independent of each other.
- **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 (as shown in Eq. 5, where X and Y are vectors in input space), which performs an implicit mapping from the input feature vector to high-dimension feature space for identifying a clearer margin. For solving the problem of classifying multiple audio classes, the 1 versus 1 (binary classification) method is used. In general, the support vector machine algorithm performs well given sufficiently large training data and also has low running times.

$$K(X, Y) = (1 + \sum_{i=1}^{N} x_i y_i)^2 \quad (5)$$

- **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

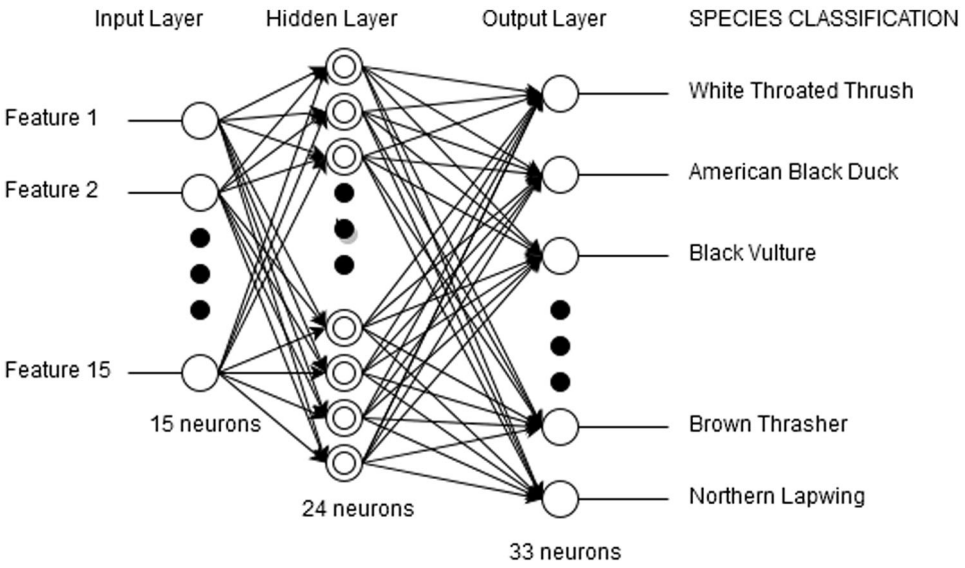


Fig. 6 The structure of a neural network

organized into layers, namely input, hidden and output layers which converts an input vector into output. The Fig. 6 describes the ANN classifier used in the experiment to classify bird species, which uses a single hidden layer with 24 neurons.

4 Results and discussion

In this section we discuss in detail about our proposed methods—the setup, features and classifiers used in each of the five classification tasks. These tasks include predicting size of the bird, classify songs and calls of birds, classify different types of bird calls and predicting habitat of the bird. Finally, we combine features from the bird recordings, with the results of preceding classification tasks to predict the bird species itself.

4.1 Predicting size of the bird

In order to facilitate experiments on predicting the size of a bird, the database is organized into five different categories based on bird weights. Although information about bird weights is openly available (Dunning 2013), there are many variations of the weight of a bird between seasons, climates and even between male and female birds of the same species. The average value over these variations is calculated so that each bird species belongs to exactly one weight category. Table 1 gives the classification of birds based on their weight. The underlying hypothesis for this experiment is that body mass of a bird is closely related to the pitch and amplitude of it's sounds, which has been widely investigated in the past (Clark 1979; Hall et al. 2013; Linhart and Fuchs 2015). Therefore, the features used in this classification task are limited to pitch and energy.

Table 1 Classification of birds based on their weight

Weight class	Range of bird weights (g)	Example
A	0–20	Anna's Hummingbird
B	20–60	Cedar Waxwing
C	60–500	American Robin
D	500–2000	Barred Owl
E	>2000	Wild Turkey

Table 2 Classification accuracy of different classifiers for predicting bird size

Classifier	Classification accuracy (%)
Support vector machines	87.29
J-48 decision tree	96.13
Naive Bayes	91.16
Neural networks	93.93

To compute the average frequency of each recording, the fast Fourier transform method is applied on bird sound recordings. The frequency values are the pitch components of the particular bird. The pitch values and energy values are fed as input to multiple classifiers and it can be seen from Table 2 that the decision tree classifier gives the best results. Size of a bird has a close correlation with its pitch. This is illustrated in Fig. 7 which shows the relation between the weight of a bird and its pitch features (average pitch and maximum pitch). It can be seen that using average pitch alone, it is possible to distinguish the weight classes, although not completely. In order to achieve better distinction, additional features like maximum pitch, energy features(maximum energy, average

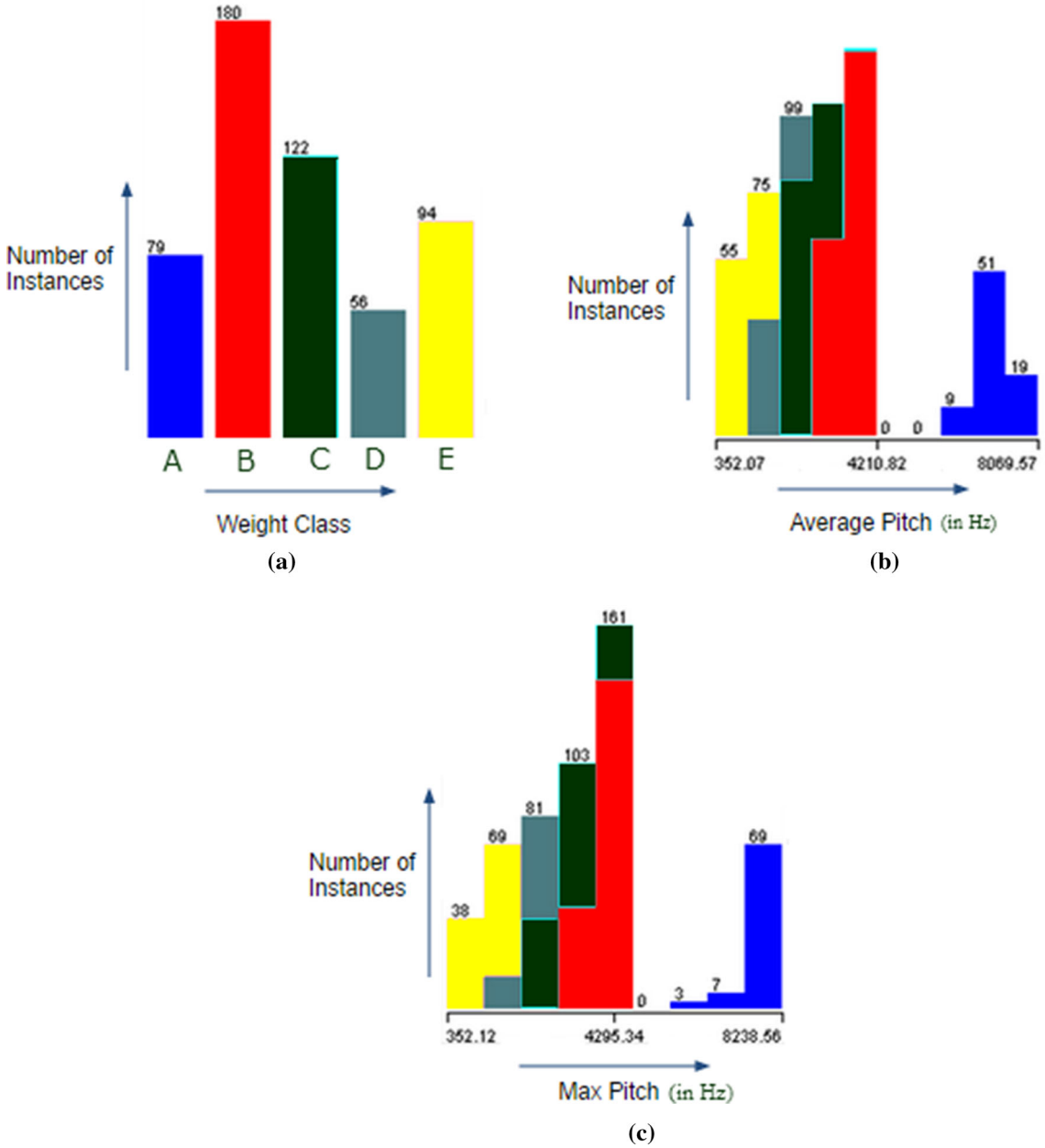


Fig. 7 **a** Distribution of number of birds in different classes. **b** Number of birds falling in each of the weight categories based on the average pitch values. **c** Number of birds falling in each of the weight categories based on their max pitch values

Table 3 Confusion matrix for predicting bird size using J-48 classifier where A, B, C, D and E are weight classes

	A	B	C	D	E
A	24	0	0	0	0
B	0	64	4	0	0
C	0	1	36	2	0
D	0	0	0	14	0
E	0	0	0	0	36

energy) can be used. Table 2 gives the classifier accuracies for different classifier and the highest classifier accuracy is 96.13 %. Table 3 represents the confusion matrix for J48 classifier trained using 351 instances with five attributes namely weight, average pitch, maximum pitch, average energy and maximum energy.

It may be observed from the table that out of 68 sound clips of class B, 64 are correctly classified and 4 are misclassified as of class C (observe third row and third column of table). Another important observation is that whenever incorrect classification has occurred in the matrix, it is always classified to the weight classes adjacent to the original weight class. This is expected as the weight classes are not perfect and it considers the average weight of a bird.

Table 4 Song versus call classification for all weight classes

Weight class	Classification accuracy (%)
A	92.2
B	88.0
C	79.6
D	95.6
E	100.0
Average	93.87

4.2 Classifying bird songs and calls

In this task, we attempt to differentiate bird songs from calls given that we have already classified the weight class of the bird. Bird songs are relatively longer in duration, often melodious and complex series of notes usually associated with some sense of courtship while calls are brief sounds of simple acoustic structure. This leads to the hypothesis that songs are of longer duration and higher energy overall compared to calls of the same bird. The duration feature nicely fits this use case and is used along with pitch and energy(used in previous task). These features are fed as input to the decision tree classifier to output one of the two classes-song or call.

The reason for retaining the decision tree classifier is simple. Firstly, the feature set is almost the same as the previous classification task except for the added duration feature. Secondly, duration can be thought of as a binary variable(long or short) which suits the attribute selection process in a decision tree. Table 4 shows the average classification of songs and calls in each of the five weight categories using average duration, max duration, average energy, max energy, average pitch and max pitch values as attributes for classification. J48 classifier yields an average accuracy of 93.87 % in classification over all five weight categories of the birds.

4.3 Classifying different types of bird calls

There are several different types of bird calls that have been identified and studied by ornithologists. Four of these that are important and frequently observed are,

- Companion Call: Call notes are short communications, often single notes. They are generally used between mates or members of a flock to signal each others whereabouts and to point out food.
- Juvenile Calls: These are calls made by baby birds in a nest that typically sound like a racket as they beg for food.
- Bird to Bird Aggression Calls: Birds often compete for territory, mates, and food sources. These are the calls to fight against competitor.
- Alarm Call: This type of call by a bird indicates a potential danger/predator on the landscape.

This task of identifying bird calls is carried out specific to each weight class. In this case, the differentiation in the types of calls is based on the structure and intonation of these sounds. Therefore, we use the previously used pitch, energy and duration features along with tempo and pulse clarity, which are based on identifying peaks in the audio. Table 5 summarizes the results for average class or song classification with each weight classes using different classifiers. It can be observed that J48 classifier, which is a decision tree based classifier gives the best classification result because the number of feature vectors used for classification is less.

4.4 Predicting bird habitat

The environmental condition of the habitat, restricts the communication efficiency and efficacy between a sender and a receiver. The sound reaction to habitat conditions is not constant in different situations. Frequency and structure of the acoustic repertoire are the plastic traits that get modified according to the environmental constraints. According to the acoustic adaptation hypothesis, the environment is an important cause of modification and alteration of acoustic signals. Dominant frequencies of the bird species and ability of long-distance calls are the results of an interaction between the birds and their surrounding environment to maximize the reach of the emitted sound. The range of frequencies at which birds call in an environment varies with the quality and type of habitat and the ambient sounds (Laiolo 2010; Kight and Swaddle 2011). For example, one such study found that in urban environments which tend to be noisy, the birds used higher frequency sounds in order to avoid the environment noise which is mainly composed of lower frequency components

Table 5 Average call or song classification within each weight classes using different classifiers

Classifier/weight class	A (<20 g)	B (20–60 g)	C (60–500 g)	D (500–2000 g)	E (>2000 g)
J48	69.5	67.3	70.4	91.3	90.5
Naive Bayes	60.86	55	43.78	82.9	80.2
Bayes net	86.6	80.3	75.4	90.2	90
Neural net	85.9	78.6	72.1	89.7	88.5

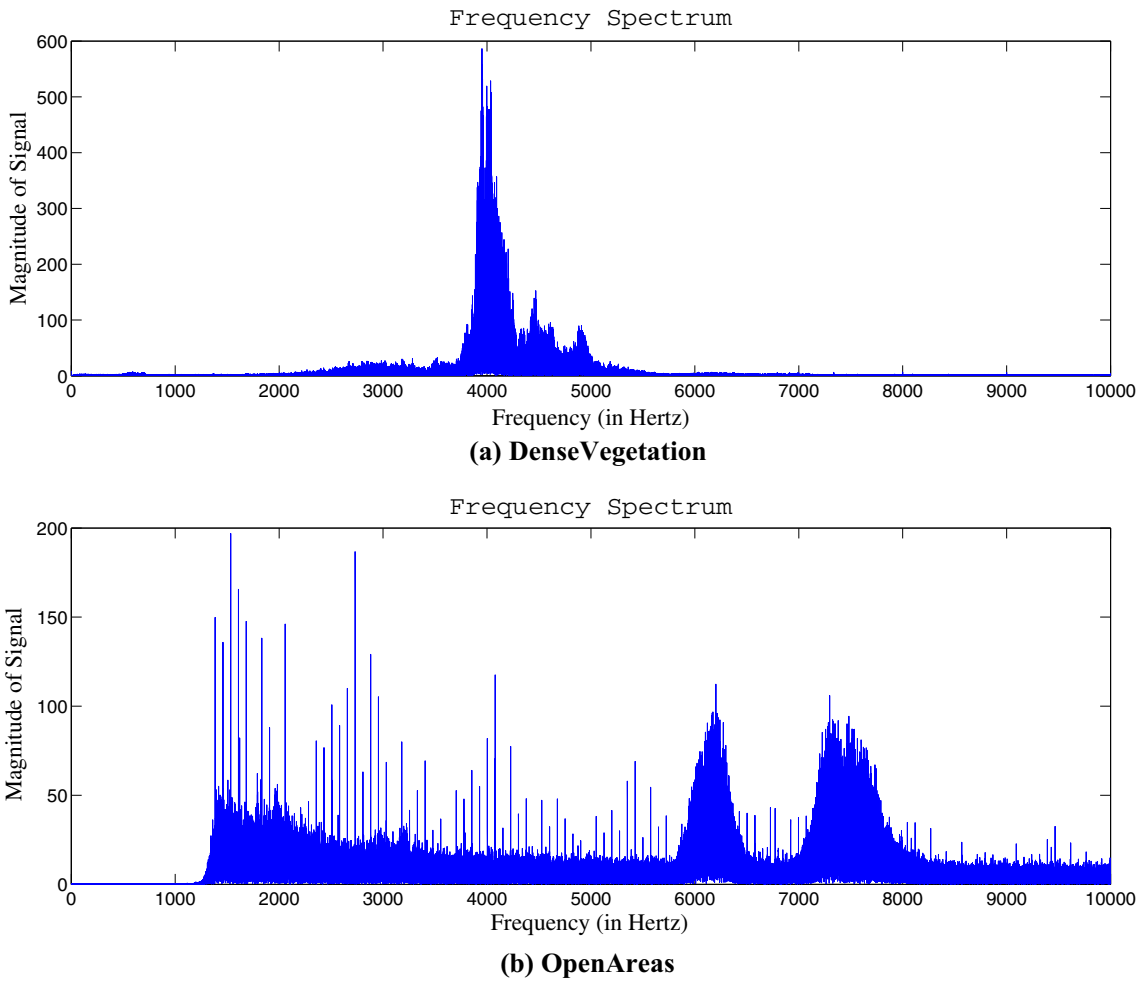


Fig. 8 Comparison of histograms of frequency components for sound of birds present in dense vegetation and open area

(Slabbekoorn and Peet 2003). This type of information can be used to derive bird's habitat based on their sounds.

Figures 8 and 7 shows the comparison of histograms sound of bird from dense vegetation and open area respectively. It can be observed that birds in dense vegetation produce sounds which primarily contain high frequency components in a narrow frequency band while birds in open area produce sounds with lower frequency components in a wider frequency band.

To study habitat prediction of birds from their sounds, the bird database is divided into water birds, desert birds, urban birds and forest birds based on their habitat.

- Forest : Forest is a large land area which is covered with dense woody vegetation like trees. They are the major form of terrestrial ecosystem on Earth.
- Water : Water habitats better known as marine habitats are ecosystems that support marine life, and these include habitats like oceans, rivers, lakes etc.
- Urban : The Urban environment refers to environments dominated by high-density residential and commercial buildings, paved surfaces and other urban-related factors.
- Desert : Desert is a barren land area where little precipitation occurs and hence is a dry terrain.

Table 6 Bird habitat classification for different classifiers

Features/classification accuracy	Prosodic features	MFCC features	Prosodic + MFCC features
J48	66.6	72.6	74.7
Naive Bayes	55.8	62.4	66.6
Bayes net	72.6	70.3	80.2
Neural net	62.4	72.6	74.7

Table 7 Confusion matrix for predicting bird habitat

Habitat class/classified as	Forest	Water	Urban	Desert
Forest	16	1	2	0
Water	0	11	0	2
Urban	0	1	36	2
Desert	0	0	0	14

Experiments are conducted using 149 training instances and 51 testing instances separately for different classifiers, namely J48, Naive Bayes, Bayesian network and neural network. The experiments are conducted separately using prosodic features, MFCC features and the combination of both. Results are shown in Table 6. Table 7 shows the confusion matrix of classification of bird habitat into different habitat, namely forest, water, desert and urban. The prosodic features used are metrics that measure the distribution of the Pitch and Energy features(mean, standard deviation, max/min etc.). These features are extracted on the basis of the acoustic adaptation theory as described earlier and the MFCC features(most widely used features in audio signal processing) were used to boost the classification accuracy. From the Table 6 we can observe that prosodic or standard MFCC features alone are not sufficient to achieve reasonable classification accuracy. By combining both these features, we can see an improvement in the classification accuracy for all classifiers. The Bayesian networks based classifier results in the highest classification accuracy- 80.2 % when combined with prosodic and MFCC features. The Bayesian networks classifier is independent of the ordering of the features and performs well when the features used are related and these relationships vary with different classes. From the confusion matrix, we can see that there is an overlap between forest and urban classes (dense vegetation/close habitat), and between water and desert classes (less vegetation/open habitat).

4.5 Predicting bird species

There are over nine thousand different bird species in the world. However, due to its large number as well as a multitude of different bird sounds and different background noise conditions in the recordings, there are many challenges in doing certain useful tasks using bird sounds. The proposed method of identifying the bird species from their sound is divided into two stages. In the first stage, bird's weight class, call type and their habitat are identified along with extracting robust features from the audio recordings. In the second stage, the results of weight class, call type and habitat done using previously proposed methods, along with earlier measured prosodic features (pitch, duration and energy) are fed to the bird species classifier as shown in Fig 9. Totally, fifteen features are used in this task—weight, call type, habitat, statistics (average, standard deviation, max and min) of pitch, energy and duration.

Experiments are conducted using 701 instances of thirty three different bird species. Weight class of the bird, type of call, habitat, minimum pitch, maximum pitch, minimum energy, maximum energy, minimum duration and maximum duration are used as attributes in the same order. Random forest classifier has been used for classification giving an average classification of 83.88 % for thirty three different bird species. Random forest consisting of 100

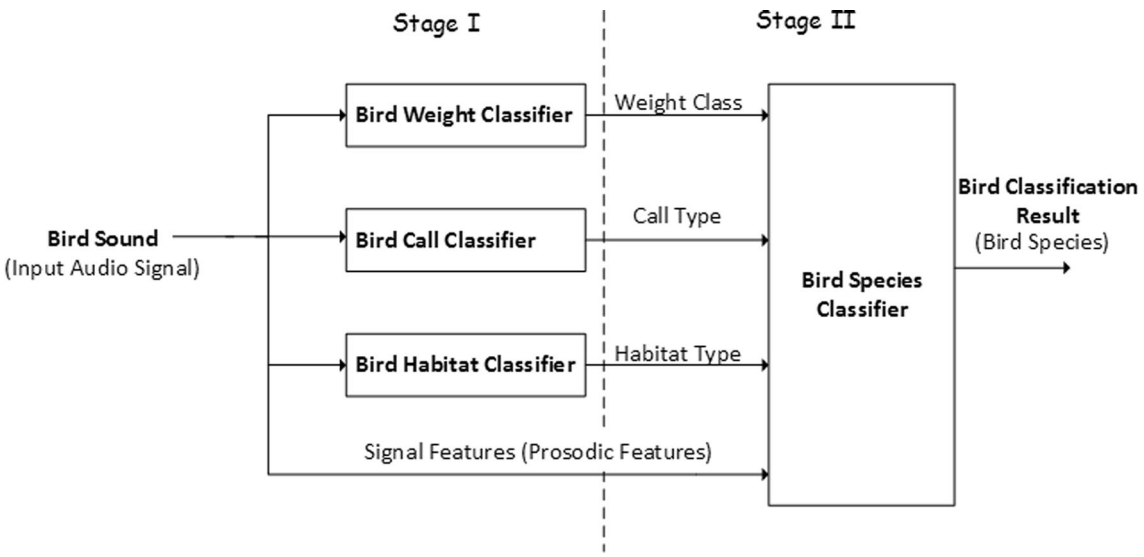


Fig. 9 Block diagram of species classification task

Table 8 Classification result for predicting bird specie

Classifier	Accuracy (%)
J-48 decision tree	79.60
Random forest	83.88
Bayes net	69.90
Support vector machines	70.75
Neural network	78.60

Table 9 Bird species results for random forest classifier—true positives, false positives, precision and recall

Species class	TP rate	FP rate	Precision	Recall
White-throated Thrush	1	0	1	1
American Black Duck	0.5	0	1	0.667
Black Vulture	0.923	0.003	0.857	0.889
Jabiru	0.75	0.001	0.857	0.8
Wild Turkey	0.8	0.007	0.706	0.75
Brown Thrasher	0.606	0.015	0.667	0.635
Northern Lapwing	0.953	0.033	0.744	0.836
American Robin	0.447	0.017	0.607	0.515
Common Myna	0.481	0.013	0.591	0.531
Common Sandpiper	1	0	1	1
Aberts Towhee	1	0	1	1
Barred Owl	1	0	1	1
Gyrfalcon	0	0.003	0	0
Indian Peafowl	0.667	0.004	0.727	0.696
Asian Brown Flycatcher	0.5	0.006	0.6	0.545
Great Kiskadee	1	0	1	1
Himalayan Snowcock	0	0	0	0
Ferruginous Hawk	0.333	0.001	0.5	0.4
Common Snipe	1	0	1	1
Snowy Owl	0.75	0.004	0.5	0.6
Dark-sided Flycatcher	0	0	0	0
Wood Stork	0.75	0.001	0.857	0.8
Eastern Bluebird	1	0.003	0.952	0.976
Bald Eagle	0.947	0.003	0.9	0.923
Oriental Cuckoo	0.931	0.007	0.844	0.885
Anhinga	0.667	0.007	0.615	0.64
Black-and-white Warbler	0.708	0.012	0.68	0.694
Annas Hummingbird	0.611	0.012	0.579	0.595
Fan-tailed Warbler	0.818	0	1	0.9
Nelsons Sparrow	0.625	0.004	0.625	0.625
Roseate Spoonbill	0.75	0.004	0.75	0.75
Bananaquit	1	0.006	0.778	0.875

shallow trees where each one is constructed by considering 4 random features is used. Other classifiers show lesser classification performance. The results are shown in the Table 8.

Table 9 gives an understanding of the results of bird species classification using random forest classifier. The confusion matrix has been omitted due to its size(33×33 matrix) and instead the analysis is presented using the concepts of precision and recall. In the context of bird species classification, precision is the probability that given bird species class, its recording is correctly classified. On the other hand, given a classification of a recording, recall is the probability that the classification is correct. Equations 6 and 7 define precision and recall. TP, FP and FN rate are the fraction of true positives, false positives, and false negatives respectively for each species class.

$$Precision = \frac{TP\ Rate}{TP\ Rate + FP\ Rate} \tag{6}$$

$$Recall = \frac{TP\ Rate}{TP\ Rate + FN\ Rate} \tag{7}$$

Low precision/recall measure for a bird species class indicates difficulty in correctly classifying the species. Particularly in cases of bird species Ferruginous Hawk, Snowy Owl, Common Myna and Anna's Hummingbird the precision and recall measure are significantly lower. This indicates weak training models developed for them because of lack of robust training instances/features. Also a higher false positive rate can be observed among the small birds like American Robin, Northern Lapwing, Brown Thrasher etc. It means that these birds are versatile with respect to the features extracted and are often the result of mis-classification of other species classes. The above observation again points towards the need for a robust feature set and diverse set of training examples for those bird species.

5 Conclusion

In this paper we presented a method of categorizing real time bird sound recordings. We have compared several approaches for improving the performance of the classification tasks. It can be established that there is a direct correlation between bird size and the pitch of the sound it produces for the birds present in the data set. Hence bird size can be predicted based on pitch and energy based features. Bird vocalization can generally be divided into two categories, namely call and song. The proposed method uses pitch-based, energy-based and duration-based features to perform call versus song classification. Bird call can be further categorized based on pitch, energy, duration and tempo features. The bird habitat can be predicted efficiently by using prosodic and MFCC features together.

Performance of the proposed bird identification system still leaves scope for improvement. Based on the results of this study, birds species may be identified with an accuracy

of around 83 %. However variety of sounds of other species demand further analysis. To make the results more convincing and the approach generalized, one has to collect more data and find more efficient and robust classification techniques and features to improve classification performance in the near future. For future works, this approach can be easily scaled up to large number of bird species provided the availability of audio clips for them. Useful directions for further research include processing audio recordings in noisy environments with multiple bird species simultaneously and, extending this study to other important animal species such as mammals which can aid in understanding complex and sensitive ecosystems.

References

Acevedo, A., Corrada-Bravo, C., Corrada-Bravo, H., Villanueva-Rivera, L., & Aide, T. (2009). Automated classification of bird and amphibian calls using machine learning: A comparison of methods. *Ecological Informatics*, *4*, 206–214.

Bardeli, R., Wolff, D., Kurth, F., Koch, M., Tauchert, K., & Frommolt, K. (2010). Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring. *Pattern Recognition Letters*, *31*, 1524–1534.

Beckers, G. J. (2011). Bird speech perception and vocal production: A comparison with humans. *Human Biology*, *83*(2), 191–212.

Bermúdez-Cuamatzin, E., Ríos-Chelén, A. A., Gil, D., & Garcia, C. M. (2010). Experimental evidence for real-time song frequency shift in response to urban noise in a passerine bird. *Biology Letters*, *3*, 368–370.

Bolhuis, J. J., Okanoya, K., & Scharff, C. (2010). Twitter evolution: Converging mechanisms in birdsong and human speech. *Nature Reviews Neuroscience*, *11*(11), 747–759.

Brandes, T. S. (2008). Automated sound recording and analysis techniques for bird surveys and conservation. *Bird Conservation International*, *18*(S1), S163–S173.

Briggs, F., Lakshminarayanan, B., Neal, L., Fern, X. Z., Raich, R., Hadley, S., et al. (2012). Classification of multiple bird species. *Journal of Acoustic Society of America*, *131*, 4640–4650.

Chen, Z., & Maher, R. C. (2006). Semi-automatic classification of bird vocalizations using spectral peak tracks. *The Journal of the Acoustical Society of America*, *120*, 2974–2984.

Clark, G. A. (1979). Body weights of birds: A review. *The Condor*, *81*(2), 193–202.

Davis, S. B., & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences, In *Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing* (Vol. 28, pp. 357–366).

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience*, *22*(1), 567–631.

Dowling, J., Luther, D., & Marra, P. (2012). Comparative effects of urban development and anthropogenic noise on bird songs. *Behavioral Ecology*, *23*(1), 201–209.

Dunning, J. (2013). Updates to the second edition of the CRC handbook of avian body masses. https://ag.purdue.edu/fnr/documents/BodyMassesBirds.pdf.

Fagerlund, S. (2007). Bird species recognition using support vector machines. *Journal on Advances in Signal Processing*, *7*, 64–71.

Hall, M. L., Kingma, S. A., & Peters, A. (2013). Male songbird indicates body size with low-pitched advertising songs. *PLoS One*, *8*(2), e56717.

Juang, C., & Chen, T. (2007). Birdsong recognition using prediction-based recurrent neural fuzzy networks. *Neurocomputing*, *71*, 121–130.

Kight, C. R., & Swaddle, J. P. (2011). How and why environmental noise impacts animals: An integrative, mechanistic review. *Ecology Letters*, *14*(10), 1052–1061.

Kwan, C., Mei, G., Zhao, X., Ren, Z., Xu, R., Stanford, V., Rochet, C., Aube, J., & Ho, K. (2004). Bird classification algorithms: Theory and experimental results, In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'04)* (vol. 5, pp. 289–292), Montreal, Canada.

Laiolo, P. (2010). The emerging significance of bioacoustics in animal species conservation. *Biological Conservation*, *143*(7), 1635–1645.

Lartillot, O., & Toiviainen, P. (2007). A matlab toolbox for musical feature extraction from audio, In *International Conference on Digital Audio Effects* (pp. 237–244).

Lartillot, O., Eerola, T., Toiviainen, P., & Fornari, J. (2008). Multi-feature modeling of pulse clarity: Design, validation and optimization., In *ISMIR* (pp. 521–526), Citeseer.

Lathi, B. P. (2004). *Signal processing and linear systems*. Oxford: Oxford University Press.

Lee, C.-H., Han, C.-C., & Chuang, C.-C. (2008). Automatic classification of bird species from their sounds using two-dimensional cepstral coefficients. *IEEE Transactions on Audio, Speech, and Language Processing*, *16*(8), 1541–1550.

Linhart, P., & Fuchs, R. (2015). Song pitch indicates body size and correlates with males' response to playback in a songbird. *Animal Behaviour*, *103*, 91–98.

Lopes, M. T., Gioppo, L. L., Higushi, T. T., Kaestner, C. A. A., Silla, Jr., C. N., & Koerich, A. L. (2011). Automatic bird species identification for large number of species, In *IEEE International Symposium on Multimedia*.

Lopes, M. T., Koerich, A. L., Kaestner, C. A. A., Silla, Jr., C. N. (2011). Feature set comparison for automatic bird species identification, In *IEEE International Conference on Systems, Man, and Cybernetics, Anchorage, Alaska*.

Luther, D., & Baptista, L. (2010). Urban noise and the cultural evolution of bird songs. *Proceedings of the Royal Society of London B: Biological Sciences*, *277*(1680), 469–473.

Mellinger, D., & Bradbury, J. W. (2007). Acoustic measurement of marine mammal sounds in noisy environments, In *Proceedings of the International Conference on Underwater Acoustical Measurements: Technologies and Results*.

Mitchell, T. M. (1997). *Machine learning*. Maidenhead: McGraw-Hill.

Rickwood, P., & Taylor, A. (2008). Methods for automatically analyzing humpback song units. *Journal of the Acoustical Society of America*, *123*, 1763–1772.

Silla, C. N., & Kaestner, C. A. (2013). *Hierarchical classification of bird species using their audio recorded songs* (pp. 1895–1900). Washington, DC: IEEE Computer Society.

Slabbekoorn, H., & Peet, M. (2003). Ecology: Birds sing at a higher pitch in urban noise. *Nature*, *424*(6946), 267–267.

Somervuo, P., Harma, A., & Fagerlund, S. (2006). Parametric representations of bird sounds for automatic species recognition. *IEEE Transactions on Audio, Speech and Language Processing*, *14*, 2252–2263.

Sun, R., Marye, Y. W., & Zhao, H. (2013). Wavelet transform digital sound processing to identify wild bird species, In *Proceedings of the 2013 International Conference on Wavelet Analysis and Pattern Recognition*.

Tsai, W.-H., Xu, Y.-Y., & Lin, W.-C. (2013). Bird species identification based on timbre and pitch features, In *IEEE International Conference on Multimedia and Expo* (pp. 1–6).

Vilches, E., Escobar, I., Vallejo, E., & Taylor, C. (2006). Data mining applied to acoustic bird species recognition, In *Proceedings of the 18th IEEE International Conference on Pattern Recognition (ICPR'06)*.

Witten, I. H., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques*. San Francisco: Morgan Kaufmann Publishers.