Classifier Technical Specifications: European Nightjar

1. Introduction

The BTO Acoustic Pipeline's **European Nightjar** classifier (version 1) detects *churr* and *gro-eek* calls of European Nightjar *Caprimulgus europaeus*. This document provides technical notes on the derivation of the classifier, its use and how to interpret results. The rest of this document is arranged into five parts:

- Training data sample sizes
- Precision and Recall statistics on withheld data
- False positive rates on independent data
- Known issues

2. Training data sample sizes

This classifier is trained on strongly labelled audio recordings compiled by BTO and collaborators, with some additional recordings from xeno-canto. We are grateful to our collaborators and the sound recordists who share recordings via xeno-canto. The following table gives the number of audio samples used to train and evaluate the classifier. The Background class encompasses ambient and anthropogenic sounds plus non-target wildlife (i.e. other bird species, mammals, amphibians and insects).

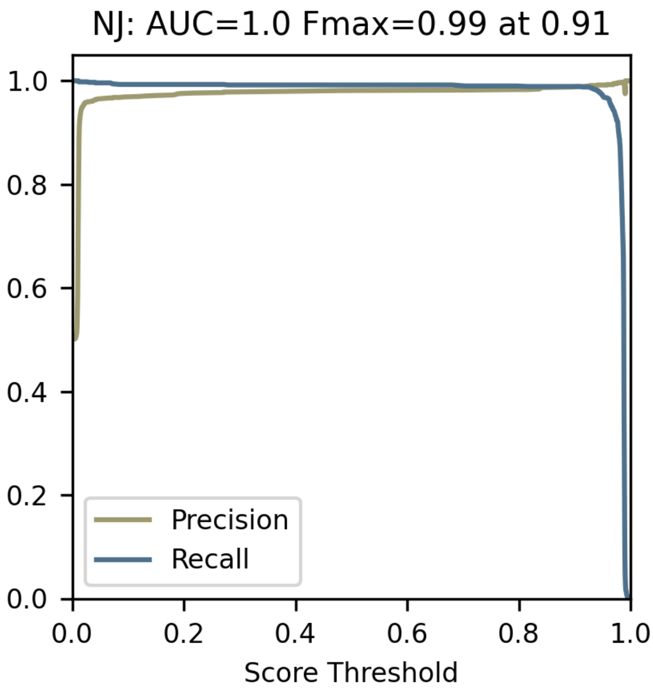
Class	Train sample size	Evaluation sample size
Background	5000	1000
Nightjar	5000	843

3. Precision and Recall

Classifier performance is typically evaluated using the metrics *Precision* and *Recall*:

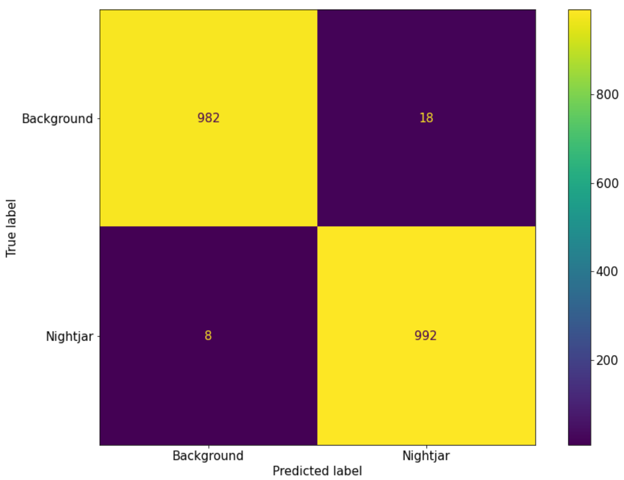
- **Precision** is the % of detections returned by a classifier that truly are of that species. If the classifier thinks 100 clips contain Nightjar calls, Precision is the percentage of these that actually do contain Nightjar calls.
- **Recall** is the percentage of true instances that are detected. For example, if 100 clips contain Nightjar calls, Recall is the percentage of these that are found by the classifier.

These metrics are *threshold dependent*. If we say that all classifier scores greater than 0.5 constitute a detection we will get different Precision and Recall values than if we use a more stringent score threshold of 0.9. Figure 1 shows how Precision and Recall vary with score threshold for this classifier. This figure is based on application of the classifier to withheld training data. The subsequent table provides Precision and Recall values for commonly used score thresholds. Additionally, *Best T* gives the threshold at which the F-score is maximised, which is a way of optimising both Precision and Recall. The Precision and Recall statistics for that threshold (*P|R Best*) are also shown. Detections exported from the Acoustic Pipeline by default use this best threshold.



Species	Scientific	P|R 0.50	P|R 0.90	P|R 0.95	P|R 0.99	Best T	P|R Best
Nightjar	Caprimulgus europaeus	98.2 | 99.2	98.9 | 98.9	99.2 | 97.1	98.9 | 9.1	0.91	99.0 | 99.0

As an illustration, the figure below shows the confusion matrix (number of true and false positives and negatives by class) for a sample of independent data, using a score threshold of 0.5.



4. False positive rates

The false positive rate indicates how often the classifier suggests a species is present when it is not. For this test we use a benchmark dataset of 21,000 ambient sound clips that have been manually checked to confirm they contain no bird records. We run the classifier against this dataset and summarise the percentage of clips that are falsely assigned to a species. As for performance metrics, this measure is threshold dependent, with typically fewer errors when a more stringent score threshold is applied. False positive rates for Nightjar are shown below. Note that this is a simple test: in reality false positive rates may be higher in natural soundscapes, for example, where a distant call of one species is mistaken for another species.

Species	Scientific	0.50	0.90	0.95	0.99	Best
Nightjar	Caprimulgus europaeus	0.537	0.176	0.119	0.000	0.166

5. Known issues

Performance of this classifier is generally high. Scores returned by the classifier reflect the likelihood that a Nightjar is present, but also convey some component of distance, with distant Nightjar song generally having a lower score. This phenomenon has been shown for related nightjar species (e.g. Common Nighthawk, Yip et al. 2020 https://doi.org/10.1002/rse2.118).