



A machine learning framework to classify Southeast Asian echolocating bats

Natalie Yoh^{a,*}, Tigga Kingston^{b,c}, Ellen McArthur^d, Oliver E. Aylen^e, Joe Chun-Chia Huang^f, Emy Ritta Jinggong^d, Faisal Ali Anwarali Khan^d, Benjamin P.Y.H. Lee^g, Simon L. Mitchell^a, Jake E. Bicknell^a, Matthew J. Struebig^a

^a Durrell Institute of Conservation and Ecology (DICE), School of Anthropology and Conservation, University of Kent, Canterbury, United Kingdom

^b Department of Biological Sciences, Texas Tech University, Lubbock, TX, USA

^c Southeast Asian Bat Conservation Research Unit, Lubbock City, TX, USA

^d Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia

^e University of Otago, Department of Zoology, Dunedin, Otago, New Zealand

^f Taiwan Forestry Research Institute, Taipei, Taiwan

^g Wildlife Management Division, National Parks Board, Singapore

ARTICLE INFO

Keywords:

Acoustic monitoring
Chiroptera
Echolocation
Southeast Asia
Machine learning
Supervised algorithm

ABSTRACT

Bats comprise a quarter of all mammal species, provide key ecosystem services and serve as effective bio-indicators. Automated methods for classifying echolocation calls of free-flying bats are useful for monitoring but are not widely used in the tropics. This is particularly problematic in Southeast Asia, which supports more than 388 bat species. Here, sparse reference call databases and significant overlap among species call characteristics makes the development of automated processing methods complex. To address this, we outline a semi-automated framework for classifying bat calls in Southeast Asia and demonstrate how this can reliably speed up manual data processing. We implemented the framework to develop a classifier for the bats of Borneo and tested this at a landscape in Sabah. Borneo has a relatively well-described bat fauna, including reference calls for 52% of all 81 known echolocating species on the island. We applied machine learning to classify calls into one of four call types that serve as indicators of dominant ecological ensembles: frequency-modulated (FM; forest-specialists), constant frequency (CF; forest-specialists and edge/gap foragers), quasi-constant frequency (QCF; edge/gap foragers), and frequency-modulated quasi constant frequency (FMqCF; edge/gap and open-space foragers) calls. Where possible, we further identified calls to species/sonotype. Each classification is provided with a confidence value and a recommended threshold for manual verification. Of the 245,991 calls recorded in our test landscape, 85% were correctly identified to call type and only 10% needed manual verification for three of the call types. The classifier was most successful at classifying CF calls, reducing the volume of calls to be manually verified by over 95% for three common species. The most difficult bats to classify were those with FMqCF calls, with only a 52% reduction in files. Our framework allows users to rapidly filter acoustic files for common species and isolate files of interest, cutting the total volume of data to be processed by 86%. This provides an alternative method where species-specific classifiers are not yet feasible and enables researchers to expand non-invasive monitoring of bat species. Notably, this approach incorporates aerial insectivorous ensembles that are regularly absent from field datasets despite being important components of the bat community, thus improving our capacity to monitor bats remotely in tropical landscapes.

1. Introduction

Biodiversity monitoring is critical to informing conservation practice. Still, multi-taxon assessments are frequently constrained by

resources, time, and survey bias (Gardner et al., 2008). Focusing survey efforts on biological indicators is one way to ameliorate these challenges so long as these species or groups reflect the needs of others in the system, particularly in the way they respond to environmental change

* Corresponding author.

E-mail address: tallyyoh@gmail.com (N. Yoh).

<https://doi.org/10.1016/j.ecolind.2022.108696>

Received 4 September 2021; Received in revised form 11 February 2022; Accepted 14 February 2022

Available online 21 February 2022

1470-160X/© 2022 The Authors.

Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).