Ecological Informatics 89 (2025) 103172

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Advancing invasive species monitoring: A free tool for detecting invasive cane toads using continental-scale data

Franco Ka Wah Leung *, Lin Schwarzkopf, Slade Allen-Ankins

College of Science and Engineering, James Cook University, Townsville 4814, Queensland, Australia

A R T I C L E I N F O

Keywords:
Australian acoustic observatory
BirdNET
Cane toad
Invasive species
Machine learning
Passive acoustic monitoring

A B S T R A C T

Invasive species pose a significant threat to global biodiversity and ecosystem health, necessitating effective monitoring tools for early detection and management. Here, we present the development and assessment of a user-friendly and transferable monitoring tool for the invasive cane toad (*Rhinella marina*) using passive acoustic monitoring (PAM) and machine learning algorithms. Leveraging a continental-scale PAM dataset (Australian Acoustic Observatory), we trained a cane toad classifier using the BirdNET algorithm, a convolutional neural network architecture capable of identifying acoustic events. We validated thousands of BirdNET predictions across Australia, and our classifier achieved over 90 % accuracy even at many sites outside the areas from which the training data were obtained. Additionally, because cane toads typically call for long periods, we significantly enhanced detection accuracy by incorporating contextual information from time-series data, essentially checking if other calls occurred around each detection (an optimized threshold approach using conditional inference trees). This method substantially reduced false positives and improved overall performance in cane toad detection at sites across Australia. Overall, our method will allow others to develop accurate and precise automated acoustic monitoring tools tailored to their situation, with minimal training data, addressing the critical need for accessible solutions in biodiversity monitoring, control of invasive species and conservation.

1. Introduction

Biological invasions present a significant threat to biodiversity, ecosystem structure and function, human health, and the global economy (Bradshaw et al., 2016; Ehrenfeld, 2010). Effective conservation and management are enhanced by monitoring invasive species, enabling early detection of new incursions, assessment of population trends, and implementation of timely control measures to mitigate adverse impacts on native biodiversity and ecosystems (Pyšek et al., 2020; Vander Zanden et al., 2010).

Passive acoustic monitoring (PAM) uses acoustic recorders to capture environmental sounds, including those of invasive species, making it a valuable tool for tracking and managing biological invasions (Ribeiro Jr et al., 2022). This non-invasive monitoring method uses analysis of recorded audio to detect species (Blumstein et al., 2011; Gibb et al., 2019). Studies indicate that PAM is comparable in effectiveness to traditional observer-based monitoring to assess diversity of many terrestrial vertebrate species (e.g., Hoefer et al., 2023; Melo et al., 2021). In addition, PAM enables continuous monitoring, a highly cost-effective approach to invasive species detection, particularly in remote areas and rugged environments (Hu et al., 2009).

PAM generates a very large amount of data, which often cannot be analyzed using human listening (Villanueva-Rivera and Pijanowski, 2012), and thus automated approaches become essential for its application. There is, however, a scarcity of publicly accessible, broadly applicable, and easy-to-use sound detection and classification tools (Pérez-Granados et al., 2023; Wood et al., 2023a). Machine learning algorithms, especially deep convolutional neural networks (CNNs), have shown great promise in automating sound identification for invasive species, enabling faster and more accurate analysis of large datasets (Kahl et al., 2021; Jeantet and Dufourq, 2023). Although time-consuming to develop (Knight et al., 2017), these tools can significantly expand the opportunities for monitoring over large spatial scales and extended time periods, enhancing early detection and management (Amorim et al., 2023).

Cane toads (*Rhinella marina*) invaded northern Australia in 1935 (Easteal, 1981), and expanded across Queensland, New South Wales, the Northern Territory, and Western Australia (Atlas of Living Australia,

* Corresponding author.
E-mail addresses: franco.leung@my.jcu.edu.au (F.K.W. Leung), lin.schwarzkopf@jcu.edu.au (L. Schwarzkopf), slade.allenankins@jcu.edu.au (S. Allen-Ankins).

https://doi.org/10.1016/j.ecoinf.2025.103172
Received 2 October 2024; Received in revised form 14 February 2025; Accepted 27 April 2025
Available online 28 April 2025

2024). Recognized as one of the world's top 100 worst invasive species (Lowe et al., 2000), cane toads spread rapidly (Shine et al., 2021) and have profound negative impacts on Australian native biodiversity (Shine, 2010). Notably, anurans, including cane toads, rely heavily on vocalizations for courtship, making them ideal candidates for detection via passive acoustic monitoring (PAM) (e.g., Pérez-Granados et al., 2023; Wood et al., 2023a).

Although cane toads are a highly vocal invasive species, there are no publicly accessible, user-friendly acoustic classifiers tailored specifically for their detection. Despite efforts dating back to the 1990s to integrate PAM with automated detection algorithms for cane toads (Hu et al., 2009; Taylor et al., 1996, 2017), these tools have not been widely adopted. One barrier to adoption is a need for sensor and algorithm flexibility (Roe et al., 2018). For instance, the original algorithms and software were designed and applied within custom-built acoustic sensor networks (e.g., Taylor et al., 1996, 2017), limiting their broad applicability. In addition, variation among background noises and intra-specific call variation can cause variation in the success of automated species recognition in different locations (Cole et al., 2022; Lauha et al., 2022; Metcalf et al., 2022). Cane toads range across a very wide area, and their advertisement calls vary among Australian populations (Muller et al., 2016; Yasumiba et al., 2016). Consequently, automated detection tools developed for specific regions or study areas may not work well in other areas. In light of these challenges, there is a critical need for innovative solutions to cane toad detection over wide areas.

In this study, we introduce a free and user-friendly cane toad acoustic classifier designed for analyzing broad spatio-temporal datasets across Australia. By utilizing a machine-learning algorithm and an extensive audio dataset, we aimed to create a freely accessible, flexible cane toad classifier with high detection accuracy with the ability to accommodate regional variability in call characteristics, thereby facilitating repeatable and cost-efficient monitoring efforts for the invasive toads across Australia. Given the widespread nature of cane toad populations, we incorporated data post-processing techniques to optimize detection performance over a wide area.

2. Materials and methods

2.1. BirdNET

We chose the BirdNET algorithm, a CNN architecture designed to identify acoustic events by analyzing visual patterns in spectrograms (Kahl et al., 2021), to develop our cane toad acoustic recognizer. BirdNET is a freely available, pretrained algorithm can be used to train a custom classifier that can detect species outside its original training data. It reliably identifies over 6000 species worldwide (https://github.com/kahst/BirdNET-Analyzer), including more than 40 frog species (Pérez-Granados et al., 2023; Wood et al., 2023a), with proven effectiveness across various recording conditions (Kahl et al., 2021; Manzano-Rubio et al., 2022). Here, we want to highlight the user-friendliness of BirdNET software's graphical user interface (GUI), which simplifies audio analysis and recognizer training without necessitating advanced programming expertise, thus facilitating its use by a wide range of stakeholders. Additionally, custom trained BirdNET recognizers can be readily shared as a portable TensorFlow Lite file, ensuring easy access, compatibility across different sensor networks, recording equipment, and study designs, and enabling widespread adoption across various parties.

2.2. Audio dataset

To develop and train our cane toad recognizer, we used audio from the Australian Acoustic Observatory (A2O), a continent-wide acoustic sensor network, consisting of 62 active sites, covering seven major ecoregions in Australia. Each site is equipped with four acoustic recording units (ARUs) placed in various habitats, ranging from rainforests to arid landscapes, providing real-world sound data (see Roe et al., 2021 for full details). We selected all A2O sites in Queensland, along with specific sites in New South Wales, the Northern Territory, Western Australia and Victoria (Fig. S1), and used audio recordings from 18:00 to 06:00, as adult cane toads in Australia are generally nocturnal (Doody et al., 2019). These sites were characterized by multiple occurrence records of cane toads in the Atlas of Living Australia (2024), and the acquired audio data from these A2O sites served as the basis for model training, testing, and performance evaluation.

2.3. Training data generation

Cane toads typically produce advertisement calls lasting, on average, eight seconds, with a dominant frequency ranging from 500 to 600 Hz and a call frequency spanning 110–1180 Hz (Bleach et al., 2015; Muller et al., 2016). Creating a classifier using BirdNET involves three key steps: generating training data that accurately represents the target species' acoustic characteristics, training the classifier with the preprocessed training dataset, and evaluating the classifier using a testing dataset (Fig. 1). While this workflow follows the tools and methods used in our study, each step can be adapted using alternative software, codes, or algorithms, depending on the specific research needs and available resources.

In this workflow, training data was generated using the *monitoR* R package (Hafner and Katz, 2018), a tool for template-based acoustic detection. This step involved constructing a binary acoustic template based on previously recorded cane toad calls (Fig. S2). *monitoR* was selected for its compatibility with our workflow; however, alternative signal processing methods can yield similar results, such as using feature embeddings (Allen-Ankins et al., 2025). According to BirdNET's developers, the algorithm's performance plateaus after 3500 samples per class (Kahl et al., 2021). To account for regional variations in detecting toad calls, particularly in the absence of specific guidelines on training data allocation per location for individual classes or species, we ensured a minimum of 3500 training samples in total, with an average sample count close to this threshold per recording site. Because the BirdNET algorithm processes only three-second audio snippets, our template was three seconds long (Fig. S2). Employing a single template is effective at identifying toad calls for generating training data, achieving reasonable performance even for species with geographical variation and highly variable harmonics in their calls (Balantic and Donovan, 2020; Katz et al., 2016). While a template can detect the target species, it also produces a high number of false positives, making it insufficient as a standalone classifier. However, these false positives can be valuable for training the classifier. Therefore, we created a single binary-point template with the *makeBinTemplate* function, using a typical cane toad call (Fig. S2). The binary-point template maps signal ('on' points) and non-signal ('off' points) areas within a spectrogram, disregarding other values. We set the amplitude cut-off to 'interactive selection' and used a rectangle around the call to identify potential 'on' points during template creation (see Hafner and Katz, 2018 for full details).

We then conducted template detection using the *binMatch* function from the *monitoR* package on audio data (18:00–06:00) from the wet site ARUs at eight study sites, totaling 21,491 h (Fig. 1, S1). This analysis was carried out using R version 4.2.3 (R Core Team, 2023) and leveraged the James Cook University High-Performance Computing (JCU HPC) resources. However, this step can also be processed using a standard computer, depending on the available computing resources and the amount of audio recordings one needs to analyze. Due to the substantial number of predictions generated from the template analysis, we focused our validation efforts on two-hour recordings with over 200 predictions and checked the top-scoring 100 predictions for each month from each ARU. This labeled cane toad audio spanned a latitude range of −13.16 to −26.50 and encompassed all months and hours (Fig. S1; Table 1). Consequently, our training dataset consisted of both true positives (cane toad calls) and highly repeated false-positive sounds (Table 1). To

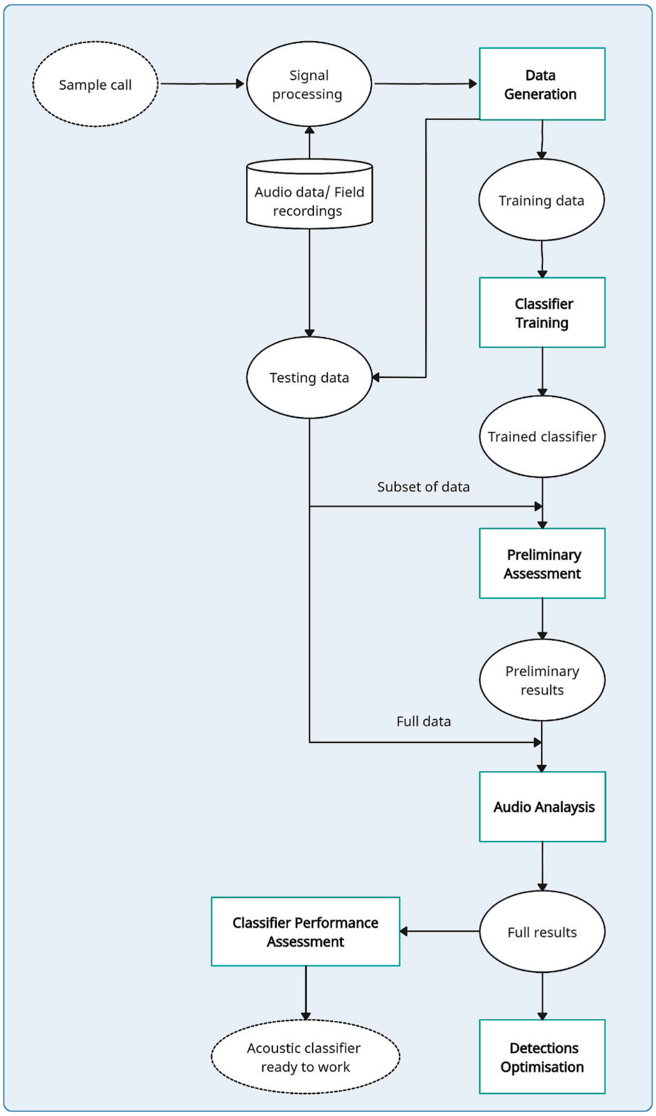

Fig. 1. Workflow for developing the machine-learning acoustic classifier for cane toad detection. The process begins with signal processing, where sample calls and field recordings are processed for Data Generation to produce training and testing data, which the previous one is then used in Classifier Training to develop a custom classifier. The trained classifier undergoes Preliminary Assessment using a subset of testing data, followed by Audio Analysis on the full testing dataset. Performance is further evaluated in Classifier Performance Assessment, where classifier performance is evaluated before potential deployment in broader studies. The final step, Detections Optimization, is optional and involves refining detection outputs to further enhance accuracy based on study-specific requirements.

prepare the training dataset for use with BirdNET, we segmented all labeled audio into three-second audio segments using the *AudioSegment* module (Hu and Wang, 2007) in Python version 3.11.6 (Van Rossum and Drake Jr, 1995).

2.4. Classifier development

Before training the model, we added noise and background sounds identified during template detection, labeling them as 'Background', and included any other unidentified vocalizations labeled as 'Unidentified Sounds' to the training dataset (Fig. 1; Table 1), following the recommendation to include a non-event class (https://github.com/kahst/BirdNET-Analyzer). Additionally, we incorporated non-target species' vocalizations to align with BirdNET's training methodology and developer recommendations (Kahl et al., 2021; https://github.com/kahst/BirdNET-Analyzer), which suggest that including non-target sounds enhances target species' predictive performance. Including these vocalizations could improve the classifier's ability to differentiate cane toad calls from spectrally and temporally similar sounds, enhancing its robustness and utility. We further refined the training data and enhanced detection accuracy by applying a low-pass filter to the cane toad training data using *scipy.signal* in Python (filtering out frequencies above 1300 Hz) (MacCallum et al., 2011). This step helped to prevent BirdNET from learning to recognize commonly co-occurring sounds, such as other amphibians and insect noise. Finally, we used the full dataset to train the cane toad classifier with the custom training function in the BirdNET software [Epochs = 100, Batch size = 32, Learning rate = 0.001].

Table 1
Details of the training dataset generated by template analysis from the Australian Acoustics Observatory (A2O). NSW stands for New South Wales, NT for Northern Territory, and QLD for Queensland and WA for Western Australia.

Sound class	State	Months	Recording time (Per 2 h)	N (3 s audio snippet)
Background	NT, QLD	Jan, April Jun, Jul, Sep	18:00, 22:00, 02:00, 04:00	623
Canis lupus	NT, QLD, WA	Mar–Jun, Aug–Dec	18:00–04:00	153
Centropus phasianinus	QLD	Feb, Mar, Jul, Sep–Dec	18:00, 04:00	323
Cyclorana asutralia	WA	Nov	22:00, 00:00	925
Cyclorana cryptotis	QLD	Nov	04:00	241
Cyclorana novaehollandiae	QLD	Nov, Dec	20:00–04:00	10,959
Dacelo novaeguineae	NSW, NT, QLD	Jan–May, Jul–Dec	18:00, 04:00	324
*Rhinella marina**	NT, QLD, WA	Jan–Dec	18:00–04:00	22,006
Ninox boobook	QLD	Feb–May, Jul–Nov	18:00–04:00	1218
Notaden melaoscaphus	QLD	Jan–Mar	18:00, 20:00	2291
Unknown	NT, WA	Aug–Nov	18:00, 22:00, 02:00, 04:00	60

Note: *The training dataset for *Rhinella marina* was generated using recordings from the following A2O stations, listed from west to east: Uunguu Indigenous Protected Area, Litchfield Savanna, Staaten River National Park, Mitchell Grass Rangeland, Moorrinya National Park, Undara National Park, Wambiana Cattle Station, and Fletcherview Research Station (Fig. S1).

2.5. Preliminary assessment

To ensure the reliability of our trained classifier, a preliminary performance assessment was conducted before running the classifier over all A2O sites (Fig. 1), which would have taken hundreds of hours. For this initial assessment, we created a test dataset consisting of 20 two-hour recordings from nine sites within or close to suitable habitat for cane toads (suitability ≥0.4; Kelly et al., 2023). We used audio data from recorders not used in the training dataset (Fig. S2; Table S1). To ensure a diverse representation of available soundscapes, these recordings were selected randomly across the temporal distribution of available recordings, encompassing both wet and dry seasons, as well as a range of times between 18:00 and 06:00. Each three-second segment (n = 45,720) of the selected recordings was manually labeled as either 'Cane Toad' or 'Not Cane Toad'. The 20 selected recordings were then analyzed using the custom-trained classifier and the species list (including all the sound classes from the training data) using BirdNET software. This analysis can also be performed using the *analyze.py* script, available at https://github.com/kahst/BirdNET-Analyzer. The performance of the classifier was evaluated using the *eventEval* function from *monitoR* (Hafner and Katz, 2018). This function allowed us to categorize the detected events as true positives (TP), true negatives (TN), false positives (FP), or false negatives (FN). We then computed precision and recall metrics, with precision representing the proportion of our classifier's detections that correctly identified cane toad calls [precision = TP/(TP + FP), and recall representing the proportion of actual cane toad detections captured by our classifier, recall = TP/(TP + FN)]. Recognizing that different projects may have different desired temporal resolutions at which cane toad detections are required (e.g., most projects will not need to detect every call at three-second intervals, but rather, for example, hourly presence), we conducted precision and recall analyses across a wide range of temporal scales, from seconds (3 s) to an hour (3600 s). This approach provided an understanding of our classifier's performance under diverse conditions and temporal resolutions, thus enhancing its applicability for studies with different objectives.

2.6. Audio analysis & classifier performance assessment

We processed the audio data from all selected A2O sites ($n = 40$) on the JCU HPC with our trained classifier (Fig. 1), adjusting the sensitivity parameter to 1.5 [0–1.5] and minimum confidence score to 0.1 [0–1] to optimize cane toad detection (Kahl et al., 2021). The confidence score represents BirdNET's 'confidence' in its predictions, with higher values generally indicating greater prediction accuracy, although this relationship varies across species (Wood and Kahl, 2024). Setting a high sensitivity performs better in acoustically dense environments by detecting more sound events, whereas setting a lower minimum confidence score generates more predictions, although there may be more false positives, and both are necessary for threshold performance evaluation (Kahl et al., 2021). To provide a standardized threshold and assess classifier accuracy, we calculated probabilistic scores following Wood and Kahl (2024).

With over 3,500,000 detections, manual validation of each one was impractical, necessitating their treatment as putative observations subject to a probabilistic threshold (e.g., Brunk et al., 2023). Using the *segments.py* script from BirdNET-Analyzer, we randomly sampled 270 cane toad detections per site across confidence scores (0.1–1) (Barré et al., 2019; Metcalf et al., 2022), validating 30 detections per 0.1 interval. Some sites with fewer than 30 detections in specific confidence score intervals had smaller sample sizes, and only sites with at least one validated true-positive detection were included in further analyses. Subsequently, we manually validated 8623 predictions using Kaleidoscope Lite version 5.6.6 (Wildlife Acoustics®, Manyard, MA, USA) for all 40 selected sites and retained 6546 predictions from 26 toad-positive sites for subsequent analysis. We used logistic regression to evaluate the classifier's spatio-temporal transferability, using validated predictions (correct vs. incorrect) and BirdNET confidence scores, with season (dry/wet) and site as covariates, to determine the probabilistic thresholds (see Wood et al., 2023b and Wood and Kahl, 2024 for detailed methodology). To evaluate potential site-specific seasonal accuracy, we performed another logistic regression analysis incorporating interaction terms between season and site as covariates. We then extracted the coefficients for these interaction terms, computed estimated marginal means, and conducted pairwise comparisons using the *emmeans* package in R (Lenth, 2024) to quantify the differences.

2.7. Detection optimization using time-aggregated features

Given the extensive temporal range of the data, non-perfect precision is likely to generate many false positives (FP), making it difficult to determine cane toad site presence directly from recognizer output without further validation. As cane toads typically call intensively over extended periods during chorusing events (e.g., Brodie et al., 2020, 2022), isolated calls are uncommon. Therefore, we anticipated that incorporating time-series features would enhance the accuracy of the classifier in distinguishing true positives (TP) from FP. To optimize the classifier's detection results for studies requiring high precision and recall at fine temporal resolution, a thresholding framework suggested by Singer et al. (2024) was applied. This approach integrates contextual information from aggregated time-series data, including the quality (i.e., raw model score) and quantity of detections at varying temporal intervals.

In this detection optimization approach (Fig. 1), statistical parameters aggregating detection quality and quantity across 12 time intervals were calculated (Table 2). Along with the original BirdNET confidence score, a total of 169 predictors were modeled using conditional inference trees (CIT) (Hothorn et al., 2006). These CIT models identified threshold values that maximized differentiation between true and false positives for the 6546 validated cane toad detections. By allowing interactions among predictor variables, the models used all 169 variables in combination. With a tree depth set to two, the threshold rules incorporated up to two conditions. These conditions included either (1) a minimum confidence score per time interval, (2) a minimum number of detections per time interval, or (3) a combination of both. This resulted in 14,364 model combinations.

CIT models were ranked based on a performance metric, calculated as the weighted sum of precision (p) and recall (r): model performance $= p \times w + r \times (1 - w)$. The weighting factor ($w$) was set to 0.75 (Singer et al., 2024), as low precision is more likely to bias ecological inferences than low recall (Metcalf et al., 2022). To assess whether the optimized thresholds enhanced detection performance, the selected CIT models were compared against three universal thresholds (filtering above a certain BirdNET confidence score; UNI10: confidence score $\geq$ 0.1, UNI50: confidence score $\geq$ 0.5, UNI90: confidence score $\geq$ 0.9) (Wood et al., 2021) using precision, recall, and model performance. Due to multicollinearity among the aggregated time-series features, multiple candidate models with identical performance arose. We used only the simplest optimized threshold in the main text, figures, and tables. Furthermore, to evaluate the optimized threshold approach, we

Table 2
Aggregated time-series features, included as predictor variables for threshold modelling with conditional inference trees. All features are calculated for 12 different time intervals (±3 s, ±6 s, ±9 s, ± 12 s, ±10mins, ±20mins, ±30mins, ±40mins, ±12 h, ±24 h, ±48 h, ±72 h), resulting in 169 different predictor variables per validated detection.

Type	Abbreviation	Description
BirdNET default	conf	Original BirdNET confidence score
Aggregated time series features	ndets≥0.1	Number of detections with confidence score ≥ 0.1
	ndets≥0.2	Number of detections with confidence score ≥ 0.2
	ndets≥0.3	Number of detections with confidence score ≥ 0.3
	ndets≥0.4	Number of detections with confidence score ≥ 0.4
	ndets≥0.5	Number of detections with confidence score ≥ 0.5
	ndets≥0.6	Number of detections with confidence score ≥ 0.6
	ndets≥0.7	Number of detections with confidence score ≥ 0.7
	ndets≥0.8	Number of detections with confidence score ≥ 0.8
	ndets≥0.9	Number of detections with confidence score ≥ 0.9
	ndets≥0.99	Number of detections with confidence score ≥ 0.99
	avgconf	Average confidence score
	medconf	Median confidence score
	maxconf	Maximum confidence score
	minconf	Minimum confidence score

Note: Adapted from Singer et al. (2024), Table S2. Abbreviations are modified.

compared the ability of the best-performing CIT model and the highest precision universal threshold (i.e., UNI90) in determining cane toad presence at specific sites across the 40 A2O sites at which cane toads occurred, according to the Atlas of Living Australia, (2024). To further confirm that cane toads were absent at sites where the initial 270 validation detections returned no true positives, all detections with confidence scores of ≥0.8 at non-detected sites were manually reviewed. Finally, the individual impact of each aggregated time-series predictor on the optimized threshold was assessed using backward selection. For more detailed methodology, see Singer et al. (2024). All analyses were conducted in R, with the classifier and acoustic materials available at https://github.com/Leptobrachium/Gpshing.

3. Results

3.1. Preliminary performance assessment

In the preliminary performance assessment, a test dataset of 20 two-hour recordings from nine sites was analyzed using the classifier. The classifier successfully detected cane toad vocalizations in 9/20 recordings, resulting in 9474 detections. Manual validation of all 45,720 three-second segments confirmed that cane toads were present in 9 of the 11 recordings where they actually occurred. However, the classifier missed two recordings where cane toad activity was extremely low (90.75 detections per hour, compared to 650.79 detections per hour in the other recordings). The remaining 18 recordings were correctly classified as either containing or not containing cane toads, indicating that the classifier did not produce any false positives. To assess detection performance at different temporal resolutions, we analyzed precision and recall across multiple time scales. Across all tested scales, precision remained above 95 % for the three selected confidence levels (0.1, 0.5, and 0.9; Fig. 2). Notably, the classifier achieved perfect precision from the two-to-five-minute level onward, meaning that no false positives were detected at or beyond this resolution. For recall, the classifier consistently detected at least 30 % of cane toad calls across all confidence levels, increasing to approximately 75 % between the 15-to-30-min intervals (Fig. 2). At the 0.9 confidence threshold, recall peaked at around 80 % at the 45-min level before slightly decreasing to 75 %, primarily due to missed detections in two recordings with low cane toad activity. However, it is important to keep in mind that these results are based on a dataset of 20 randomly selected two-hour recordings.

3.2. Classifier spatio-temporal performance

Overall, our trained classifier identified 3,542,224 cane toad detections across 778,039 h of A2O recordings (~88.75 years) from 40 A2O sites. Cane toad presence was confirmed at 26 sites (n = 6546) based on validation of 8623 randomly selected detections spanning a confidence score range of 0.1 to 1 (Fig. S3). We observed considerable spatial variation in the distribution of confidence scores among the validated detections (Fig. S3). At some sites, including Chillagoe, Litchfield, Moorrinya, and Spyglass, all validated detections were true positives, regardless of confidence score. In contrast, other sites displayed a gradual increase in the proportion of true positives as confidence scores approached the maximum (e.g., Boodjamulla, Doonan Creek). However, at certain sites, true positives were rare, with only one or a few scattered across the confidence score range (e.g., Minjerribah, Mourachan) (Fig. S3).

There was a positive correlation between the confidence score and the probability of a correct BirdNET prediction (hereafter referred to as accuracy) using a logistic model (intercept = 2.980, SE = 0.177; $p <$ 0.001; Table S2). Overall, accuracy reached over 80 % across all sites and seasons at the highest confidence score (0.99). However, there was significant spatial variability in accuracy (Fig. 3; Table S2). With a confidence score above 0.9, 17/26 sites achieved high accuracy levels exceeding 90 %. Among these, Chillagoe, Litchfield, Moorrinya, and Spyglass achieved perfect accuracy (100 %) across all scores (Fig. 3). Additionally, BirdNET predictions had a < 5 % chance of being incorrect at the highest confidence scores at 15/26 sites (Fig. 3). Notably, 13/26 sites maintained consistently high accuracy (≥90 %) even at lower confidence thresholds (0.5). In contrast, 6/26 sites exhibited low accuracy (<70 %) even at the highest confidence score (Fig. 3). There was also a significant difference in accuracy between seasons, such that predictions for the wet season were more accurate than the dry season across confidence scores (intercept = 0.530, SE = 0.102; p < 0.001; Table S2). This seasonal difference, however, was consistently less than 10 % (Fig. 4). Furthermore, 10/25 sites displayed site-specific seasonal variability in accuracy, with seven of them achieving higher accuracy in the dry season compared to the wet season (Table S3).

3.3. Detection performance using time-series aggregated features

Because cane toads tend to call within lengthy choruses, the context within which a detection occurs can include important information which might increase the probability of a correct detection. We noticed the trend for prolonged calling earlier, in section 3.1, in which increasing the temporal scale of detection to periods longer than the 3 s single detection made it more likely we would correctly detect calling. Thus, we used a statistical method applying the aggregated time-series data (Singer et al., 2024), to process the data. Sixteen candidate CIT

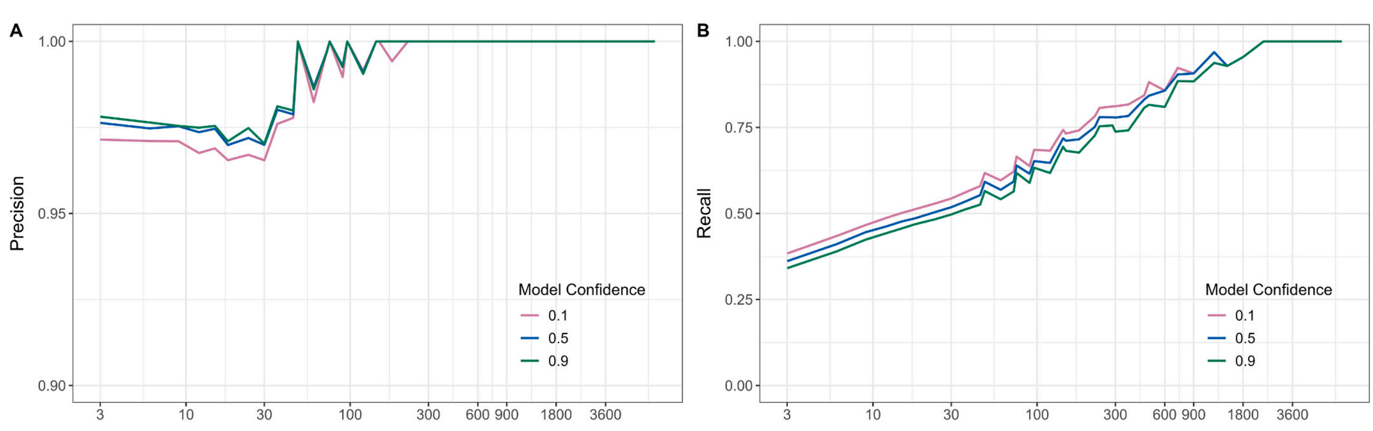

Fig. 2. Preliminary assessment of the trained cane toad classifier on 20 randomly selected testing data from nine A2O sites. The lines depict the classifier's precision and recall across various temporal resolutions measured in seconds. Note that model confidence and precision converge at 100 % at a resolution of 300 s, while recall is optimal at 2700 s. A resolution of approximately 10 min (~600 s) strikes a balance, achieving both high precision and recall.

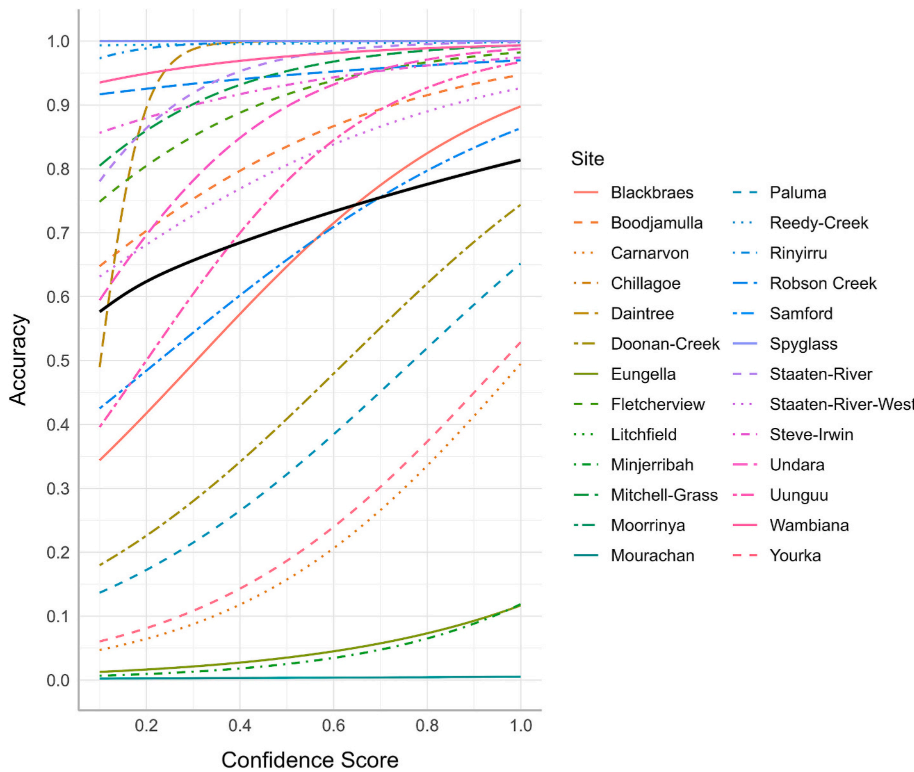

Fig. 3. The relationship between the probability of accurate BirdNET predictions (accuracy) of cane toad detections across BirdNET confidence scores in 26 toad-presence A2O sites. Each line represents a distinct site, with accuracy plotted as a function of the confidence score. The solid black line shows the overall trend, indicating the general relationship between confidence score and accuracy across all sites. The lines for Chillagoe, Litchfield, and Moorrinya are covered by Spyglass's.

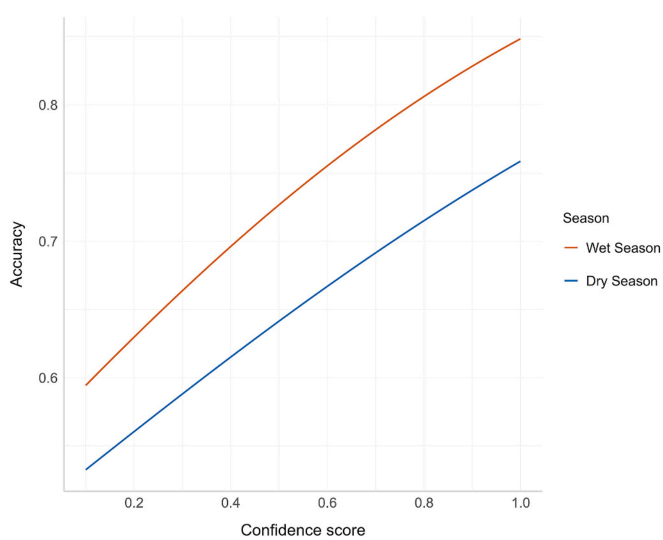

Fig. 4. The relationship between probability of accurate BirdNET predictions (accuracy) of cane toad detections across BirdNET confidence scores in wet and dry seasons.

models demonstrated the highest achievable performance (precision: 93.7 %, recall: 79.8 %, model performance: 0.902). Among these, the simplest optimized threshold was derived from the model formula, which required the ±12-h average confidence score to be greater than 0.52 and at least five detections with confidence score ≥ 0.1 in that period (Table S4). We compared the success of applying this method with another way to increase the probability of correct detections, which is applying a high confidence threshold. When applying universal thresholds, precision levels were 67.7 %, 71.9 %, and 79.6 % for UNI10, UNI50, and UNI90, respectively (Fig. 5; Table S5). At the three-second level, the optimized threshold improved precision by 14.1 % over the highest-precision universal threshold (UNI90). For recall, the optimized thresholds showed an 21.3 % improvement over UNI50, second only to the perfect recall rate of UNI10 (achieved with a minimum confidence of 0.1 in BirdNET audio analysis, see section 2.6). Overall, the optimized threshold outperformed other approaches to increasing model performance (Fig. 5; Table S5).

To better characterize the best approach to optimization, we examined the performance of models with different features included. The performance of the optimized thresholds declined as aggregated time-series features were progressively removed (Fig. 6). Precision remained at its peak of 93.7 % when predictors incorporating time intervals longer than ±12 h were included, while recall maintained its maximum of 79.8 % with predictors integrating intervals longer than ±20 min. Excluding all time intervals resulted in only a minor decrease in precision (~4 %) compared to using the original BirdNET confidence score alone, but it led to a significant drop in recall by over 70 % (i.e., more correct calls were missed). Model performance also showed a notable decline of approximately 0.2 without incorporating time intervals. On the other hand, removing the 12 least informative statistical parameters had no effect on model performance, precision, or recall. However, the average confidence score was critical, as its removal caused the most substantial decreases in model performance (~20 %), precision (~1.5 %), and recall (~70 %).

CIT optimized thresholds increased detection accuracy of cane toad occurrences compared to UNI90 across various sites (Fig. 7). Using the UNI90 threshold, 13 site-level false positives were identified, predominantly in southern and western regions, whereas the CIT optimized threshold produced only three false positive (Fig. 7). The application of the CIT optimized threshold remarkably improved the performance of

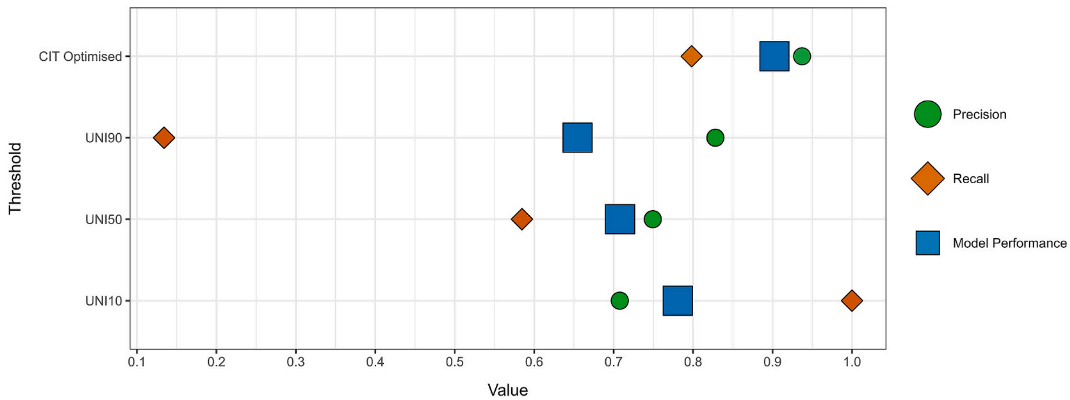

Fig. 5. Comparison of precision, recall and model performance among three universal thresholds (filtering above a certain BirdNET confidence score; UNI10: confidence score ≥ 0.1, UNI50: confidence score ≥ 0.5, UNI90: confidence score ≥ 0.9), and optimized thresholds derived from conditional inference trees (CIT Optimized: ±12 h > 0. 516,334,615 and at least five detections with confidence ≥0.1). Model performance is calculated as the weighted sum of the precision (p) and the recall (r): model performance = p × 0.75 + r × (1–0.75).

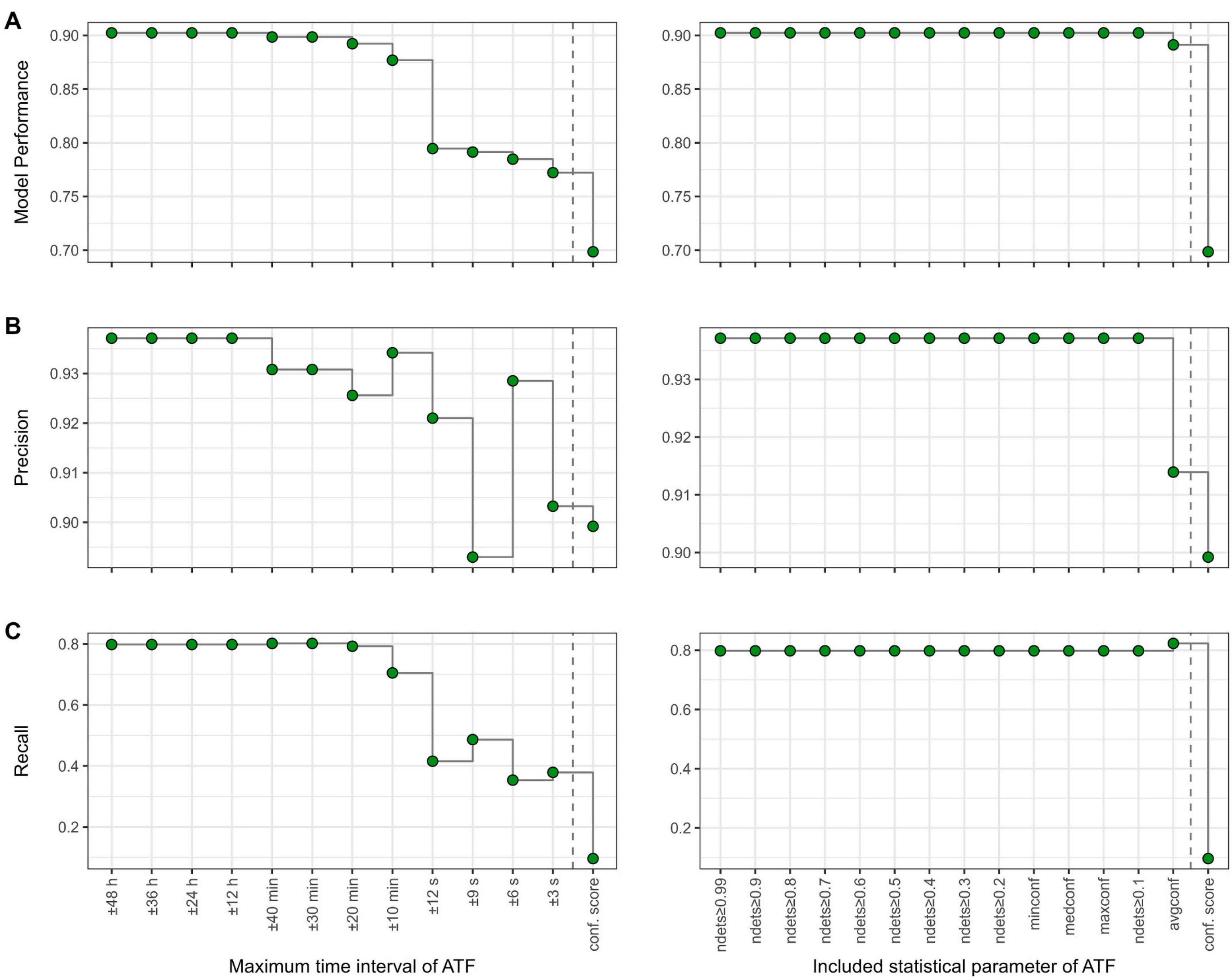

Fig. 6. Effects of stepwise reduction of aggregated time-series features (ATF) on the optimized threshold models' performance: [A: model performance; B: precision; C: recall]. Predictors are ordered according to time interval length and average effect on the model, derived from bootstrapping with 999 permutations. The dashed line separates models that include only the original BirdNET confidence score, based on a three-second interval, from those that incorporate aggregated time-series features.

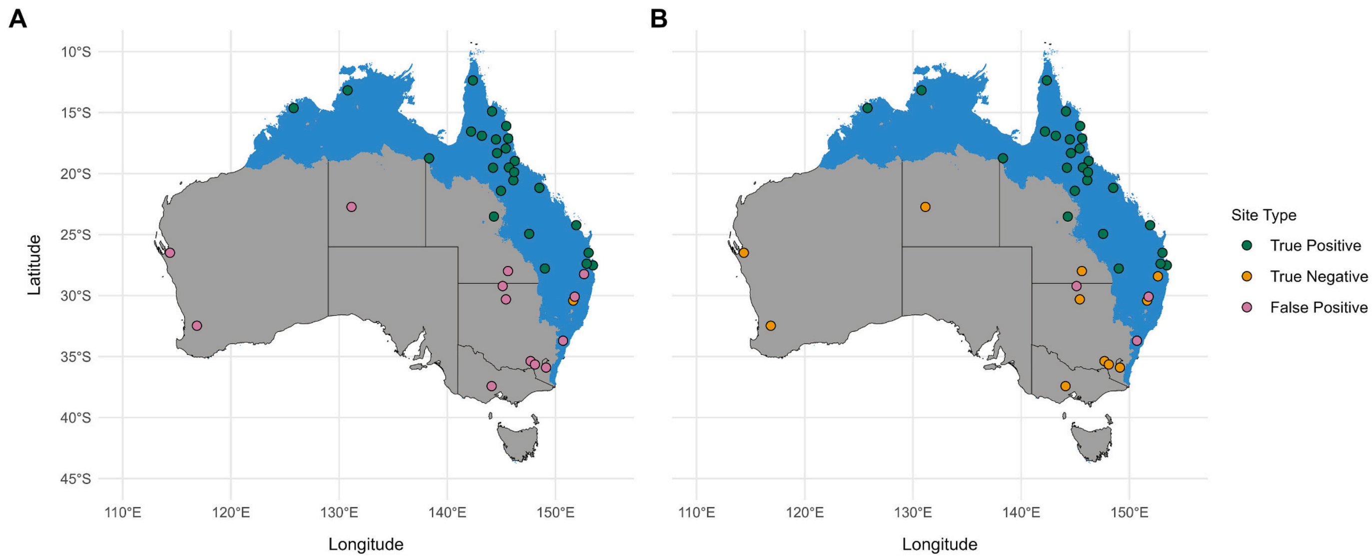

Fig. 7. Comparison of cane toad occurrences across Australia based on two different thresholding approaches. Map A represents occurrences using a universal threshold UNI90 (BirdNET confidence score $\geq$ 0.9); map B represents occurrences using a conditional inference tree (CIT) optimized threshold ($\pm$12 h > 0. 516,334,615 and at least five detections with confidence $\geq$0.1). Each point on the map corresponds to an A2O study site, with colors indicating the detection outcome: true positive (green), true negative (orange) and false positive (pink). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

cane toad detection at site-level, compared to relying solely on a BirdNET confidence score of $\geq$0.9.

4. Discussion

We provide step-by-step information on building our acoustic classifier for cane toads (Fig. 1). Every component of this workflow, including the BirdNET algorithm, the programming tools (R and Python), Kaleidoscope Lite and the A2O audio data were free to use, ensuring that the classifier development process was cost-effective but also widely accessible to researchers and practitioners with varying levels of expertise and resources. While we used a high-performance computing (HPC) cluster to process the large-scale A2O recordings efficiently, BirdNET can also run on standard desktop computers, allowing users to analyze smaller datasets from fewer sites without requiring advanced computing infrastructure. The trained classifier, along with all training materials and code, is available in an online repository (https://github.com/Leptobrachium/Gpshing). We hope that this standardized approach will serve as a reference model for developing similar classifiers for other invasive species.

We found that the trained classifier reliably detected cane toad vocalizations, consistently distinguishing cane toad calls from background noise, achieving high precision and recall across various temporal scales at all confidence levels (Fig. 2). However, while the recall rate was generally strong, it varied with confidence level and temporal resolution (Fig. 2). Nonetheless, the slight decline in recall at the two-hour level occurred because some vocalizations, particularly in low-activity recordings, were missed. This variation underscores the importance of carefully selecting thresholds and temporal resolutions tailored to the specific objectives of the study.

In ecological research, false positives can be more critical mistakes than false negatives, as they suggest a species is present where it is not (Cole et al., 2022; Tolkova et al., 2021). However, missing some vocalizations (false negatives) does not necessarily hinder site-level detection, as the extended survey periods afforded by PAM typically compensate for occasional missed detections (Hoefer et al., 2023; MacKenzie et al., 2002; Tyre et al., 2003). Keeping these criteria in mind, our classifier showed substantial improvement in detecting cane toad occurrences across Australia after incorporating time-aggregated features, reducing false positives from 13 (under a high universal threshold of 0.9) to just three at the site level (Fig. 7). While some false positives remain, this refinement significantly enhances classifier reliability compared to relying solely on the confidence score threshold. Our optimized CIT thresholds maintained high precision, while significantly improving the recall rate for our classifier at even the three-second level across a broad spatio-temporal area (Fig. 5; Table S5). Although we did not explicitly test the spatio-temporal variation of detection results after applying the optimized threshold, the classifier's performance at 17 sites was strong (Fig. 3), with minimal temporal variation (Fig. 4). Therefore, we suggest that our classifier, with the optimized threshold, could meet the precision and recall requirements necessary for application to a range of different study objectives, not just site-level occurrence detection.

Most classifiers are trained and tested in specific study locations or areas (e.g., Manzano-Rubio et al., 2022; Wood et al., 2023b), and may not be accurate when applied to sites not included in the training data (Allen-Ankins et al., 2024). Similarly, our classifier varied in detection accuracy in and across different sites and seasons (Figs. 3, 4; Table S2, S3). However, despite the observed variation in accuracy, our classifier, trained on a subset of audio data from eight A2O sites, still demonstrated robust performance across many locations, including those geographically distant from the original training sites (Fig. 3, S2). It also showed that our training data preparation method was efficient (section 2.3). During manual validation of top detections from underperforming or false-positive A2O sites, we found that many of these sites were located in ecoregions not represented in the training dataset, specifically temperate grasslands, savannas, and shrublands, as well as tropical and subtropical moist broadleaf forests. In contrast, our training data included cane toad sounds, environmental noises, and other vocalizations of spectrally overlapped species from tropical and subtropical grasslands, savannas, and shrublands and temperate broadleaf and mixed forests. This lack of representation from temperate grasslands, savannas, shrublands, and tropical and subtropical moist broadleaf forests likely contributed to the classifier's difficulty in generalizing to these underrepresented sites. Although it may not be necessary to collect training data from every site, including samples from representative areas that cover a range of ecoregions, latitudes, and longitudes greatly increases the robustness of the recognizer at different sites. Thus, strategically selecting training data from diverse and representative regions is both time-efficient and effective in enhancing classifier function across different environments.

The application of optimized CIT thresholds significantly improved the classifier's performance compared to relying solely on the BirdNET confidence score (Fig. 5; Table S5), demonstrating the effectiveness of integrating contextual information from time-series data into automatic call detection (Madhusudhana et al., 2021; Singer et al., 2024). Notably, the reduction in false positives at site level, especially in regions where misclassifications were common under the UNI90 threshold (Fig. 7), demonstrates the practical use of CIT optimization to refine classifier outputs and minimize erroneous detections. In addition, the optimized thresholds were most effective when incorporating time intervals no shorter than ± 12 h (Fig. 6), suggesting that temporal context is important for accurately classifying cane toad occurrences. The decline in performance when the length of time intervals was excluded marks the importance of maintaining temporal granularity in the analysis, particularly for studies that monitor species over extended periods and across broad geographical regions. Among the statistical parameters, average confidence scores and the number of detections with a confidence score ≥ 0.1 provided the most valuable information. Average confidence scores at various time intervals were also the most influential parameter in previous studies on birds (Singer et al., 2024; Wood et al., 2021), implying that average scores could be a more reliable threshold than simply using minimum confidence scores in acoustic classification.

While our study demonstrates promising detection results from our acoustic classifier, some limitations should be acknowledged. Firstly, spatial performance was not flawless, likely because we limited training data to particular ecoregions to keep our approach efficient. Researchers applying this classifier to new study systems should retrain it with additional data from their specific sites to account for variations in the soundscape, ensuring optimal performance, broadening its applicability and minimizing site-specific biases. Apart from spatial characteristics, detection quality appears to depend on call abundance. True-positive detections in peripheral distribution areas and sites with lower cane toad activity (e.g., Mourachan, Minjerribah) become valuable for enhancing the classifier's recall performance in these regions. It is worth noting that our classifier and the entire training dataset have been made publicly available on an online repository. The above cases could be used as additional training data in future updates. We also recommend that users always assess a subset of detection results for performance, even after applying the CIT approach, as the adaptation to spatio-temporal variation following the application of optimized thresholds has not yet been fully tested.

In light of our findings, several promising opportunities for future research and development arise. Although our classifier was trained on data from Australia, it has the potential to detect cane toads in other regions where they have been introduced. Expanding its application to countries such as French Guiana, the Philippines, and the United States (Harvey et al., 2021; Shine et al., 2021) could provide valuable insights into its performance in even more diverse environments. Collaborating with researchers in these regions would both test the classifier's adaptability and enhance its utility by incorporating additional vocalizations from local species, ultimately improving effectiveness for global cane toad monitoring and management. Secondly, our classifier has potential applications beyond monitoring. By providing a reliable cane toad call detection tool, it can contribute to broader efforts to understand the species' adaptation, breeding ecology, and ecological impact in Australia. For example, our classifier can aid in examining spatio-temporal patterns and environmental factors that influence cane toad calling activity. Such studies could reveal key aspects of their reproductive behavior, including chorus timing and frequency (e.g., Bolitho et al., 2023; Wood et al., 2023a). Broadly, analyzing calling activity patterns across the diverse locations sampled by the A2O could help identify breeding hot spots and habitat preferences, furthering our understanding of cane toad ecology and distribution. Despite extensive research on cane toad invasion and its impact on native species (e.g., Greenlees et al., 2006; Lampo and De Leo, 1998; Shine, 2010), toads' effects on the acoustics of local frog communities remain underexplored. While playback experiments have shown that cane toad calls influence the timing, rate, and inter-call intervals of some Australian frogs (Bleach et al., 2015; Hopkins et al., 2023; Taylor et al., 2017), a comprehensive understanding of their impact on acoustic resources, especially across larger spatial scales, is lacking. Using our classifier to analyze large-scale audio data could determine whether cane toads disrupt the acoustic space of local frog communities in real-world settings, shedding light on their effects on microhabitat use by native species.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author (Franco Ka Wah Leung) used ChatGPT to check grammar and improve readability. After using this tool/service, all authors reviewed and edited the content as needed and F.K.W. Leung takes full responsibility for the content of the publication.

CRediT authorship contribution statement

Franco Ka Wah Leung: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Lin Schwarzkopf:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization. **Slade Allen-Ankins:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Project administration, Methodology, Conceptualization.

Acknowledgement

We extend our sincere thanks to Dr. Anthony Truskinger for his invaluable assistance in resolving technical issues with the Australian Acoustic Observatory. We also express our gratitude to Jacqueline Richmond-Clay and Layne Dellevergini for their crucial help in validating acoustic detections. Additionally, we acknowledge the Australian Research Council (LE170100033) for funding the Australian Acoustic Observatory.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2025.103172.

Data availability

The codes and training data used to develop the classifier are freely and publicly available on GitHub repository: https://github.com/Leptobrachium/Gpshing.

References

Allen-Ankins, S., Mula, L.J., Schwarzkopf, L., 2024. Towards automated detection of the endangered southern black-throated finch (*Poephila cincta cincta*). Wildl. Res. 51, WR23151. https://doi.org/10.1071/WR23151.

Allen-Ankins, S., Hoefer, S., Bartholomew, J., Brodie, S., Schwarzkopf, L., 2025. The use of BirdNET embeddings as a fast solution to find novel sound classes in audio recordings. Front. Ecol. Evol. 12, 1409407. https://doi.org/10.3389/fevo.2024.1409407.

Amorim, M.C.P., Wanjala, J.A., Vieira, M., Bolgan, M., Connaughton, M.A., Pereira, B.P., Fonseca, P.J., Ribeiro, F., 2023. Detection of invasive fish species with passive acoustics: discriminating between native and non-indigenous sciaenids. Mar. Environ. Res. 188, 106017. https://doi.org/10.1016/j.marenvres.2023.106017.

Atlas of Living Australia, 2024. Species Occurrence Data for *Rhinella marina*. https://bie.ala.org.au/species/https://biodiversity.org.au/afd/taxa/5c8e5d04-49dd-43ec-b9e0-f4e55ee82407 (accessed 22 March 2024).

Balantic, C., Donovan, T., 2020. Statistical learning mitigation of false positives from template-detected data in automated acoustic wildlife monitoring. Bioacoustics 29, 296–321. https://doi.org/10.1080/09524622.2019.1605309.

Barré, K., Le Viol, I., Julliard, R., Pauwels, J., Newson, S.E., Julien, J.F., Claireau, F., Kerbiriou, C., Bas, Y., 2019. Accounting for automated identification errors in acoustic surveys. Methods Ecol. Evol. 10, 1171–1188. https://doi.org/10.1111/2041-210X.13198.

Bleach, I., Beckmann, C., Both, C., Brown, G., Shine, R., 2015. Noisy neighbours at the frog pond: effects of invasive cane toads on the calling behaviour of native Australian frogs. Behav. Ecol. Sociobiol. 69, 675–683. https://doi.org/10.1007/s00265-015-1879-z.

Blumstein, D., Mennill, D., Clemins, P., Girod, L., Yao, K., Patricelli, G., Deppe, J., Krakauer, A., Clark, C., Cortopassi, K., Hanser, S., McCowan, B., Ali, A., Kirschel, A., 2011. Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus. J. Appl. Ecol. 48, 758–767. https://doi.org/10.1111/j.1365-2664.2011.01993.x.

Bolitho, L., Newell, D., Hines, H., 2023. Bioacoustic monitoring reveals the calling activity of an endangered mountaintop frog (*Philoria kundagungan*) in response to environmental conditions. Diversity 15, 931. https://doi.org/10.3390/d15080931.

Bradshaw, C.J., Leroy, B., Bellard, C., Roiz, D., Albert, C., Fournier, A., Barbet-Massin, M., Salles, J., Simard, F., Courchamp, F., 2016. Massive yet grossly underestimated global costs of invasive insects. Nat. Commun. 7, 12986. https://doi.org/10.1038/ncomms12986.

Brodie, S., Allen-Ankins, S., Towsey, M., Roe, P., Schwarzkopf, L., 2020. Automated species identification of frog choruses in environmental recordings using acoustic indices. Ecol. Indic. 119, 106852. https://doi.org/10.1016/j.ecolind.2020.106852.

Brodie, S., Towsey, M., Allen-Ankins, S., Roe, P., Schwarzkopf, L., 2022. Using a novel visualization tool for rapid survey of long-duration acoustic recordings for ecological studies of frog chorusing. Front. Ecol. Evol. 9, 761147. https://doi.org/10.3389/fevo.2021.761147.

Brunk, K., Gutiérrez, R., Peery, M., Cansler, C., Kahl, S., Wood, C.M., 2023. Quail on fire: changing fire regimes may benefit mountain quail in fire-adapted forests. Fire Ecol. 19, 19. https://doi.org/10.1186/s42408-023-00180-9.

Cole, J., Michel, N., Emerson, S., Siegel, R., 2022. Automated bird sound classifications of long-duration recordings produce occupancy model outputs similar to manually annotated data. Ornithol. Appl. 124, duac003. https://doi.org/10.1093/ornithapp/duac003.

Doody, J., McHenry, C., Rhind, D., Clulow, S., 2019. Novel habitat causes a shift to diurnal activity in a nocturnal species. Sci. Report. 9, 230. https://doi.org/10.1038/s41598-018-36384-2.

Easteal, S., 1981. The history of introductions of *Bufo marinus* (Amphibia: Anura); a natural experiment in evolution. Biol. J. Linn. Soc. 16, 93–113. https://doi.org/10.1111/j.1095-8312.1981.tb01645.x.

Ehrenfeld, J., 2010. Ecosystem consequences of biological invasions. Annu. Rev. Ecol. Evol. Syst. 41, 59–80. https://doi.org/10.1146/annurev-ecolsys-102209-144650.

Gibb, R., Browning, E., Glover-Kapfer, P., Jones, K., 2019. Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. Methods Ecol. Evol. 10, 169–185. https://doi.org/10.1111/2041-210X.13101.

Greenlees, M., Brown, G., Webb, J., Phillips, B., Shine, R., 2006. Effects of an invasive anuran [the cane toad (*Bufo marinus*)] on the invertebrate fauna of a tropical Australian floodplain. Anim. Conserv. 9, 431–438. https://doi.org/10.1111/j.1469-1795.2006.00057.x.

Hafner, S., Katz, J., 2018. A Short Introduction to Acoustic Template Matching with monitoR.

Harvey, J.A., Ambavane, P., Williamson, M., Diesmos, A., 2021. Evaluating the effects of the invasive cane toad (*Rhinella marina*) on island biodiversity, focusing on the Philippines. Pac. Conserv. Biol. 28, 199–210. https://doi.org/10.1071/PC21012.

Hoefer, S., McKnight, D., Allen-Ankins, S., Nordberg, E., Schwarzkopf, L., 2023. Passive acoustic monitoring in terrestrial vertebrates: a review. Bioacoustics 32, 506–531. https://doi.org/10.1080/09524622.2023.2209052.

Hopkins, J., Bower, D., Edwards, W., Schwarzkopf, L., 2023. Effects of invasive toad calls and synthetic tones on call properties of native Australian toadlets. J. Herpetol. 57, 437–446. https://doi.org/10.1670/23-004.

Hothorn, T., Hornik, K., Zeileis, A., 2006. Unbiased recursive partitioning: a conditional inference framework. J. Comput. Graph. Stat. 15, 651–674. https://doi.org/10.1198/106186006X133933.

Hu, G., Wang, D., 2007. Auditory segmentation based on onset and offset analysis. IEEE Trans. Audio Speech Lang. Process. 15, 396–405. https://doi.org/10.1109/TASL.2006.881700.

Hu, W., Bulusu, N., Chou, C., Jha, S., Taylor, A., Tran, V., 2009. Design and evaluation of a hybrid sensor network for cane toad monitoring. ACM Trans. Sensor Netw. 5, 1–28. https://doi.org/10.1145/1464420.1464424.

Jeantet, L., Dufourq, E., 2023. Improving deep learning acoustic classifiers with contextual information for wildlife monitoring. Ecological Informatics 77, 102256. https://doi.org/10.1016/2.eco.inf.2023.102256.

Kahl, S., Wood, C., Eibl, M., Klinck, H., 2021. BirdNET: a deep learning solution for avian diversity monitoring. Eco. Inform. 61, 101236. https://doi.org/10.1016/j.ecoinf.2021.101236.

Katz, J., Hafner, S., Donovan, T., 2016. Assessment of error rates in acoustic monitoring with the R package monitoR. Bioacoustics 25, 177–196. https://doi.org/10.1016/j.ecoinf.2021.101236.

Kelly, C., Schwarzkopf, L., Christy, T., Kennedy, M., 2023. The toad less travelled: comparing life histories, ecological niches, and potential habitat of Asian black-spined toads and cane toads. Wildl. Res. 51, WR22111. https://doi.org/10.1071/WR22111.

Knight, E., Hannah, K., Foley, G., Scott, C., Brigham, R., Bayne, E., 2017. Recommendations for acoustic recognizer performance assessment with application to five common automated signal recognition programs. Avian Conserv. Ecol. 12, 14. https://doi.org/10.5751/ACE-01114-120214.

Lampo, M., De Leo, G.A., 1998. The invasion ecology of the toad *Bufo marinus*: from South America to Australia. Ecol. Appl. 8, 388–396. https://doi.org/10.1890/1051-0761(1998)008[0388:TIEOTT]2.0.CO;2.

Lauha, P., Somervuo, P., Lehikoinen, P., Geres, L., Richter, T., Seibold, S., Ovaskainen, O., 2022. Domain-specific neural networks improve automated bird sound recognition already with small amount of local data. Methods Ecol. Evol. 13, 2799–2810. https://doi.org/10.1111/2041-210X.14003.

Lenth, R., 2024. Emmeans: Estimated Marginal Means, Aka Least-Squares Means. R Package Version 1.10.4. https://doi.org/10.32614/CRAN.package.emmeans (accessed 1 September 2024).

Lowe, S., Browne, M., Boudjelas, S., De Poorter, M., 2000. 100 of the World's Worst Invasive Alien Species: A Selection from the Global Invasive Species Database. Invasive Species Specialist Group, Auckland.

MacCallum, J.K., Olszewski, A.E., Zhang, Y., Jiang, J.J., 2011. Effects of low-pass filtering on acoustic analysis of voice. J. Voice 25, 15–20.

MacKenzie, D., Nichols, J., Lachman, G., Droege, S., Andrew, R., Langtimm, C., 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology 83, 2248–2255. https://doi.org/10.1890/0012-9658(2002)083[2248:ESORWD]2.0.CO;2.

Madhusudhana, S., Shiu, Y., Klinck, H., Fleishman, E., Liu, X., Nosal, E.M., Helble, T., Chloewiak, D., Gillespie, D., Širović, A., Roch, M.A., 2021. Improve automatic detection of animal call sequences with temporal context. J. R. Soc. 18, 20210297. https://doi.org/10.1098/rsif.2021.0297.

Manzano-Rubio, R., Bota, G., Brotons, L., Soto-Largo, E., Pérez-Granados, C., 2022. Low-cost open-source recorders and ready-to-use machine learning approaches provide effective monitoring of threatened species. Eco. Inform. 72, 101910. https://doi.org/10.1890/0012-9658(2002)083[2248:ESORWD]2.0.CO;2.

Melo, I., Llusia, D., Bastos, R.P., Signorelli, L., 2021. Active or passive acoustic monitoring? Assessing methods to track anuran communities in tropical savanna wetlands. Ecol. Indic. 132, 108305. https://doi.org/10.1016/j.ecolind.2021.108305.

Metcalf, O., Barlow, J., Bas, Y., Berenguer, E., Devenish, C., França, F., Marsden, S., Smith, C., Lees, A., 2022. Detecting and reducing heterogeneity of error in acoustic classification. Methods Ecol. Evol. 13, 2559–2571. https://doi.org/10.1111/2041-210X.13967.

Muller, B., Pike, D., Schwarzkopf, L., 2016. Defining the active space of cane toad (*Rhinella marina*) advertisement calls: males respond from further than females. Behaviour 153, 1951–1969. https://doi.org/10.1163/1568539X-00003404.

Pérez-Granados, C., Feldman, M., Mazerolle, M., 2023. Combining two user-friendly machine learning tools increases species detection from acoustic recordings. Can. J. Zool. 102. https://doi.org/10.1139/cjz-2023-0154.

Pyšek, P., Hulme, P., Simberloff, D., Bacher, S., Blackburn, T., Carlton, J., Dawson, W., Essl, F., Foxcroft, L., Genovesi, P., Jeschke, J., Kuhn, I., Liebhold, A., Mandrak, N., Meyerson, L., Pauchard, A., Pergl, J., Roy, H., Seebens, H., Kleunen, M., Vila, M., Wingfield, M., Richardson, D., 2020. Scientists' warning on invasive alien species. Biol. Rev. 95, 1511–1534. https://doi.org/10.1111/brv.12627.

R Core Team, 2023. R: A Language and Environment for Statistical Computing (Version 4.2.3). R. http://www.R-project.org/.

Ribeiro Jr., J.W., Harmon, K., Leite, G.A., de Melo, T.N., LeBien, J., Campos-Cerqueira, M., 2022. Passive acoustic monitoring as a tool to investigate the spatial distribution of invasive alien species. Remote Sens. 14, 4565. https://doi.org/10.3390/rs14184565.

Roe, P., Ferroudj, M., Towsey, M., Schwarzkopf, L., 2018. Catching toad calls in the cloud: Commodity edge computing for flexible analysis of big sound data. In: 2018 IEEE 14th International Conference on e-Science (e-Science). IEEE, Amsterdam, Netherland, pp. 67–74.

Roe, P., Eichinski, P., Fuller, R.A., McDonald, P.G., Schwarzkopf, L., Towsey, M., Truskinger, A., Tucker, D., Watson, D., 2021. The Australian acoustic observatory. Methods Ecol. Evol. 12, 1802–1808. https://doi.org/10.1111/2041-210X.13660.

Shine, R., 2010. The ecological impact of invasive cane toads (*Bufo marinus*) in Australia. Q. Rev. Biol. 85, 253–291. https://doi.org/10.1086/655116.

Shine, R., Alford, R.A., Blennerhasset, R., Brown, G.P., DeVore, J.L., Ducatez, S., Finnerty, P., Grenlees, M., Kaiser, S.Q., McCann, S., Petit, L., Pizzatto, L., Schwarzkopf, L., Ward-Fear, G., Phillips, B.L., 2021. Increased rates of dispersal of free-ranging cane toads (*Rhinella marina*) during their global invasion. Sci. Rep. 11, 23574. https://doi.org/10.1038/s41598-021-02828-5.

Singer, D., Hagge, J., Kamp, J., Hondong, H., Schuldt, A., 2024. Aggregated time-series features boost species-specific differentiation of true and false positives in passive acoustic monitoring of bird assemblages. Remote Sens. Ecol. Conserv. https://doi.org/10.1002/rse2.385.

Taylor, A., Grigg, G., Watson, G., McCallum, H., 1996. Monitoring frog communities: An application of machine learning. In: Proceedings, Eighth Innovative Applications of Artificial Intelligence Conference. AAAI Press, Portland, Oregon, USA, pp. 1564–1569.

Taylor, A., McCallum, H., Watson, G., Grigg, G., 2017. Impact of cane toads on a community of Australian native frogs, determined by 10 years of automated identification and logging of calling behaviour. J. Appl. Ecol. 54, 2000–2010. https://doi.org/10.1111/1365-2664.12859.

Tolkova, I., Chu, B., Hedman, M., Kahl, S., Klinck, H., 2021. Parsing birdsong with deep audio embeddings. ArXiv Preprint. https://doi.org/10.48550/arXiv.2108.09203.

Tyre, A., Tenhumberg, B., Field, S., Niejalke, D., Parris, K., Possingham, H., 2003. Improving precision and reducing bias in biological surveys: estimating false-negative error rates. Ecol. Appl. 13, 1790–1801. https://doi.org/10.1890/02-5078.
Van Rossum, G., Drake Jr., F.L., 1995. Python Reference Manual (Version 3.11.6). Python.
Vander Zanden, M., Hansen, G., Higgins, S., Kornis, M., 2010. A pound of prevention, plus a pound of cure: early detection and eradication of invasive species in the Laurentian Great Lakes. J. Great Lakes Res. 36, 199–205. https://doi.org/10.1016/j.jglr.2009.11.002.
Villanueva-Rivera, L.J., Pijanowski, B.C., 2012. Pumilio: a web-based management system for ecological recordings. Bull. Ecol. Soc. Am. 93, 71–81. https://doi.org/10.1890/0012-9623-93.1.71.
Wood, C., Kahl, S., 2024. Guidelines for appropriate use of BirdNET scores and other detector outputs. J. Ornithol. 165, 1–6. https://doi.org/10.1007/s10336-024-02144-5.
Wood, C.M., Kahl, S., Chaon, P., Peery, M.Z., Klinck, H., 2021. Survey coverage, recording duration and community composition affect observed species richness in passive acoustic surveys. Methods Ecol. Evol. 12, 885–896. https://doi.org/10.1111/2041-210X.13571.
Wood, C., Kahl, S., Barnes, S., Van Horne, R., Brown, C., 2023a. Passive acoustic surveys and the BirdNET algorithm reveal detailed spatiotemporal variation in the vocal activity of two anurans. Bioacoustics 32, 532–543. https://doi.org/10.1080/09524622.2023.2211544.
Wood, C., Barceinas, C., Kahl, S., 2023b. Pairing a user-friendly machine-learning animal sound detector with passive acoustic surveys for occupancy modeling of an endangered primate. Am. J. Primatol. 85, e23507. https://doi.org/10.1002/ajp.23507.
Yasumiba, K., Duffy, R., Parsons, S., Alford, R., Schwarzkopf, L., 2016. Rapid differentiation of sexual signals in invasive toads: call variation among populations. Sci. Rep. 6, 28158. https://doi.org/10.1016/j.jglr.2009.11.002.