Ecological Informatics 95 (2026) 103764

Contents lists available at ScienceDirect

Ecological Informatics

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

Two-stage approach to automatic detection with machine learning for improved surveillance of the invasive Cuban treefrog

Kaitlin Huber [a,*], J. Hardin Waddle [b], Brad M. Glorioso [c], Therese Donovan [d]

[a] Vermont Cooperative Fish and Wildlife Research Unit, University of Vermont, Burlington, VT, USA
[b] U.S. Geological Survey, Wetland and Aquatic Research Center, 7920 NW 71 St, Gainesville, FL, USA
[c] U.S. Geological Survey, Wetland and Aquatic Research Center, 700 Cajundome Blvd, Lafayette, LA, USA
[d] U.S. Geological Survey, Vermont Cooperative Fish and Wildlife Research Unit, Rubenstein School of Environment and Natural Resources, University of Vermont, Burlington, VT, USA

ARTICLE INFO

Keywords:
Cuban treefrog
Bioacoustics
Automated detection
Machine learning
Template matching
Supervised learning

ABSTRACT

The Cuban treefrog (*Osteopilus septentrionalis*), as an invasive species in the southern United States, presents a need for effective surveillance. Automated detection expedites processing of audio data for large-scale surveillance and monitoring programs. However, current available methods commonly used for anuran species have not been sufficient to detect Cuban treefrogs. Here, we present results from a two-stage method for automated detection that employs both cross-correlation template matching and secondary supervised learning classifiers. In the first stage, audio data are screened for initial detections using template matching, in which the detections contain both true and false positives. In the second stage, the false positives are screened out using classifier algorithms. We used this method to process 139,985 audio recordings, consisting of 596,046 total minutes, collected at 13 locations in Louisiana and Florida from 2014 to 2022. From the stage 1 template matching, we detected 83,191 Cuban treefrog signals across recordings. The stage 2 machine learning model was able to identify stage 1 false positive detections with a testing accuracy of 98.46% and a testing false positive rate of 1.116%. After pruning false positive detections, a total of 20,271 individual Cuban treefrog detections remained, distributed mainly across 3 sites in an area with known presence. Locations with presumed absence had an easily verifiable number of false positive detections ($n = 109$ across all other sites). The two-stage methodology utilizing both template matching and machine learning algorithms can be integrated into wildlife surveillance or monitoring programs for species with distinctive, conserved calls as an effective way to achieve sensitive species detection with a low incidence of false positives.

1. Introduction

The Cuban treefrog (*Osteopilus septentrionalis*) is an invasive species native to Cuba, the Cayman Islands, and The Bahamas that has spread to other islands in the Caribbean Sea and to the United States. A stable population was confirmed in Louisiana in 2017, spreading from previously long-established populations in Florida (Brosnan et al., 2025; Glorioso et al., 2018; Schwartz, 1952). Its success in Florida and surrounding states is attributed largely to its dietary plasticity, high fecundity, short generation times, ability to produce eggs year-round, and ability to thrive in urban settings within subtropical climates (Glorioso et al., 2018; Meshaka, 2001). Cuban treefrogs are known to consume native treefrogs and other wildlife (Glorioso et al., 2010; Meshaka, 2001), and they compete with native treefrogs both acoustically (Tennessen et al., 2016) and for resources such as refugia or prey (Meshaka, 2001; Rice et al., 2011). As such, Cuban treefrogs threaten biodiversity in areas where they become established (Waddle et al., 2010). To combat further dispersal and establishment of this species, accurate data on its current distribution and prompt response to new establishments are essential.

The Cuban treefrog emits a distinctive call that can be captured with audio recorders and used to acquire species occurrence data. Consequently, they can be monitored with autonomous recording units (ARUs), digital recorders programmed to capture audio data on a set schedule (Acevedo and Villanueva-Rivera, 2006; Hutto and Stutzman, 2009; Venier et al., 2012). Monitoring with ARUs streamlines the data

---

* Corresponding author.
E-mail address: kehuber@uvm.edu (K. Huber).

https://doi.org/10.1016/j.ecoinf.2026.103764
Received 4 August 2025; Received in revised form 7 April 2026; Accepted 8 April 2026

collection process, but at the same time it produces vast amounts of acoustic data that take much time to process before any management decisions can occur (Shonfield and Bayne, 2017). Machine learning presents a promising solution with its ability to quickly process large amounts of data and also generate automated detections of target species in ongoing monitoring programs (Acevedo et al., 2009; Godfray et al., 2004).

Many machine learning methodologies exist and continue to develop within the realm of modern bioacoustics, each with benefits and drawbacks specific to research goals, soundscape characteristics, and the acoustic features of the target species (Stowell et al., 2016; Towsey et al., 2012). Unsupervised learning methods (e.g. clustering) do not require labelled datasets, which could benefit monitoring projects with little time or budget for manual audio tagging. Historically, they have been better suited to soundscape analyses, rather than species-specific detection (Coffey et al., 2019; Ozanich et al., 2021; Guerrero et al., 2023), although recent platforms have shown success with individual species detection with the added benefit of delineating between the species' different calls (Schneider et al., 2022). Supervised learning models such as convolutional neural networks (CNNs) (Colonna et al., 2016; Smirnov, 2013) or Hidden Markov Models (Yao et al., 2009) perform very well for some individual species detection when trained on sufficient data. These models, however, often require a high level of expertise in machine learning, in addition to large, labelled training datasets beyond the scope of individual monitoring projects (Gibb et al., 2019). However, advances are made each year, and some can now achieve promising results with smaller training datasets (Eichinski et al., 2022; Otesbelgue et al., 2025). Transfer learning, where pre-trained models can be improved with a small amount of additional regional or species-specific data, and training data augmentation are two methods that have become increasingly popular in automated detection and can achieve desirable results (Ghani et al., 2023; Lauha et al., 2022; MacIsaac et al., 2024).

Template-based methods can also be highly useful for automated detection of individual species or calls, especially where large training datasets and machine learning expertise are lacking (Balantic and Donovan, 2020; Mellinger and Clark, 2000). By converting audio files into visual spectrograms, a signal from a target species can be selected and used as a template. A template is a spectrogram matrix giving information on frequency, time, and amplitude of the signal. Algorithms can score new audio data based on similarity to the template in a moving window analysis, and peaks in scores above a given threshold are considered “detections” (Fig. 1 (A)). For target species with calls that are largely conserved across individuals and populations, template-based detection strategies require only a single acceptable example of the sound, thus allowing monitoring projects to screen their audio data and accumulate predicted detections immediately and without the need for prior examination by human taggers. Template-based methods have been used for various species-specific projects and in several software packages (Cortopassi and Bradbury, 2000; Mellinger and Clark, 2000; Grießmann and Naguib, 2002; Aide et al., 2013; Katz et al., 2016; Bravo et al., 2017).

Detections accumulated using computer-automated methods, including template-based strategies, often have an unacceptable number of false positives, or signals falsely attributed to the target species (Acevedo et al., 2009). Two-stage approaches to automated species detection involving post-processing of initial detections have been shown to be useful for minimizing false positives (Balantic and Donovan, 2020; Thompson et al., 2025; Clarfeld et al., 2025a). An effective two-stage process using both template matching and supervised learning classifiers can be implemented to achieve low incidence of false positives while maintaining sufficient true detections for accurate species occurrence data (Balantic and Donovan, 2020). In this approach, Stage 1 begins as data are first screened using template matching to accumulate initial detections. During template screening, audio segments with correlation scores exceeding the threshold are identified as candidate detections, and each output includes the bounding box corresponding to the detected signal and its associated correlation score. Each “template detection” is therefore an audio segment defined by its bounding box from the screening process, which serves as a key input to the Stage 2 classifier models. A subset of these template detections is manually verified as true detections or false positives. Once a sufficient number has been labelled, these verified detections are used to train and test supervised learning classifiers that then filter out false positives in the remaining data (Fig. 2). Supervised learning classifiers can predict the likelihood of a false positive using various acoustic features of the points on each audio segment's spectrogram, such as amplitude (or sound intensity), binned zero-crossing rates, time and frequency contours of the amplitude probability mass function, and summary statistics of the frequency spectrum (Sueur et al., 2008) (Fig. 1 (B)). The two-stage approach of screening with template matching and then filtering the false positives with well-trained classifier models can be integrated into a monitoring program that has sensitive species detection with a low incidence of false positives.

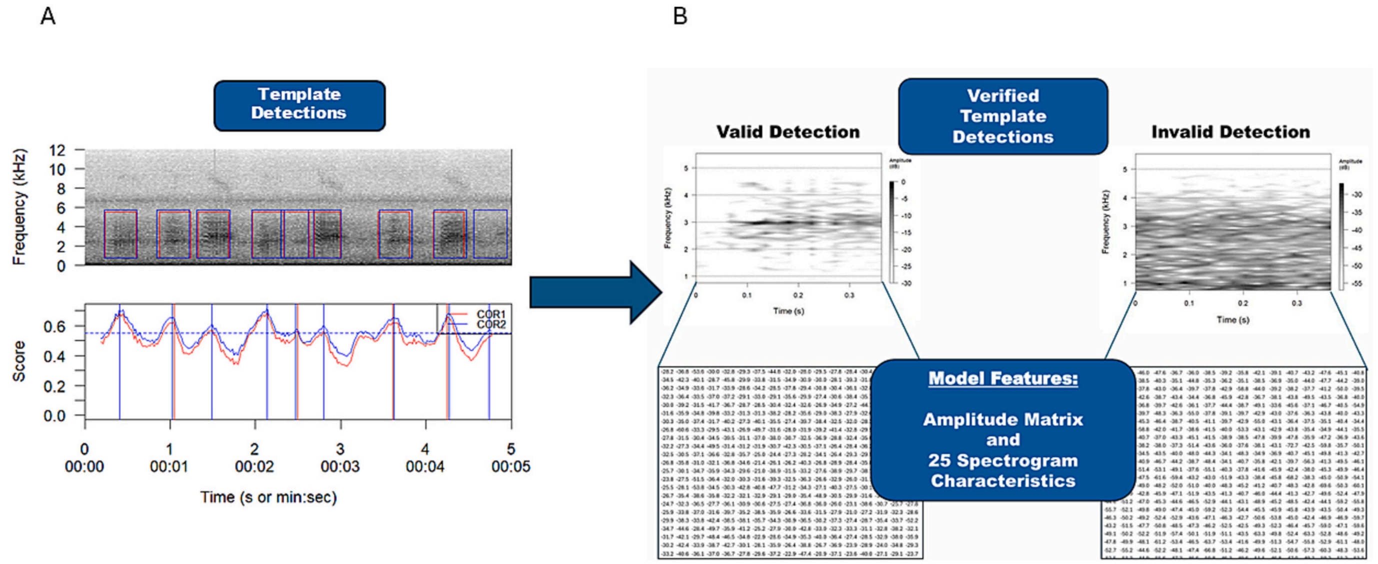

Fig. 1. Scores and detections (where scores meet the given threshold) produced during template matching in moving window analysis (A) and verified template detections and extracted amplitude matrix, illustrating the process of extracting sonogram characteristics from the labelled dataset (B).

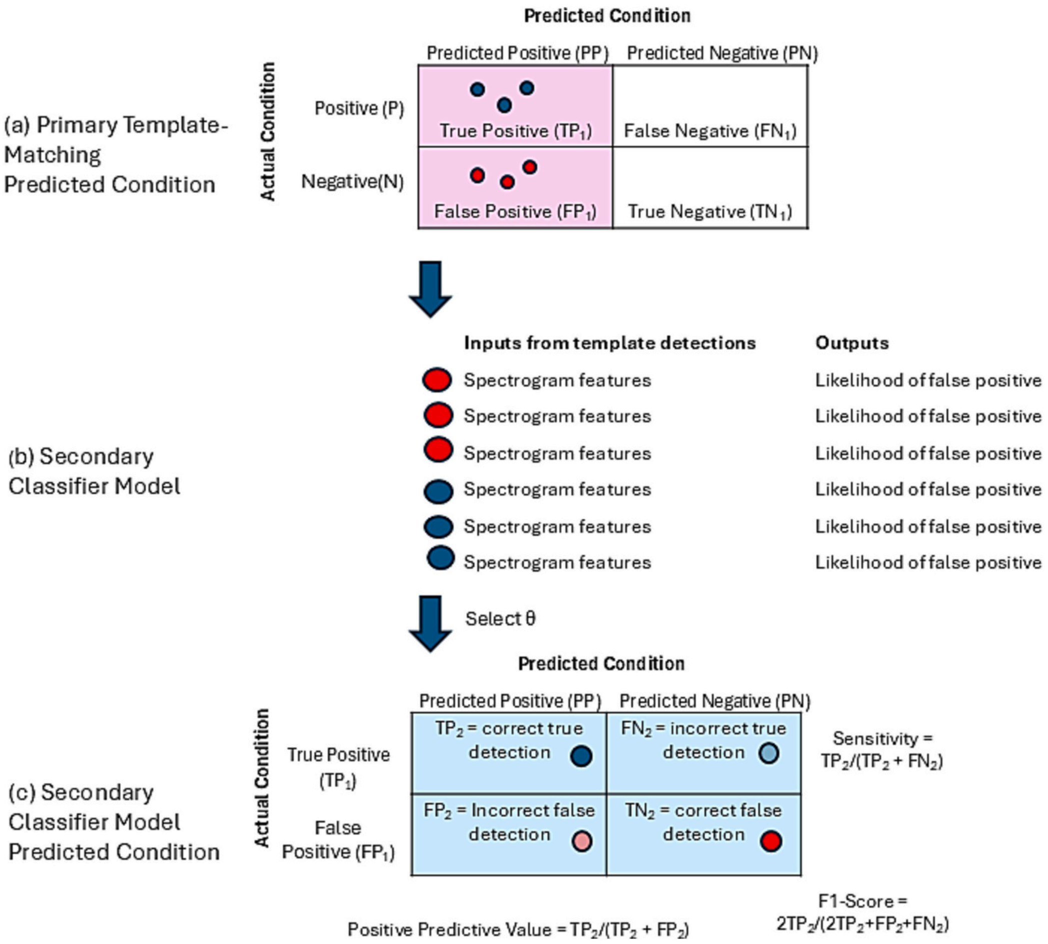

Fig. 2. Two-stage process using both template-matching (Stage 1) and secondary classifier models (Stage 2). Initial template detections are accumulated in Stage 1. Then, a subset of true (valid) detections and false positives (invalid detections) are used to train secondary machine learning classifiers to identify false positives based on audio features (Stage 2). Figure modified from Clarfeld et al., 2025a.

The Cuban treefrog, as a prominent invasive species in the southern United States, presents a need for timely surveillance and monitoring. Without an efficient, effective, and rapid method for detecting this species, management programs may be blind to where their eradication efforts are needed most. Dispersal of the Cuban treefrog across the United States and Canada is facilitated by human activities (Krysko et al., 2011; Morningstar et al., 2025), and new stable populations outside its current range in the coastal south are inevitable without prompt management actions. Effective Cuban treefrog surveillance programs using automated detection could narrow the gap between data collection and management decisions. Here, we employ the two-stage process using both template matching and supervised learning classifiers to improve the automated detection of the Cuban treefrog. Our objectives were to: 1) Create cross-correlation templates to screen all data for initial detections; 2) Verify a subset of these initial detections; 3) Train and test supervised learning classifier algorithms to filter false positives; and 4) Run the best-performing classifiers on the remaining model outputs. Throughout, we illustrate a simple yet effective method for automated detection of the Cuban treefrog and other sound-emitting species.

2. Methods

2.1. Objective 1: Create templates and screen data for initial detections

2.1.1. Data collection

A total of 139,985 recordings between 3 and 5 min in length (596,046 total minutes) were captured on a set schedule every hour for various survey periods between 2014 and 2022, by autonomous

Table 1
Summary of each location and its associated recordings.

Area	Location	ID	Years of Record	Number of Recordings
Audubon Park	New Orleans, Louisiana	AUDB01	2018–2022	14,414
		AUDB02	2018–2022	13,634
		AUDB03	2018–2022	11,176
Bayou Segnette State Park	Westwego, Louisiana	BASE01	2021–2022	4940
		BASE02	2021–2022	4921
Brechtel Park	New Orleans, Louisiana	BRPA01	2021–2022	3596
		BRPA02	2021–2022	4941
Gulf Islands National Seashore Naval Live Oaks Nature Preserve	Gulf Breeze, Florida	GUIS01	2014–2022	23,211
Jean Lafitte National Historical Park and Preserve	Marrero, Louisiana	JELA01	2014–2022	25,267
		JELA02	2014–2022	23,052
Atchafalaya National Wildlife Refuge	Lottie, Louisiana	LABY01	2020–2022	5465
Woodlands Conservancy	Belle Chasse, Louisiana	WOCO01	2022	2012
		WOCO02	2021–2022	3356

recording units placed across 12 locations in southern Louisiana and at 1 additional location near Pensacola, Florida (Table 1, Fig. 3).

2.1.2. *Create templates for Cuban treefrog*

The primary call of the Cuban treefrog has been described as a raspy snarl or rubbery snore (Dorcas and Gibbons, 2008; U.S. Fish and Wildlife Service, 2019; Morningstar et al., 2025), occurring in the approximate frequency range 1-6 kHz (Figure 4). As this call is largely conserved across individuals, populations, and locations, Cuban treefrogs are an ideal species to attempt to detect using template matching. Correlation templates were created via Fast Fourier Transform and use a matrix of amplitude values, taken directly from the spectrogram of the recording with the chosen example of the target species' emitted sound (Katz et al., 2016). Similarity of any new audio files to each template can be scored based on Pearson correlation coefficient, and any scores above a predetermined score threshold are considered “detections” of the target species (Figure 1 (A)). These detections, however, can include both true detections and false positives. Score thresholds are determined individually for each template during the creation process while assessing scores on test audio files, where a threshold is chosen that maintains as many true detections as possible while also excluding many false positives.We used the *create_templates* app in the *AMMonitor* R package's Shiny interface (Clarfeld et al., 2024; Clarfeld et al., 2025b) to create cross-correlation templates from examples of both distant and nearby Cuban treefrogs. Within the *create_templates* app, we generated spectrograms from 10 s clips at a 24 kHz sampling rate using short-time Fourier transforms with 512-sample Hann (Hanning) windows, 0% overlap, and an unfiltered 0–12 kHz analysis band. We developed several cross-correlation templates and corresponding score thresholds using randomized recordings in the database that already had Cuban treefrog calls annotated. While creating the templates, we manually adjusted the score thresholds by visually inspecting test detections on spectrograms from these recordings in the *create_templates* app and choosing values that appeared to balance true detections against obvious false positives, particularly in recordings that also contained other species' calls, wind, or anthropogenic noise. Finally, we compared templates based on how well they detected true Cuban treefrog calls in the example audio clips and how effectively they reduced false positives from similar sounds.

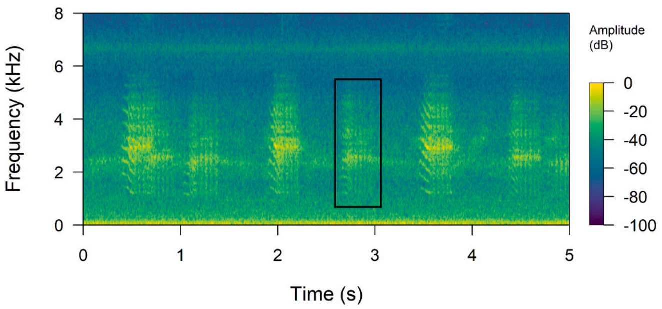

Fig. 4. Vocalization examples of two Cuban treefrogs (*Osteopilus septentrionalis*), with a bounding box around one note of a single frog's call.

The templates that best detected Cuban treefrog while achieving a balance of fewer false positives were saved as R objects in an *AMModels* library (Donovan and Katz, 2018), which included their score thresholds and other metadata. These templates were clear examples of Cuban treefrog calls, containing less noise than other templates in the sample, and also seemed to achieve a balance between identifying true detections of Cuban treefrog while maintaining a relatively low rate of false positives (in the example recordings used to test the templates). Score thresholds of 0.55 were assigned to both templates, meaning scored audio segments would need to exceed that threshold to be considered a detection. Further instructions for creating templates with *AMMonitor* can be found in the “models” tutorial of the R package.

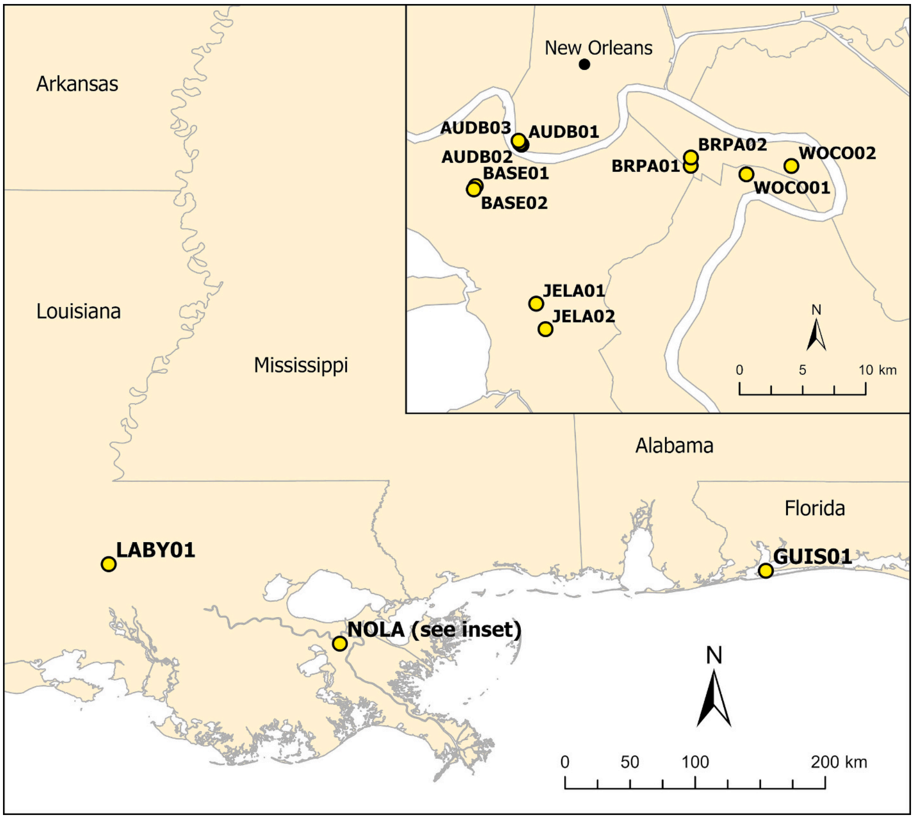

Fig. 3. Study locations where active monitoring occurred in Louisiana and Florida, during the period between 2014 and 2022.

2.1.3. *Accumulate initial detections using templates*

We used the *AMMonitor* function *scoresDetect()* to accumulate scores and detections across all collected recordings. Correlation scores were generated using the Pearson correlation coefficient (Katz et al., 2016) during a moving window analysis of each audio file. Peaks in the scores that exceeded the specified threshold for each template were considered detections (including both true detections and false positives) (Fig. 1 (A)). Detections for each template, including their scores and sonogram bounding boxes, were imported into the *modeloutputs* table of the *AMMonitor* database.

2.2. *Objective 2: verify subset of initial detections*

2.2.1. *Manually verify a subset of detections*

A portion of the detections from the initial template-screening process (template detections) were manually verified (labelled) by the authors (n > 3000). For each template, a random sample from each location was included, as well as all detections with correlation scores above 0.7 to include high-confidence detections (those detections that most-strongly correlated with the template). Further verifications were done to include representative samples of challenging false positive types.

We used the *AMMonitor* Shiny interface to mark the subset of template detections as either “valid” (a true positive detection; a Cuban treefrog call) or “invalid” (a false positive detection; not a Cuban treefrog) after listening to the audio clip and/or visually inspecting the sonogram. All verifications were saved to the *AMMonitor* database, where they were used to train and test the stage-two classification models (Objective 3).

2.3. *Objective 3: Train and test supervised learning classifiers to filter false positives*

2.3.1. *Train supervised learning classification algorithms*

In this objective, we trained our classifier models for the Stage 2 of this method, where audio features can be used to discern between true detections and false positives from the Stage 1 phase (screening audio data with templates) (Fig. 1 (B)). We used the package *seewave* (Sueur et al., 2008) to extract the raw amplitude matrix values and various acoustic summary characteristics (Table 2) for each labelled detection. These characteristics became the models' predictive features, or covariates, to distinguish between true and false positive detections. Values for the raw amplitude matrix, the time and frequency contours, and zero-crossing rates per time bin contained multiple values, each corresponding to a pixel in the amplitude matrix or a specific time bin. We made each individual value its own predictive feature to pass to the classifier models. Thus, more than 2240 features were used as inputs for the classifier models. We did not preprocess our data through feature scaling or transformation.

The package *caret* (Kuhn, 2008) was used to split the labelled data into the standard ratio for training (80%) and testing (20%) datasets (Gholamy et al., 2018) with even proportions of valid and invalid detections from all study locations, and to train each classifier model. For each template, we trained models with each of the following *caret* methods: regularized logistic regression (*glmnet)*, random forest (*rf*), and support vector machines with polynomial kernel (*svmPoly*). Regularized logistic regression fits a generalized linear model with a penalty on the coefficients to induce shrinkage and variable selection (Tibshirani, 1996), random forests aggregate predictions from an ensemble of decorrelated decision trees grown on bootstrap samples of the data (Breiman, 2001), and support vector machines with a polynomial kernel learn a maximum-margin decision boundary in a feature space defined by polynomial transformations of the predictors (Cortes and Vapnik, 1995). We used k-fold cross validation to train all models, with $k = 3$ to address the imbalance between the classes and after confirming similar results with k = 5. For each model type, we used a tuning grid search while training to select the optimal hyperparameter values.

Table 2

Description of audio features used in supervised learning classifiers, and the corresponding functions used to extract each feature.

Feature	Description	Function
Raw Amplitude Values	Individual raw amplitude values (in dB) from the matrix of a detection, generated via Fast Fourier Transform. In a spectrogram, amplitude is represented by the degree of shading in the image.	seewave::spectro()
Correlation Score	Correlation score of the detection, generated during template-matching via moving window analysis.	monitor::corMatch()
Time Contour for Each Time Bin	Amplitude probability mass function for each time bin, computed via sum aggregate function.	seewave::acoustat()
Frequency Contour for Each Frequency Bin	Amplitude probability mass function for each frequency bin, computed via sum aggregate function.	seewave::acoustat()
Time.P1	Time initial percentile, computed from time probability mass function.	seewave::acoustat()
Time.M	Time median, computed from time probability mass function.	seewave::acoustat()
Time.P2	Time terminal percentile, computed from time probability mass function.	seewave::acoustat()
Time.IPR	Time interpercentile range, computed from time probability mass function.	seewave::acoustat()
Frequency.P1	Frequency initial percentile, computed from Frequency probability mass function.	seewave::acoustat()
Frequency.M	Frequency median, computed from Frequency probability mass function.	seewave::acoustat()
Frequency.P2	Frequency terminal percentile, computed from Frequency probability mass function .	seewave::acoustat()
Frequency.IPR	Frequency interpercentile range, computed from Frequency probability mass function.	seewave::acoustat()
Zero-Crossing Rate for Each Time Bin	Zero-crossing rate (ZCR) per time bin, computed by:: zcr = 0.5 * mean(abs(sgn (x(t + 1)) – sgn(x(t)))) with: N the length of the signal x, and where: sgn(x(t)) = 1 if x(t) $\geq$ 0 and sgn(x(t)) = −1 if x(t) < 0. This measures how frequently a signal crosses the zero-amplitude axis. Tonal sounds or vocalizations tend to have lower ZCR, and noisy signals such as rainfall or wind have higher ZCR due to rapid amplitude fluctuations.	seewave::zcr ()
Spectral Mean	Mean frequency across detection's frequency spectrum.	seewave::specprop()
Spectral Standard Deviation	Standard deviation of the mean frequency.	seewave::specprop()
Spectral SEM	Standard error of the mean frequency.	seewave::specprop()
Spectral Median	Median frequency.	seewave::specprop()
Spectral Mode	Mode (dominant) frequency.	seewave::specprop()
Q1: First Quartile	First quartile (0.25 quantile) of the frequency spectrum.	seewave::specprop()
Q3: Third Quartile	Third quartile (0.75 quantile) of the frequency spectrum.	seewave::specprop()

2.3.2. *Assess classifier performance on the testing data*

Models with optimized hyperparameter values were then evaluated on the unseen testing data (20%), the portion of verified template detections that were not used to train the models. Performance was assessed by inspecting the receiver-operating characteristic (ROC) curve, which plots the true positive rate (sensitivity) against the specificity (1 - false positive rate), and on seven metrics of performance calculated from the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) in the confusion matrix:

overall accuracy, sensitivity (also called recall), specificity, positive predictive value (or precision), negative predictive value, F1 score, and Matthews Correlation Coefficient (MCC).

Accuracy ($\frac{TP+TN}{TP+TN+FP+FN}$) describes the overall rate of correct predictions made by a model. Sensitivity ($\frac{TP}{TP+FN}$) measures how well a model accurately identifies valid detections, out of all the actually valid detections in the dataset. Specificity ($\frac{TN}{TN+FP}$) can be thought of as the true negative rate, or the rate of correctly identified invalid detections. Positive predictive value, or precision, ($\frac{TP}{TP+FP}$) denotes how many of the predictions classified as “valid” were correct, while negative predictive value ($\frac{TN}{TN+FN}$) describes the same for “invalid” classifications. We were particularly interested in each model's F1 score and MCC. The MCC ($\frac{TP\times TN-FP\times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$) provides a balanced representation of the model performance, especially with imbalanced training datasets such as those we were working with (Chicco and Jurman, 2023). A high MCC score (close to 1) represents a model that maximizes performance in all four basic rates in the confusion matrix (sensitivity, specificity, positive predictive value, and negative predictive value). The F1 score ($\frac{2TP}{2TP+FP+FN}$), on the other hand, describes the balance between high sensitivity and high positive predictive value. Essentially, a high F1 score represents a model that catches the majority of detections while maintaining a low rate of false positives. This is an important metric to examine when improving automated detection, as the aim is to maintain high sensitivity to valid detections with a low rate of false positives (Balantic and Donovan, 2020).

2.4. *Objective 4: Run best-performing classifiers on remaining model outputs*

2.4.1. *Run classifiers on remaining template detections*

We chose the classifier models with the highest F1 score for each template for this final step. We extracted the audio features of the remaining (unverified) template detections using the *seewave* package (Sueur et al., 2008), then fed these features to each of the final classifier models. Final detections were considered those model outputs predicted to be class “valid” by the classifier models.

3. Results

3.1. *Objective 1: create templates and screen data for initial detections*

After testing various templates and corresponding score thresholds, two correlation templates (COR1 and COR2) were saved and used to initially process all audio files. A total of 83,191 detections and 173,530 detections were generated from the collected media with the COR1 and COR2 templates, respectively. These template detections were present across 12 locations and 2499 (COR1) and 4472 (COR2) media files.

3.2. *Objective 2: verify subset of initial detections*

The verified datasets for COR1 and COR2 consisted of 3250 (1089 valid detections) and 3241 (515 valid detections) template detections, respectively. It took approximately 15 combined hours to manually verify these two subsets. All of the valid detections from these verifications were found only at the three units deployed in Audubon Park-within this sample, where Cuban treefrog presence had been previously confirmed: AUDB01, AUDB02, and AUDB03. Common false positives at these sites and the others included car or airplane noises, other treefrogs, bird calls, human voices or music, fireworks, and rain.

3.3. *Objective 3: train and test supervised learning classifiers to filter false positives*

Optimal hyperparameters were selected via tuning grid when training the regularized logistic regression (*glmnet)*, support vector machine with polynomial kernel (*svmPoly),* and random forest (*rf*) models for each correlation template (Table 3). ROC curves and their accompanying area-under-the-curve (AUC) values were generated using the testing data and showed excellent separation for most of the classifiers, especially for those of the COR1 template (Fig. 5). The best models by AUC for COR1 were *svmPoly* (AUC = 0.998) and *rf* (AUC = 0.998). For COR2, the best model was *svmPoly* (AUC = 0.982).

Overall, the classifiers for the COR1 template had higher performance across most metrics (Table 4). For this template's classifiers, all models had quite high specificity with values above 98%. The *svmPoly* classifier outperformed the other two models in all metrics, especially in sensitivity (Sensitivity = 0.9771), F1 score (F1 = 0.9771), and MCC (MCC = 0.9655). Overall, the COR1 models seemed to discern between valid detections and false positives very well. The *svmPoly* model for COR1 had a false positive rate of 1.116%. The probability of missing a true detection, or the false negative rate, was 2.29%.

The COR2 template's classifiers overall did not perform as well as those of COR1. They did well in the specificity metric, with all models measuring above 96% (Table 4). Negative predictive value was also relatively high, however other metrics like positive predictive value, sensitivity, and F1 score were lower. The best model by F1 score and MCC was the *svmPoly* (F1 = 0.8517; MCC = 0.8233), which had a false positive rate of 3.112%. This best model's probability of missing a true detection (false negative rate) was much higher than the best COR1 model, at 13.59%. All of these suggest the models for this template had an easier time accurately identifying the invalid detections, rather than the valid, and that it missed a larger proportion of true detections while including more false positives in predicted detections.

3.4. *Objective 4: run best-performing models on remaining model outputs*

Because the *svmPoly* models achieved the highest F1 scores for both templates, these classifiers were used to process the remaining template-screened detections. The two templates (COR1 and COR2) were built from different audio clips and therefore produced different detection rates and false positives, so the performance metrics were only comparable between classifiers applied to the same template.

The final number of predicted detections in the remaining dataset were 20,271 detections (COR1) and 18,173 detections (COR2). Predicted detections were found, combining both models, across 10 locations and 1,114 media files. The detections were distributed mainly across the 3 Audubon locations with confirmed Cuban treefrog presence (AUDB01, AUDB02, AUDB03) (Figure 6). Across all other study sites combined, where Cuban treefrog presence had yet to be confirmed, COR1 predicted only 109 detections, and COR2 predicted a mere 58 detections. AUDB01, AUDB02, and AUDB03 were the only sites with valid detections following manual validation of all putative detections at the other sites. These results match expectations from in-person field work at all sites; Cuban treefrogs were, as of this study, only ever confirmed at the three Audubon locations in this sample.

Table 3
Hyperparameter values for the best classifier models, found while training using a tuning grid search.

Model	Best Tune
COR1	
glmnet	alpha = 0.3; lambda = 0.01
svmPoly	degree = 2; scale = 0.05; C = 0.1325
rf	mtry = 25
COR2	
glmnet	alpha = 0.3; lambda = 0.01
svmPoly	degree = 2; scale = 0.05; C = 0.255
rf	mtry = 25

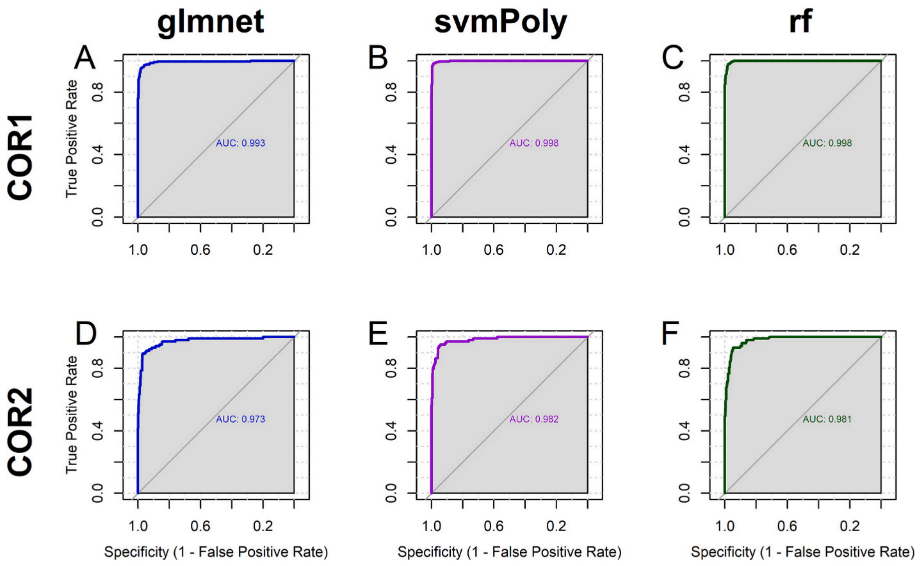

Fig. 5. Receiver-operating characteristic (ROC) curves for each model, generated using the testing dataset, and their associated Area Under the Curve values (AUC). COR1 and COR2 refer to the classifiers trained on data from the COR1 or COR2 template outputs, respectively. Regularized logistic regression (*glmnet*) model results are shown in panels A (COR1) and D (COR2). Support vector machine with polynomial kernel (*svmPoly*) results are shown in panels B (COR1) and E (COR2). Random forest (*rf*) results are shown in panels C (COR1) and F (COR2).

Table 4

Testing performance of supervised learning classifiers trained to filter invalid detections from both templates (COR1 and COR2). Performance metrics include overall Accuracy, Sensitivity (or recall), Specificity, Positive Predictive Value (PPV, also known as precision), Negative Predictive Value (NPV), F1 Score, and Matthew's Correlation Coefficient (MCC).

Model	Accuracy	Sensitivity	Specificity	PPV	NPV	F1	MCC
COR1							
glmnet	0.9692	0.9404	0.9838	0.9670	0.9703	0.9535	0.9307
svmPoly	0.9846	0.9771	0.9884	0.9771	0.9884	0.9771	0.9655
rf	0.9692	0.9358	0.9861	0.9714	0.9682	0.9533	0.9307
COR2							
glmnet	0.9444	0.7476	0.9817	0.8851	0.9537	0.8105	0.7821
svmPoly	0.9522	0.8641	0.9688	0.8396	0.9742	0.8517	0.8233
rf	0.9398	0.7282	0.9798	0.8721	0.9502	0.7937	0.7630

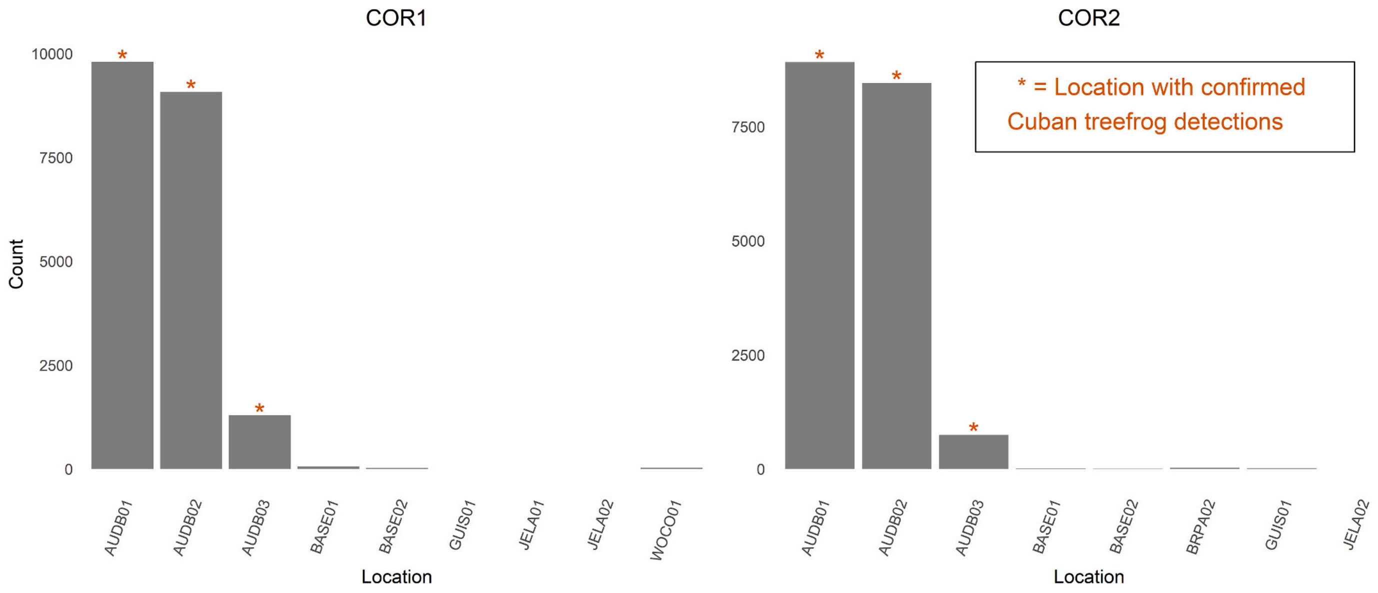

Fig. 6. Number of Cuban treefrog *(Osteopilus septentrionalis*) detections predicted by the two-stage (template and classifier) model approach at each location in the final dataset. Study locations without detections were excluded.

4. Discussion

Understanding the current Cuban treefrog distribution is vital to conservation efforts that aim to prevent further establishment of this species in new locations. The two-stage method, using both template matching and machine learning classifiers, that we have demonstrated here vastly reduces the effort needed to progress from collected audio recordings to verified presence-absence data. For surveillance, only a brief investment of time is required to verify detections at each location and generate a labelled dataset for training the classifier algorithms—in our study, this process took approximately 15 h in total. More importantly, this method proved highly effective for our use-case by providing high-confidence detections of this invasive species across monitored locations. Notably, these results were achieved using a small training dataset and without the need for complex neural network architecture, computationally intensive training, or paid subscription services for user-friendly machine learning software.

Overall, the two-stage method here produced accurate detections of Cuban treefrogs where present, and a very low and easily verifiable count of false positives at sites where they were apparently absent. While we cannot fully quantify the improvement of this method on template-matching alone, since we do not know how often the templates missed true detections (false negative rate), we can estimate the secondary classifiers effectiveness from metrics like F1 score. The best-performing classifier (*svmPoly* for the COR1 template) had an excellent testing F1 score of 0.9771, meaning the model correctly identified almost all true positives within the template-screened testing data, while also minimizing false positives and false negatives. When applying the two-stage approach to the full dataset, nearly all detections—20,271 from COR1 and 18,173 from COR2—were concentrated at the three sites where Cuban treefrog presence was confirmed. Only 109 detections (COR1) and 58 detections (COR2) occurred at other sites, and these were rapidly and easily verified as false positives in under an hour. We used the R package, *AMMonitor*, as the primary vehicle for utilizing the two-stage approach. *AMMonitor* is a freely available, open-source R package for storing and processing wildlife monitoring data (Balantic and Donovan, 2020; Clarfeld et al., 2024). *AMMonitor* builds on other R packages, including *monitoR* (Hafner and Katz, 2018; Katz et al., 2016) to incorporate binary or cross-correlation templates and *AMModels* (Donovan and Katz, 2018) to store models and their metadata so they can be easily reused in the future. Generating and using templates can be done with a user-friendly app (*create_templates*) and function (*scoresDetect()*). Software alternatives to *AMMonitor* certainly exist for template matching, such as another freely available software platform *ARBIMON* (Aide et al., 2013). However, the *AMMonitor* package allows consolidation of the entire workflow involved with the two-stage method we present. *AMMonitor*'s Shiny interface facilitates efficient verification of template-screened detections used for stage 2, and the *caret* machine learning models can be stored in accompanying *AMModel* libraries for continued use. All model outputs are stored in a relational database with a standardized structure that provides an organized data management system and naturally facilitates collaboration with other *AMMonitor* projects in a time when autonomous monitoring is increasingly prevalent.

Several considerations remain. First, although our methodology included a wide range of false positive sounds and produced relatively accurate species detection across these locations, it is important to note that our training and testing data were from the same sites. For example, one site in our sample had frequent fireworks, which were scored with high confidence as Cuban treefrog detections during the template screening phase. Had our training/testing data not included firework sounds from that one site, there may have been a much higher false positive rate of the final detections there. Further testing of our best models on data from locations that were not included in the training data will show how well the models generalize to new sounds.Second, while effective for Cuban treefrog, this method may not apply well to all species. The Cuban treefrog has a distinctive call that is largely conserved across individuals, locations, and environments. Some species exhibit individual acoustic signatures, as seen in fish (Amorim and Vasconcelos, 2008), anurans (Bee and Gerhardt, 2001), birds (Terry et al., 2005), and mammals (Taylor and Reby, 2010), which require additional techniques to address (Stowell et al., 2019). Many species have large vocal repertoires, requiring models to be trained on multiple calls. Others may have different repertoires per population, such as the differing dialects across orca whale (*Orcinus orca*) pods (Deecke et al., 2000) or populations of white-crowned sparrow (*Zonotrichia leucophrys*) (Nelson, 2000). Some species may also overlap in frequency with other common sounds in the environment (such as some forms of anthropogenic noise), so representative training data that contains different types of background noise will make any models more robust.

The two-stage method using templates and machine learning classifiers that we have highlighted here is one of many approaches to automated detection with modern technology, and continuing research may illuminate how effective and efficient this approach is compared with other contemporary tools. The benefits of this approach, namely a free, open-source, and user-friendly platform for template creation and rapid model output validation, and simplicity of machine learning methods, are similar to other user-friendly software available at little to no cost to researchers. A formal benchmarking comparison against available tools like Arbimon (Aide et al., 2013), BirdNET (Kahl et al., 2021), and Perch 2.0 (van Merriënboer et al., 2025), which do not require programming skills or machine learning expertise, may show if the template-classifier method in *AMMonitor* is an attractive option for small wildlife monitoring projects. Additionally, further work exploring the false negative rate with a fully labelled fresh dataset would help quantify how many detections might be missed with this approach to expand its use from invasive wildlife surveillance to true population monitoring, which may greatly improve the impact of this technique in ecological research.

We have demonstrated that this two-stage method for automated detection can be an efficient tool in surveillance programs for Cuban treefrog in the southern United States. The best template-classifier model pairs from this work can be used moving forward to rapidly detect any movement into new areas and allow managers to promptly respond with removal actions. This fast and efficient method reduces barriers to adoption of a larger network of early detection recorders for Cuban treefrog and potentially other species. Our future goals will be to increase automation of the process and expand the number of sites monitored.

CRediT authorship contribution statement

Kaitlin Huber: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation. **J. Hardin Waddle:** Writing – review & editing, Visualization, Resources, Funding acquisition, Data curation. **Brad M. Glorioso:** Writing – review & editing, Visualization, Resources, Data curation. **Therese Donovan:** Writing – review & editing, Validation, Supervision, Software, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Therese Donovan reports financial support was provided by U.S. Geological Survey. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to thank Katherina Gieder, Mark Henderson, and Laura May-Collado for their review and feedback on drafts of this manuscript. We also thank the many biologists who have helped collect these recordings in the field. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U. S. Government. The Vermont Cooperative Fish and Wildlife Research Unit is jointly supported by the U.S. Geological Survey, University of Vermont, Vermont Department of Fish and Wildlife, U.S. Fish and Wildlife Service, and Wildlife Management Institute. This is contribution number 986 of the U.S. Geological Survey Amphibian Research and Monitoring Initiative (ARMI).

Data availability

All data and media files from this project are available for download from the USGS ScienceBase Repository at https://www.sciencebase.gov/catalog/item/691cee88d4be021d1d89b3fd (Waddle et al., 2025). Code to run the analysis can be found at https://code.usgs.gov/cooperativeresearchunits/vermont/cuban-treefrog (Huber and Donovan, 2025).

References

Acevedo, M.A., Villanueva-Rivera, L.J., 2006. From the field: using automated digital recording systems as effective tools for the monitoring of birds and amphibians. Wildl. Soc. Bull. 34, 211–214. https://doi.org/10.2193/0091-7648(2006)34%5B211:UADRSA%5D2.0.CO;2.

Acevedo, M.A., Corrada-Bravo, C.J., Corrada-Bravo, H., Villanueva-Rivera, L.J., Aide, T. M., 2009. Automated classification of bird and amphibian calls using machine learning: a comparison of methods. Eco. Inform. 4, 206–214. https://doi.org/10.1016/j.ecoinf.2009.06.005.

Aide, T.M., Corrada-Bravo, C., Campos-Cerqueira, M., Milan, C., Vega, G., Alvarez, R., 2013. Real-time bioacoustics monitoring and automated species identification. PeerJ 1, e103. https://doi.org/10.7717/peerj.103.

Amorim, M.C.P., Vasconcelos, R.O., 2008. Variability in the mating calls of the Lusitanian toadfish *Halobatrachus didactylus*: cues for potential individual recognition. J. Fish Biol. 73, 1267–1283. https://doi.org/10.1111/j.1095-8649.2008.01974.x.

Balantic, C.M., Donovan, T.M., 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.

Bee, M.A., Gerhardt, H.C., 2001. Neighbour–stranger discrimination by territorial male bullfrogs (*Rana catesbeiana*): I. Acoustic basis. Anim. Behav. 62, 1129–1140. https://doi.org/10.1006/anbe.2001.1851.

Bravo, C.J.C., Berríos, R.Á., Aide, T.M., 2017. Species-specific audio detection: a comparison of three template-based detection algorithms using random forests. PeerJ Comput. Sci. 3, e113. https://doi.org/10.7717/peerj-cs.113.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Brosnan, E.B., Paniagua Torres, K.A., Martin, K.R., Atkinson, M.S., Glorioso, B.M., Waddle, J.H., Mendyk, R.W., Savage, A.E., 2025. Tracing invasion routes of Cuban treefrogs into Louisiana using mitochondrial DNA. Front. Amphib. Reptile Sci. 3. https://doi.org/10.3389/famrs.2025.1646714.

Chicco, D., Jurman, G., 2023. The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification. BioData Mining. 16, 4. https://doi.org/10.1186/s13040-023-00322-4.

Clarfeld, L., Tang, C., Huber, K., Balantic, C., Dillon, K., Donovan, T., 2024. AMMonitor: Remote monitoring of biodiversity in an adaptive framework. In: Version 2.1.0: U.S. Geological Survey Software Release. Reston, VA. https://doi.org/10.5066/P13MRDRV.

Clarfeld, L.A., Gieder, K.D., Abrams, R., Bernier, C., Cahill, J., Staats, S., Wixsom, S., Donovan, T.M., 2025a. Two-stage models improve machine learning classifiers in wildlife research: a case study in identifying false positive detections of ruffed grouse. Ecol. Inform., 103166 https://doi.org/10.1016/j.ecoinf.2025.103166.

Clarfeld, L., Tang, C., Huber, K., Balantic, C., Donovan, T., 2025b. AMMonitor 2: remote monitoring of biodiversity in an adaptive framework in r. Methods Ecol. Evol. 16, 477–485. https://doi.org/10.1111/2041-210X.14487.

Coffey, K.R., Marx, R.E., Neumaier, J.F., 2019. DeepSqueak: a deep learning-based system for detection and analysis of ultrasonic vocalizations. Neuropsychopharmacology 44, 859–868. https://doi.org/10.1038/s41386-018-0303-6.

Colonna, J., Peet, T., Ferreira, C.A., Jorge, A.M., Gomes, E.F., Gama, J., 2016. Automatic classification of anuran sounds using convolutional neural networks. In: C3S2E ‘16. Association for Computing Machinery, New York, NY, USA, p. 7378. https://doi.org/10.1145/2948992.2949016.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297. https://doi.org/10.1007/BF00994018.

Cortopassi, K.A., Bradbury, J.W., 2000. The comparison of harmonically rich sounds using spectrographic cross-correlation and principal coordinates analysis. Bioacoustics 11, 89127. https://doi.org/10.1080/09524622.2000.9753454.

Deecke, V.B., Ford, J.K.B., Spong, P., 2000. Dialect change in resident killer whales: implications for vocal learning and cultural transmission. Anim. Behav. 60, 629–638. https://doi.org/10.1006/anbe.2000. 1454.

Donovan, T.M., Katz, J.E., 2018. AMModels: an r package for storing models, data, and metadata to facilitate adaptive management. PLoS One 13, e0188966. https://doi.org/10.1371/journal.pone.0188966.

Dorcas, M.E., Gibbons, W., 2008. Frogs and Toads of the Southeast, Wormsloe Foundation Nature Book. University of Georgia Press, Athens.

Eichinski, P., Alexander, C., Roe, P., Parsons, S., Fuller, S., 2022. A convolutional neural network bird species recognizer built from little data by iteratively training, detecting, and labeling. Front. Ecol. Evol. 10. https://doi.org/10.3389/fevo.2022.810330.

Ghani, B., Denton, T., Kahl, S., Klinck, H., 2023. Global birdsong embeddings enable superior transfer learning for bioacoustic classification. Sci. Rep. 13, 22876. https://doi.org/10.1038/s41598-023-49989-z.

Gholamy, A., Kreinovich, V., Kosheleva, O., 2018. Why 70/30 or 80/20 Relation between Training and Testing Sets: A Pedagogical Explanation. Departmental Technical Reports (CS). https://scholarworks.utep.edu/cs_techrep/1209/.

Gibb, R., Browning, E., Glover-Kapfer, P., Jones, K.E., 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.

Glorioso, B.M., Waddle, J.H., Crockett, M.E., Rice, K.G., Percival, H.F., 2010. Diet of the invasive Cuban treefrog (*Osteopilus septentrionalis*) in pine Rockland and mangrove habitats in South Florida. Caribb. J. Sci. 46 (346), 355. https://doi.org/10.18475/cjos.v46i2.a25.

Glorioso, B.M., Waddle, J.H., Muse, L.J., Jennings, N.D., Litton, M., Hamilton, J., Gergen, S., Heckard, D., 2018. Establishment of the exotic invasive Cuban treefrog (*Osteopilus septentrionalis*) in Louisiana. Biol. Invasions 20, 2707–2713. https://doi.org/10.1007/s10530-018-1732-1.

Godfray, H.C.J., Knapp, S., Gaston, K.J., O'Neill, M.A., 2004. Automated species identification: why not? Philosophical transactions of the royal society of London. Ser. B Biol. Sci. 359, 655–667. https://doi.org/10.1098/rstb.2003.1442.

Grießmann, B., Naguib, M., 2002. Song sharing in neighboring and non-neighboring thrush nightingales (*Luscina luscina*) and its implications for communication. Ethology 108, 377–387. https://doi.org/10.1046/j.1439-0310. 2002.00781.x.

Guerrero, M.J., Bedoya, C.L., López, J.D., Daza, J.M., Isaza, C., 2023. Acoustic animal identification using unsupervised learning. Methods Ecol. Evol. 14, 1500–1514. https://doi.org/10.1111/2041-210X.14103.

Hafner, S.D., Katz, J., 2018. MonitoR: Acoustic Template Detection in r. Version 1.0.7.

Huber, K., Donovan, T., 2025. Code for Two-stage Approach to Automatic Detection with Machine Learning for Improved Surveillance of the Invasive Cuban Treefrog, Version 1.0.0: U.S. Geological Survey Software Release. https://doi.org/10.5066/P14NZ2AH.

Hutto, R.L., Stutzman, R.J., 2009. Humans versus autonomous recording units: a comparison of point-count results. J. Field Ornithol. 80, 387–398. https://doi.org/10.1111/j.1557-9263.2009.00245.x.

Kahl, S., Wood, C.M., Eibl, M., Klinck, H., 2021. BirdNET: a deep learning solution for avian diversity monitoring. Eco. Inform. 61, 101236.

Katz, J., Hafner, S.D., Donovan, T., 2016. Assessment of error rates in acoustic monitoring with the r package monitoR. Bioacoustics 25, 177–196. https://doi.org/10.1080/09524622.2015.1133320.

Krysko, K.L., Burgess, J.P., Rochford, M.R., Gillette, C.R., Cueva, D., Enge, K.M., Somma, L.A., Stabile, J.L., Smith, D.C., Wasilewski, J.A., Kieckhefer, G.N.I., Granatosky, M.C., Nielsen, S.V., 2011. Verified non-indigenous amphibians and reptiles in Florida from 1863 through 2010: outlining the invasion process and identifying invasion pathways and stages. Zootaxa 3028, 1. https://doi.org/10.11646/zootaxa.3028.1.1.

Kuhn, M., 2008. Building predictive models in r using the caret package. J. Stat. Softw. 28, 126. https://doi.org/10.18637/jss.v028.i05.

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.

MacIsaac, J., Newson, S., Ashton-Butt, A., Pearce, H., Milner, B., 2024. Improving acoustic species identification using data augmentation within a deep learning framework. Eco. Inform. 83, 102851. https://doi.org/10.1016/j.ecoinf.2024.102851.

Mellinger, D.K., Clark, C.W., 2000. Recognizing transient low-frequency whale sounds by spectrogram correlation. J. Acoust. Soc. Am. 107, 35183529. https://doi.org/10.1121/1.429434.

Meshaka, W., 2001. The Cuban Treefrog in Florida– Life History of a Successful Colonizing Species. University Press of Florida.

Morningstar, C.R., Daniel, W.M., Somma, L.A., 2025. *Osteopilus septentrionalis* (Duméril and Bibron, 1841). U.S. Geological Survey, Nonindigenous Aquatic Species Database, Gainesville, FL. Last updated 12 June 2024. https://nas.er.usgs.gov/queries/FactSheet.aspx?SpeciesID=57.

Nelson, D.A., 2000. Song overproduction, selective attrition and song dialects in the white-crowned sparrow. Anim. Behav. 60, 887898. https://doi.org/10.1006/anbe.2000.1560.

Otesbelgue, A., Orth, A.J., Fong, C.D., Fassbinder-Orth, C.A., Blochtein, B., et al., 2025. Hidden Markov model for acoustic pesticide exposure detection and hive identification in stingless bees. PLoS One 20 (6), e0325732. https://doi.org/10.1371/journal.pone.0325732.

Ozanich, E., Thode, A., Gerstoft, P., Freeman, L.A., Freeman, S., 2021. Deep embedded clustering of coral reef bioacousticsa. J. Acoust. Soc. Am. 149, 2587–2601. https://doi.org/10.1121/10. 0004221.

Rice, K.G., Waddle, J.H., Miller, M.W., Crockett, M.E., Mazzotti, F.J., Percival, H.F., 2011. Recovery of native treefrogs after removal of nonindigenous Cuban treefrogs, *Osteopilus septentrionalis*. Herpetologica 67, 105117. https://doi.org/10.1655/HERPETOLOGICA-D-10-00020.1.

Schneider, S., Hammerschmidt, K., Dierkes, P.W., 2022. Introducing the software CASE (cluster and analyze sound events) by comparing different clustering methods and audio transformation techniques using animal vocalizations. Animals 12 (16), 2020. https://doi.org/10.3390/ani12162020.

Schwartz, A., 1952. *Hyla septentrionalis* duméril and bibron on the florida mainland. Copeia 117–118. https://doi.org/10.2307/1438553.

Shonfield, J., Bayne, E.M., 2017. Autonomous recording units in avian ecological research: current use and future applications. Avian Conserv. Ecol. 12. https://doi.org/10.5751/ACE-00974-120114.

Smirnov, E., 2013. North Atlantic right whale call detection with convolutional neural networks. Citeseer 7879.

Stowell, D., Wood, M., Stylianou, Y., Glotin, H., 2016. Bird detection in audio: A survey and a challenge. In: IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP), Vietri Sul Mare, Italy, pp. 1–6. https://doi.org/10.1109/MLSP.2016.7738875.

Stowell, D., Petrusková, T., Šálek, M., Linhart, P., 2019. Automatic acoustic identification of individuals in multiple species: improving identification across recording conditions. J. R. Soc. Interface 16, 20180940. https://doi.org/10.1098/rsif.2018.0940.

Sueur, J., Aubin, T., Simonis, C., 2008. Seewave, a free modular tool for sound analysis and synthesis. Bioacoustics 18, 213226. https://doi.org/10.1080/09524622.2008.9753600.

Taylor, A.M., Reby, D., 2010. The contribution of source–filter theory to mammal vocal communication research. J. Zool. 280, 221236. https://doi.org/10.1111/j.1469-7998.2009.00661.x.

Tennessen, J.B., Parks, S.E., Tennessen, T.P., Langkilde, T., 2016. Raising a racket: invasive species compete acoustically with native treefrogs. Anim. Behav. 114, 5361. https://doi.org/10.1016/j.anbehav.2016.01.021.

Terry, A.M., Peake, T.M., McGregor, P.K., 2005. The role of vocal individuality in conservation. Front. Zool. 2, 116. https://doi.org/10.1186/1742-9994-2-10.

Thompson, M.C., Ducey, M.J., Gunn, J.S., Rowe, R.J., 2025. A post-processing framework for assessing BirdNET identification accuracy and community composition. Ibis 167, 530–542. https://doi.org/10.1111/ibi.13357.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B Stat Methodol. 58, 267–288.

Towsey, M., Planitz, B., Nantes, A., Wimmer, J., Roe, P., 2012. A toolbox for animal call recognition. Bioacoustics 21, 107125. https://doi.org/10.1080/09524622.2011.648753.

U.S. Fish and Wildlife Service, 2019. Cuban treefrog (*Osteopilus septentrionalis*) Ecological Risk Screening Summary. https://www.fws.gov/sites/default/files/documents/Ecological-Risk-Screening-Summary-Cuban-Treefrog.pdf.

van Merriënboer, B., Dumoulin, V., Hamer, J., Harrell, L., Burns, A., Denton, T., 2025. Perch 2.0: The Bittern lesson for bioacoustics. arXiv preprint arXiv:2508.04665.

Venier, L.A., Holmes, S.B., Holborn, G.W., Mcilwrick, K.A., Brown, G., 2012. Evaluation of an automated recording device for monitoring forest birds. Wildl. Soc. Bull. 36, 30–39. https://doi.org/10.1002/wsb.88.

Waddle, J.H., Dorazio, R.M., Walls, S.C., Rice, K.G., Beauchamp, J., Schuman, M.J., Mazzotti, F.J., 2010. A new parameterization for estimating co-occurrence of interacting species. Ecol. Appl. 20, 1467–1475. https://doi.org/10.1890/09-0850.1.

Waddle, H., Huber, K.E., Donovan, T.M., 2025. USGS Cuban Treefrog Invasion Front Volume 1 (2014–2022): U.S. Geological Survey Data Release. https://doi.org/10.5066/P14NO6UR.

Yao, R., Johnson, M., Clemins, P., Michael, D., Osiejuk, T., Ebenezer, O.N., 2009. A framework for bioacoustic vocalization analysis using hidden markov models. Algorithms 2, 1410–1428. https://doi.org/10.3390/a2041410.