



Project Title

Optional Subtitle

Candidate Number: YPDR0¹

MSc Data Science & Machine Learning

Supervisor: Dr. Ifat Yasin

Submission date: 11 September 2023

¹**Disclaimer:** This report is submitted as part requirement for the MSc Data Science & Machine Learning at UCL. It is substantially the result of my own work except where explicitly indicated in the text. The report may be freely copied and distributed provided the source is explicitly acknowledged

Abstract

Summarise your report concisely.

Contents

1	Introduction	2
2	Background & related work	4
2.1	Evaluation metrics	4
2.2	Feature extraction	5
2.2.1	Human inspired features	5
2.2.2	Feature stacking	5
2.3	Classification	5
2.3.1	KNN	5
2.3.2	Decision trees	6
2.3.3	Gaussian Mixture Models	6
2.3.4	Hidden Markov Models	6
2.3.5	SVMs	6
2.3.6	Neural Networks	6
3	Methodology	7
4	Extensions of methodology	8
5	Conclusion	9

Chapter 1

Introduction

Birds play a crucial role in ecosystems around the world. They form an important link in the food chain, pollinate plants, and even plant trees [2]. The quantity and diversity of birds observed in an area can therefore be seen as a key indicator of the strength of its ecosystem.

Aside from being important ecological agents, birds also colour our lives with their sights, sounds and behaviours. The community of birdwatchers, known colloquially in the U.K. as ‘twitchers’, has grown considerably over the past few years. The Royal Society for the Protection of Birds (RSPB) reported a 70% increase in their website views over the first lockdown, with more than 50% of those views on pages looking at bird identification. The Bird Bird Garden Watch, an annual event which encourages people to note bird sightings in their own residences, brought in 1 million participants in January 2021, more than double the previous year’s tally. As such there is a growing commercial demand for accurate and easily accessible bird identification tools.

Birds are typically shy and protective creatures and tend to reside out of harm’s way in shrubs, trees and nests, and therefore are usually heard but not seen. The majority of their communication is done through distinctive vocalizations, known as birdsong, that are usually unique to an individual species. The consequence of this is that birds are typically identified through their birdsong rather than their visual sighting. Birdsong used for identification is usually recorded on microphones that may be running for a long time and/or located in a place that isn’t necessarily optimum to capture the vocalization. The recordings may be corrupted from a wide range of sources, such as ambient background noise, large changes in birdsong amplitude, long periods of silence, and vocalizations from other birds. Recordings captured by long running microphones may be several hours in

duration so it would be impractical to have a human expert identify the birds manually. There is therefore a need for robust birdsong identification tools that can handle these corruptions and that require minimal human intervention in order to classify unknown birdsong.

As with most machine learning classification problems, the work boils down to two key components. The first is feature extraction, that is, given an input signal usually in the form of an amplitude varying over time, how can it be transformed to a vector or matrix representation that captures the key information of the signal. The second is classification, i.e. given this representation in the feature space, how can it be used to train a model that can then perform classification on unseen samples of birdsong. There are further steps that can be taken to improve this process, such as pre-processing the signal in order to remove or reduce the influence of background noise [6].

There has been numerous research into audio classification problems both in birdsong and in other types of audio, such as speaker identification from human speech [4]. However, there remain novel methods that have been tested and have been shown to have good results in wider audio classification problems that have yet to be tried out in birdsong identification problems. These methods include both feature extraction and classification. The broad aim of this thesis is to attempt to evaluate these methods. In more detail, the aims of this thesis can be summarised as follows:

1. Explore the birdsong classification performance of feature representations shown to have promising results for non-birdsong related audio classification problems using simple classification models such as Support Vector Machines (SVM). The hypothesis is that feature representations shown to have good results in audio classification problems will have good results for birdsong identification.
2. Explore some of the hyperparameters available during the feature extraction process. The hypothesis is that using hyperparameter values more suited to birdsong classification problems will yield a higher classification accuracy.
3. Explore the performance of deep learning architectures such as Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN) and compare the results with simpler statistical models. The hypothesis is that more complex and flexible architectures such as these will have superior performance when compared to simpler statistical models, such as SVMs.

Chapter 2

Background & related work

Describe here work that is connected to your thesis. This should include references to published work. There is no fixed rule, but I would expect a student to have read around 50 published research papers and reference them in a thesis.

2.1 Evaluation metrics

Since birdsong identification can be thought of as a classification problem, the evaluation metric typically used is the classification accuracy percentage ([7],). This is calculated as the percentage of correctly identified samples from a set of labelled samples previously unseen by the model.

Acevedo et al [1] used true positive (TP) and false positive (FP) rates to evaluate their models used to classify bird and amphibian calls. This evaluation method is sometimes preferred when analysing long recordings which may include several different species to be classified. TP gives an indication as to how well the model correctly identifies species present in the recording. FP indicates how often the model erroneously identifies species absent in the recording. A high performing model will have high values for TP and low values for FP .

Potamitis et al [6] used an alternative form of evaluation presented in terms of precision (P) and recall (R) which are defined as

$$P = \frac{TP}{TP + FP}, \quad R = \frac{TP}{TP + FN} \quad (2.1)$$

where FN is the number of false negatives. Informally, P and R can be thought of as

$$P = \frac{\text{relevant retrieved instances}}{\text{all **retrieved** instances}}, \quad R = \frac{\text{relevant retrieved instances}}{\text{all **relevant** instances}} \quad (2.2)$$

It's well known that there is usually a trade-off between P and R , so usually the F -score is reported along with the P and R metrics. The F -score is defined as

$$F = \frac{(1 + \beta^2)PR}{\beta^2 P + R} \quad (2.3)$$

for some $\beta \in \mathbb{R}$.

For binary classification problems, i.e. classifying a sample as one of two classes (bird species), the Area Under Curve (AUC) metric is often preferred [5].

2.2 Feature extraction

- Description of birdsong and why it exists
- Difference between calls and songs, with example sonograph.
- Include a discussion on segmentation?

2.2.1 Human inspired features

Research has shown that birdsong and human language share many similarities in terms of the neural mechanisms employed to form the language/song, the impact of social contact in learning the language/song, and so on [3]. Given that the human auditory system has evolved over thousands of years to best process human speech, it seems reasonable to suggest that using features based on the human auditory system may be effective when it comes to birdsong.

2.2.2 Feature stacking

2.3 Classification

2.3.1 KNN

KNN used by [7]

2.3.2 Decision trees

Description of decision trees. Used by [1].

2.3.3 Gaussian Mixture Models

2.3.4 Hidden Markov Models

2.3.5 SVMs

2.3.6 Neural Networks

Feedforward Neural Network

Convolutional Neural Network

Recurrent Neural Network

Chapter 3

Methodology

Describe your method in detail and with great clarity, distinguishing it from other works (if it is indeed a novel idea). It is very important to clearly motivate your method.

Describe the results of your method here in this chapter.

Chapter 4

Extensions of methodology

It is unlikely that everything you tried worked well, so in this chapter you may wish to describe a modified version of your method and the associated results. Explain why you were motivated to try this extension and how you think it might help to address some of the shortcomings of the method in Chapter 3.

Chapter 5

Conclusion

Summarise what you have achieved and evaluate honestly if you feel the approach has been largely successful. Explain what could be improved still and perhaps why the method is not working well (if that is the case).

Bibliography

- [1] Miguel A Acevedo, Carlos J Corrada-Bravo, Héctor Corrada-Bravo, Luis J Villanueva-Rivera, and T Mitchell Aide. Automated classification of bird and amphibian calls using machine learning: A comparison of methods. *Ecological Informatics*, 4(4):206–214, 2009.
- [2] Richard K Broughton, James M Bullock, Charles George, Ross A Hill, Shelley A Hinsley, Marta Maziarz, Markus Melin, J Owen Mountford, Tim H Sparks, and Richard F Pywell. Long-term woodland restoration on lowland farmland through passive rewilding. *PloS one*, 16(6):e0252466, 2021.
- [3] Allison J. Doupe and Patricia K. Kuhl. Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience*, 22(1):567–631, 1999. PMID: 10202549.
- [4] Mineichi Kudo, Jun Toyama, and Masaru Shimbo. Multidimensional curve classification using passing-through regions. *Pattern Recognition Letters*, 20(11):1103–1111, 1999.
- [5] Yi Ren Leng and Huy Dat Tran. Multi-label bird classification using an ensemble classifier with simple features. In *Signal and information processing association annual summit and conference (APSIPA), 2014 Asia-Pacific*, pages 1–5. IEEE, 2014.
- [6] Ilyas Potamitis, Stavros Ntalampiras, Olaf Jahn, and Klaus Riede. Automatic bird sound detection in long real-field recordings: Applications and tools. *Applied Acoustics*, 80:1–9, 2014.
- [7] Murugiaya Ramashini, Pg Emeroylariffion Abas, Ulmar Grafe, and Liyanage C De Silva. Bird sounds classification using linear discriminant analysis. In *2019 4th International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE)*, pages 1–6. IEEE, 2019.