A comparison of supervised learning techniques in the classification of bat echolocation calls

David W. Armitage [a,1], Holly K. Ober [b,*]

[a] University of Florida, Department of Wildlife Ecology and Conservation, 110 Newins-Ziegler Hall, Gainesville, FL, 32611, USA
[b] University of Florida, Department of Wildlife Ecology and Conservation, North Florida Research and Education Center, 155 Research Road, Quincy, FL, 32351, USA

ARTICLE INFO

ABSTRACT

Today's acoustic monitoring devices are capable of recording and storing tremendous amounts of data. Until recently, the classification of animal vocalizations from field recordings has been relegated to qualitative approaches. For large-scale acoustic monitoring studies, qualitative approaches are very time-consuming and suffer from the bias of subjectivity. Recent developments in supervised learning techniques can provide rapid, accurate, species-level classification of bioacoustics data. We compared the classification performances of four supervised learning techniques (random forests, support vector machines, artificial neural networks, and discriminant function analysis) for five different classification tasks using bat echolocation calls recorded by a popular frequency-division bat detector. We found that all classifiers performed similarly in terms of overall accuracy with the exception of discriminant function analysis, which had the lowest average performance metrics. Random forests had the advantage of high sensitivities, specificities, and predictive powers across the majority of classification tasks, and also provided metrics for determining the relative importance of call features in distinguishing between groups. Overall classification accuracy for each task was slightly lower than reported accuracies using calls recorded by time-expansion detectors. *Myotis* spp. were particularly difficult to separate; classifiers performed best when members of this genus were combined in genus-level classification and analyzed separately at the level of species. Additionally, we identified and ranked the relative contributions of all predictor features to classifier accuracy and found measurements of frequency, total call duration, and characteristic slope to be the most important contributors to classification success. We provide recommendations to maximize accuracy and efficiency when analyzing acoustic data, and suggest an application of automated bioacoustics monitoring to contribute to wildlife monitoring efforts.

1. Introduction

Wildlife biologists and managers interested in monitoring animals often face many logistical challenges. In an effort to mediate these difficulties, a number of technologies have been developed to collect data on animals remotely, such as satellite tags, motion-detecting cameras, and acoustic recording devices (Block et al., 1998; Karanth and Nichols, 1998). For example, understanding of wildlife populations and communities has been furthered by the use of acoustic recorders for remotely monitoring vocal organisms such as whales, frogs, birds, and bats (Laiolo, 2010). Microchiropteran bats are especially good candidates for acoustic monitoring, due to challenges in surveying these secretive, nocturnal, volant mammals (Hayes et al., 2009; Parsons and Szewczak, 2009).

Three types of acoustic devices are currently suitable for the detection of bat echolocation calls: heterodyne, frequency-division, and time-expansion (Limpens, 2004). *Heterodyne* detectors take incoming waveforms and mix them with internal oscillations (Andersen and Miller, 1977). The difference in frequencies between the two waveforms is amplified and played back. Thus, researchers must prioritize a particular frequency range and tune the internal oscillators accordingly so that the detected calls fall within the range of human hearing (typically between 12 Hz and 20 kHz). *Frequency-division* detectors circumvent this problem of "selective hearing" by recording echolocation calls of all frequencies and synthesizing a pitch-shifted artificial pulse for each using an oscillation defined by the number of times the detected call's waveform crosses the zero-axis (a technique dubbed "zero-crossings analysis"). The main disadvantage of these devices is the loss of harmonic and amplitude information (only the dominant harmonic is preserved), which hinders the observer's ability to identify which bat species produced a recorded call. *Time-expansion* detectors record the entire echolocation waveform and resample it at a lower sampling rate. This

* Corresponding author. Tel.: +1 850 875 7150.
  E-mail addresses: Dave.Armitage@ufl.edu (D.W. Armitage), Holly.Ober@ufl.edu (H.K. Ober).
  [1] Tel.: +1 248 736 4174.

technique shifts the call's frequency down to the range audible to humans while preserving all information within a call, including harmonics and amplitude. Recordings made using these detectors are the only type suitable for conducting research on the audio qualities of echolocation calls. The main drawback of this system is that the detector is not recording any incoming information while calls are being resampled, which could potentially exclude rare species from detection.

Uncertainty in the classification of echolocation calls is a source of much controversy (Barclay, 1999; O'Farrell et al., 1999a,b; Fenton, 2000; Corben and Fellers, 2001). Debates have been waged on the merits of various bat detection devices as well as the merits of qualitative versus quantitative paradigms of call classification. The qualitative approach to echolocation call classification was used as early as forty years ago (Fenton, 1970; Fenton and Bell, 1981). It involves the use of expert knowledge in matching the spectrographic information of a recorded bat call to those in a library of calls from known species. O'Farrell et al. (1999a) provide an excellent overview of the assumptions and information required in this approach. Two common criticisms of qualitative identification are lack of objectivity (e.g., two observers may not come to the same identification) and the significant time investment required for analysis (Barclay, 1999; Jennings et al., 2008; Parsons and Szewczak, 2009; Skowronski and Fenton, 2009). Quantitative methods, on the other hand, are externally unbiased (i.e. bias results from poor data/incorrect techniques rather than observer differences) and fast. However, many of the computational techniques used to classify calls quantitatively require large amounts of carefully-chosen data and computing power to accurately employ, and they require a thorough understanding of the statistical assumptions at play.

Most studies that have tested quantitative classifiers of bat calls have done so using time-expanded echolocation recordings (Vaughan et al., 1997; Parsons and Jones, 2000; Kazial et al., 2001; Russo and Jones, 2002; Biscardi et al., 2004; Fukui et al., 2004; Obrist et al., 2004; Preatoni et al., 2005; Skowronski and Harris, 2006; Jennings et al., 2008; Redgwell et al., 2009). A much smaller number of studies have assessed the accuracy of quantitative techniques using zero-crossings data (Krusic and Neefus, 1996; Lance et al., 1996; Herr et al., 1997; Stocker, 1998; Britzke, 2003; Corcoran, 2007). It is much more common for zero-crossings data to be analyzed qualitatively, and researchers most often use *a priori* groupings for species whose calls cannot be reliably separated (e.g., *Myotis* spp.); (Betts, 1998; Humes et al., 1999; Kalcounis et al., 1999; Erickson and West, 2003; Menzel et al., 2005). These analyses have been aided by the use of custom filters which single out calls that fall within predefined parameter values and ignore extraneous noise (Kalcounis et al., 1999; Britzke and Murray, 2000; Britzke, 2003; Menzel et al., 2005).

Quantitative classification of bat calls has generally been achieved using either multivariate statistical techniques (e.g., MANOVA and discriminant analysis) or supervised learning algorithms (e.g., artificial neural networks, decision trees, and support vector machines). For time-expanded recordings, overall classification accuracies of >95% have been achieved using both hidden Markov models and ensembles of artificial neural networks (Skowronski and Harris, 2006; Redgwell et al., 2009). In comparison, classification accuracies when using these techniques for zero-crossings data have been lower (roughly 70% for species and 90% for genus), because these calls lack harmonic and amplitude information, which are potentially important features in call discrimination.

The objectives of this study are to (1) compare the accuracies of four quantitative classification techniques (discriminant function analysis, artificial neural networks, support vector machines, and random forests) faced with a number of bat call discrimination problems from two regional datasets and (2) determine the relative importance of discriminating features of calls among classification techniques.

2. Materials and methods

2.1. Data collection

We constructed two call libraries, one made up of calls from common southwestern United States vespertilionid bats (SW bats), and one of calls from vespertilionids we expected to find in Florida longleaf pine sandhills habitat (FL bats); (Table 1). The decision to use these species assemblages is based on range maps from each region (Best et al., 2000; Marks and Marks, 2006). These call libraries contained zero-crossings files of bats recorded either during roost emergence, hand release, or visual identification by an expert observer. Call files were either collected by the authors or generously donated by colleagues (see details in Acknowledgements section). These techniques for obtaining calls can be problematic due to factors such as observer misidentification and the effects of capture on bat vocalization behavior. Nonetheless, they are the standard methods for obtaining voucher calls from bats and have been used in all previous studies of bat call classification (Parsons and Szewczak, 2009). All files were recorded using the Anabat II+ZCAIM hardware (Titley Electronics, NSW, Australia). For this study, we defined a useable file as one which included >2 individual calls of high quality. We used a three-step process to identify high-quality calls. First, we used a filter developed by Britzke and Murray (2000) within the AnalookW software (ver. 3.7) to eliminate fragmentary calls, feeding pulses, and extraneous noise, thus retaining only identifiable search-phase calls (Fenton and Bell, 1981). Second, we examined each file and kept only those containing more than two search-phase calls of high quality (Britzke, 2003). Here, the quality of a call is defined by the AnalookW program by an index, *Qual* (available using the 'measures' feature), which is an averaged measure of the smoothness of the call

$$Qual = \frac{\left(\sum_{i=1}^{n} Q_i\right) * 100}{n-1} \qquad (1)$$

where

$$Q_i = \left| \frac{\left(F_n - F_{avg}\right)}{F_{avg}} \right| \qquad (2)$$

and

$$F_{avg} = \frac{\left(F_{n-1} + F_{n+1}\right)}{2} \qquad (3)$$

Table 1
Number of calls of species included in each bat call library.

Call library	Species	Acronym	N
SW bats	*Antrozous pallidus*	ANPA	34
	Eptesicus fuscus	EPFU	24
	Lasiurus borealis	LABO	21
	Lasiurus cinereus	LACI	45
	Myotis auriculus	MYAR	62
	Myotis californicus	MYCA	41
	Myotis thysanodes	MYTH	30
	Myotis volans	MYVO	58
	Perimyotis subflavus	PESU	28
	Pipistrellus hesperidus	PIHE	38
	Tadarida brasiliensis	TABR	30
Florida bats	*Eptesicus fuscus*	EPFU	24
	Lasiurus borealis/seminolus	LABOSE	32
	Lasiurus intermedius	LAIN	20
	Myotis austroriparius	MYAU	22
	Nycticeius humeralis	NYHU	20
	Perimyotis subflavus	PESU	42
	Tadarida brasiliensis	TABR	23

and F_n is the frequency of the call at point n. Only calls with a value of $Qual < 0.3$ were considered for analysis. Finally, to remove the effects of pseudoreplication, we randomly selected one high-quality search-phase call per file. If two files were recorded within 5 min of each other, we assumed the calls were from the same individual animal. This is a very conservative approach to call selection, adopted to minimize the chances of violating the assumption that all calls included in analyses were independent samples from different individuals (Hurlbert, 1984; Mundry and Sommer, 2007).

Once a call was randomly selected, we used the software's built-in "measures" feature to automatically extract 11 features describing the shape of the call: maximum frequency (Fmax), mean frequency (Fmean), minimum frequency (Fmin), frequency at the knee (Fk), characteristic frequency (Fc), call duration (Dur), characteristic slope (Sc), initial slope (S1), time until end of characteristic slope (Tc), time until knee (Tk), quality of the knee (Qk), and the time between two successive calls (TBC) (Fig. 1). We then grouped calls by genus and by species, and omitted from further consideration those groups with a membership of <20 individual animals to minimize sampling error. A general rule for classification algorithms is to have a greater number of individuals per group than the number of features, though arguments have been made to include a minimal number of training individuals to reduce variance between individuals of the same species (Hair et al., 2006). This type of data reduction may not be appropriate to account for the high levels of inter- and intra-specific variation typically found in large-scale echolocation monitoring studies. However, the addition of more calls increases the risk of misclassifying species with call features that overlap due to the allowed increase in the variance of the feature set. Studies similar to this one generally use as many available calls that fall within their call selection criteria as possible. Currently, there is no evidence to support a reduced training dataset in cases of bat call classification.

2.2. Data preparation

We used the R statistical computing language (R Development Core Team, 2009) in conjunction with the Classification and Regression Training package ("caret"); (Kuhn, 2009). We began by searching for near-zero-variance predictors, which are variables that are nearly identical among the classes among which we wanted to discriminate. We defined a near-zero-variance predictor as a feature with a frequency ratio >2 and a percent of unique predictor value <0.2 where

$$Frequency\ ratio = \frac{n_r}{n_{r+1}} \tag{4}$$

and

$$Percent\ unique\ predictors = \frac{n_{unique}}{n_{total}} \tag{5}$$

where n_r is the frequency of observations of the predictor value of frequency rank r, n_{unique} is the frequency of unique values of a predictor variable, and n_{total} is the total number of observations for each group. We selected a conservative 0.2 threshold to ensure that at least 80% of feature values differed for a single species. Near-zero variance predictors can become zero-variance predictors when split into training subsamples and can cause the classifier to fail. This failure may manifest in a variety of ways, but most commonly results in convergence errors when using likelihood-based estimators (e.g., in discriminant functions). Therefore, any near-zero variance predictors were removed from the datasets.

Because multicollinearity among features can lead to model overfitting and can impact the generalizability of discriminant functions and support vector machines, we constructed a correlation matrix among features and removed the columns that contributed the greatest mean absolute correlation among predictors above the threshold value of 0.85 (Bao and Cui, 2005; Zhu and Zou, 2007; Nicodemus and Malley, 2009). Finally, we centered and scaled the variables. For the artificial neural networks, we scaled predictor values between zero and one by dividing each observation by the maximum observed value. For all other models we centered the data by subtracting the column mean from each observation and scaled them by dividing each centered value by the column's root-mean-square (Mukherjee, 2006; Kuhn, 2008).

2.3. Data analysis

We trained five supervised learning algorithms (discriminant function analysis, artificial neural networks, artificial neural networks trained on principal components, support vector machines, and random forests) for each of five classification tasks:

1. All species of bats represented in the Southwestern bat library (SW bats, 11 classes)
2. All species of bats in the Southwest with *Myotis* spp. grouped together (SW bats, 8 classes)
3. All genera of bats in the Southwest (SW bats, 6 classes)
4. All *Myotis* spp. in the Southwest (SW bats, 4 classes)
5. All species of bats in the Florida bat library with *Lasiurus borealis* and *L. seminolus* grouped together (FL bats, 6 classes)

These five specific classification tasks were selected because they provide comprehensive coverage of the types of problems for which echolocation call analyses are frequently undertaken (e.g., tasks 1 and
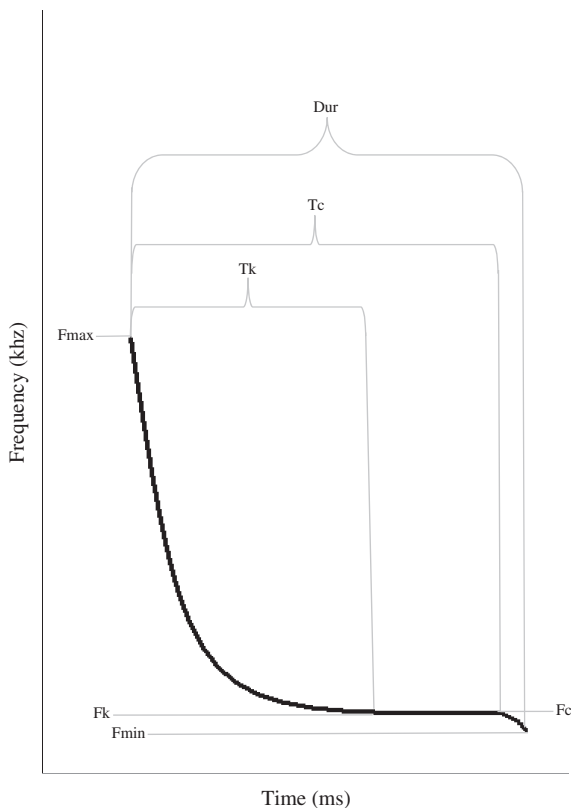
Fig. 1. Features of a bat's echolocation call used in this study. The thick black line represents the idealized call. Dur = total duration of call; Tc = duration from start of call to characteristic frequency; Tk = duration from start of call to start of knee; Fmax = highest frequency of call; Fmin = lowest frequency of call; Fk = frequency at start of knee; Fc = frequency at the end of the flattest portion of call (characteristic frequency) (O'Farrell et al., 2000; Gannon et al., 2004).

5 represent species-level classification of a particular regional assemblage, while tasks 2, 3, and 4 represent various alternative approaches employed in situations where species-specific discrimination is not possible). All classification tasks were performed using the R statistical computing language (R Development Core Team, 2009) in conjunction with the Classification and Regression Training package ("caret"); (Kuhn, 2009).

2.3.1. Discriminant function analysis

Discriminant function analysis (DFA) is a statistical discrimination method by which a number of feature classes c_i are modeled with respect to numerical features x_i and the resulting functions are used to categorize unknown cases (Hair et al., 2006). Two steps are involved in DFA. First, an F-test (Wilk's Lambda-test) is run to determine whether or not classes are separable by their constituent features. Next, discriminant functions are formulated from the best combinations of discriminating features. For two-class discrimination problems, this is accomplished with a single function. For $c_i > 2$, this can involve many functions, which are then ranked in order of relative discriminating importance by each function's eigenvalue. These discriminant functions can either take linear or quadratic forms, depending on the fulfillment of certain assumptions. Foremost is the assumption of homogeneity of variance–covariance matrices between classes and is tested with Box's M-test. If this assumption is not met, the quadratic classifier function can be used in place of the linear function, or the data can be transformed and the test rerun. Another assumption of DFA is multivariate normality among features. However, given appropriate sample sizes (>20 observations per class) and a relatively low number of classes (<8), this test is generally robust against violations of the assumption of normality. Finally, since the discriminant function technique relies on identifying and ranking the importance of features in discriminating among classes, highly correlated features can decrease classification accuracy. It is therefore a common practice to remove highly correlated variables before calculating Wilk's Lambda (Hair et al., 2006).

2.3.2. Artificial neural networks

The artificial neural network (ANN) is a non-linear, adaptive, machine learning tool built on connectionist principles (White, 1992; Veelenturf, 1995; Ripley, 1996; Lek et al., 2000; Samarasinghe, 2007). In this study, we used a three-layered back-propagation feed-forward neural network. The back-propagation algorithm is the method by which the network "trains" itself by modifying a set of initially-randomized weight functions based on the error of the network's output. Eventually, these weights are optimized resulting in a trained, highly generalizeable network. To build the network, we performed a grid search for optimal size ($1 < size < 29$, only prime numbers to disallow equal output votes between groups) and decay ($0 < decay < 1$) parameters spanning a total of 225 networks, each of which was iterated between 500 times for genus-level analyses and 2000 times for species-level analyses. The grid search automatically identified the combination of parameter values that achieved the highest classification accuracy. Next, we performed a principal components analysis on the scaled ANN data to remove the effects of correlated features and to lower the dimensionality of the feature set. We used the principal components to train the network. We performed this step believing that it would lead to more accurate predictions using a smaller feature set and less computing power.

2.3.3. Support vector machines

Support vector machines (SVM) are a supervised learning tool that constructs a linear separating hyperplane between two classes by maximizing the distance between them (Cristianini and Shawe-Taylor, 2000; Abe, 2005). The placement of the hyperplane is based on the location of support vectors, which are the marginal samples from the $\binom{|C|}{2}$ neighboring classes nearest each other. Because linear separation of bat call classes was not possible, we made use of the radial basis function (RBF) kernel to transform the feature space to enable the fitting of a maximum-margin hyperplane. This implementation of support vector machines uses a pairwise classification method in which a new feature vector is subjected to binary classifiers and is assigned to the class which receives the most votes (Anguita et al., 2004). We performed a grid search for two parameters: cost (c) and sigma (σ). The cost parameter was varied between 0.01 and 1000 in steps of one order of magnitude. The optimal value of this parameter represents a tradeoff between high classification accuracy and generalizability of the model. The parameter sigma is required by the RBF kernel and its range of testable values was selected following the methods described in Caputo et al. (2002).

2.3.4. Random forests

The random forest (RF) is a tree-based algorithm (Breiman, 2001) that to our knowledge has never been applied to bat call classification tasks. The algorithm builds a specified number of classification trees without pruning, the nodes of which are split on a random drawing of m features from the entire feature set M. A bootstrapped sample from the training set is used to build each tree, and the samples left out of the training set (dubbed "out-of-bag," or "OOB" samples) are used to calculate a running error estimate of the routine. Random forests are advantageous due to their ability to generate a metric which ranks predictors based on their relative contribution to the model's predictive accuracy (Bao and Cui, 2005). This measurement, called 'mean decrease accuracy', uses the OOB data to quantify the importance of each variable by first running the OOB data down all trees and counting the number of true positives. Next, one of the features of that OOB sample is randomly permuted and the sample is run again. The difference in the true positive rates between the complete OOB sample and the OOB sample with the randomly permuted feature averaged over all trees in the forest equals that feature's mean decrease accuracy value. This value is normalized by dividing the feature's raw score by its standard error, producing a z-score amenable to further statistical comparisons. Because the RF algorithm has no distributional assumptions, there is no need for any transformation or reduction of feature sets. It is highly robust to noisy data and large numbers of correlated predictors. We ran a grid search to find the optimal number of variables to be randomly sampled at each branch split (mtry). This value was allowed to range from 2 to 8. Each forest was grown to 1000 trees and the final mtry value was selected for the highest predictive accuracy.

2.3.5. Model training

All classifiers were trained on 80% of the data from each library. We used 10-fold cross-validation to assess each model's accuracy. This method partitions the data into 10 subsets while maintaining the proportionality of group representations. The model is then trained on 9 of the subsets and uses the remaining subset to assess its accuracy. This process is then repeated until all subsets have been used as training and test sets. Since this method can overestimate model accuracies, we then used the remaining 20% of each original library as another test of the accuracies of the final models.

Splitting the dataset in this manner also enabled computation of one-versus-all measures of the sensitivity, specificity, and positive and negative predictive powers (PPP and NPP, respectively) for each classifier using the "validation set" of data left out of the model (Cutler et al., 2007). These measures are all optimally 1, indicating perfect accuracy. These values are important in selecting the most appropriate classification technique for a given task. For example, 'positive predictive power' is a measure of the probability that an unknown call is correctly identified (Jennings et al., 2008). 'Sensitivity' and 'specificity' are the rates of correctly-assigned positive identifications and negative identifications, respectively. For RF models, we also calculated each call feature's importance measure, the 'mean decrease

accuracy', using the OOB data as described above. This was done for each classification task (with the exception of task 2) to assess the relative contributions of predictors at three levels of classifications—species, genus, and *Myotis* spp. We scaled RF importance values to lie between 0 and 100 to facilitate comparisons among the importance measures of other classifiers (Kuhn, 2008).

3. Results

The southwestern United States bat call library (SW bats) contained 11 species, seven genera, 718 files containing call sequences from individual bats, and was made up of approximately 26,000 single calls. A total of 411 calls satisfied the selection criteria outlined above and were used in the analyses. We constructed the Florida bat call library (FL bats) by selecting 155 calls from a pool of 8000 total calls from 8 species that occur in the longleaf pine sandhills habitat (Armitage, 2010). In all cases only one call was used per individual animal, and > 20 individual animals were represented per species or species group (the bats *Lasiurus seminolus* and *L. borealis* were classified as a single group, due to their near-identical frequency-divided calls).

A number of features were highly correlated with others and were subsequently removed from SVM and DFA analyses (Fmin, the minimum frequency; Fk, the frequency at the knee; and Tc, the duration from the start of the call to the end of the characteristic slope). Additionally, two features (Sc, the slope of the flattest part of the call; and Fc, the frequency at the end of the flattest portion of the call) were found to contribute to matrix rank-deficiency in a number of discriminant function models and were dropped from analyses.

Box's M-test showed significant inequalities among covariance matrices for all datasets ($p < 0.0001$). Moreover, a Shapiro–Wilks test found all features to significantly differ from normality ($p < 0.001$). Therefore, we used the quadratic form of the discriminant function (Vaughan et al., 1997; Russo and Jones, 2002). For all datasets, Wilk's Lambda test showed significant differences in the mean discriminant function scores ($p < 0.0001$), indicating that the functions were discriminating between groups.

Overall, each of the supervised learning models with the exception of DFA performed reasonably well according to the results of a 10-fold cross-validation (Table 2). Accuracy rates for all models except DFA were between 85% and $86\% \pm 2\%$ averaged over all classification tests. DFA showed considerably lower average accuracy rates of $76\% \pm 2.4\%$ across all tests. Artificial neural networks (both with and without prior principal components analyses) performed marginally better than support vector machines and random forests. However, both support vector machines and random forests outperformed the neural networks on the final classification task (Florida bats). Classification accuracies were greatest for tests 2 (*Myotis* spp. grouped) and 3 (genera) and lowest for tests 1 (all species, SW bats) and 5 (all species, FL bats). Accuracy rates in discriminating among four species of southwestern *Myotis* spp. (test 4) were lowest for DFA and highest for ANN.

We achieved similar results using the independent validation set (Table 3). Random forests and artificial neural networks both performed well in prediction of the withheld data, with classification accuracies all over 80% for SW bats (test 1), and upwards of 96% for tests 2 and 3. Random forests and support vector machines showed the highest average sensitivities, specificities, positive predictive values, and negative predictive values for each classification task. The FL bats classification task had the lowest overall accuracy metrics, with a maximum 85% classification accuracy using random forests.

In classification test 1 of southwestern bats, the models' inabilities to discriminate calls among *Myotis* spp. were the main reason for the lower accuracies. We obtained the best results by grouping together all *Myotis* species (test 2), which increased the predictive accuracies of the models a minimum of 8–10% over task 1. *Myotis auriculus* was the most problematic species to classify in tests 1 and 4. In one-versus-all

Table 2
Overall classification accuracies for datasets used to train models. Estimates ($\pm$ standard error) are based on 10-fold cross-validation of training data (80% of call libraries).

Test	Model	Training accuracy
SW bats	RF	0.829 ± 0.02
All species	SVM	0.816 ± 0.02
	NNET	**0.847 ± 0.01**
	PCA-NNET	0.839 ± 0.02
	DFA	0.639 ± 0.01
SW bats	RF	0.905 ± 0.01
All species/	SVM	0.883 ± 0.01
Myotis grouped	NNET	0.915 ± 0.01
	PCA-NNET	**0.927 ± 0.02**
	DFA	0.737 ± 0.02
SW bats	RF	0.917 ± 0.02
Genus-only	SVM	0.913 ± 0.02
	NNET	**0.925 ± 0.01**
	PCA-NNET	0.923 ± 0.01
	DFA	0.884 ± 0.02
SW bats	RF	0.843 ± 0.03
Myotis-only	SVM	0.822 ± 0.02
	NNET	**0.911 ± 0.02**
	PCA-NNET	0.859 ± 0.01
	DFA	0.751 ± 0.03
N. FL bats	RF	0.783 ± 0.03
All species	**SVM**	**0.841 ± 0.02**
	NNET	0.682 ± 0.04
	PCA-NNET	0.780 ± 0.04
	DFA	0.784 ± 0.04

Bolded values indicate the highest values ($\pm$ SE) for each classification task.

comparisons, *M. auriculus* had lower sensitivities than all other species by a minimum margin of 0.15. In test 5 of Florida bats, the most accurate model was the random forest. This model, along with SVM and NNET models all confused the calls of *L. borealis/seminolus* and *Nycticeius humeralis.* Similarly, the calls of *Tadarida brasiliensis* and *Lasiurus intermedius* were often confused.

Values of variable importance measures were similar across classification tests at different levels of taxonomic organization (species and genus) and among the two geographic regions (north Florida and the Southwestern US). Results suggest that a number of frequency characteristics (Fmax, Fmin, Fk, Fc), the characteristic slope (Sc), and the total duration of a call (Dur) were more important in call discrimination than time characteristics (Tk, Tc, TBC), the initial slope of the call (S1), or the quality of the call at the knee (Qk); (results for RF: Fig. 2). However, no features fell considerably below 20% mean decrease accuracy, and omission of the lowest-performing variables may cause unacceptable loss of predictive accuracy.

4. Discussion

4.1. Comparison of classification techniques

In this paper, we tested the utility of a number of machine learning algorithms (random forests, support vector machines, and artificial neural networks) and a classical statistical technique (discriminant function analysis) for classifying the echolocation calls of bats. All machine learning methods produced better, more generalizable results than the discriminant function analysis (Tables 2 and 3). Furthermore, most of these machine learning methods were unburdened by the distributional, covariance, and multicollinearity assumptions on which techniques such as DFA and MANOVA depend. Since the machine learners did not strongly differ in their classification accuracies, it is the prerogative of the investigator on which machine learning method to employ.

While artificial neural networks showed the highest classification accuracy of the training sets (Table 2), they were the most computationally-intensive techniques and took the longest amount of time to train. This high classification accuracy likely results from the

Table 3
Overall classification accuracies, average sensitivities, specificities, positive predictive powers (PPP), and negative predictive powers (NPP) for each model and classification task. Estimates are based on the 20% of calls left out of the training sets and include standard errors.

Test	Model	Accuracy	Sensitivity	Specificity	PPP	NPP
SW bats	RF	0.78	0.827 ± 0.063	**0.980 ± 0.008**	**0.808 ± 0.061**	**0.977 ± 0.011**
All species	SVM	0.8	**0.841 ± 0.053**	0.979 ± 0.006	0.821 ± 0.046	0.979 ± 0.010
	NNET	**0.84**	0.594 ± 0.068	0.960 ± 0.008	0.614 ± 0.060	0.961 ± 0.007
	PCA-NNET	0.77	0.548 ± 0.068	0.950 ± 0.007	0.531 ± 0.059	0.955 ± 0.008
	DFA	0.61	0.468 ± 0.040	0.943 ± 0.006	0.464 ± 0.062	0.943 ± 0.011
SW bats	RF	**0.96**	**0.935 ± 0.035**	**0.993 ± 0.004**	**0.958 ± 0.030**	**0.994 ± 0.002**
All species/	SVM	0.92	0.889 ± 0.033	0.988 ± 0.004	0.903 ± 0.044	0.989 ± 0.002
Myotis grouped	NNET	0.91	0.744 ± 0.070	0.967 ± 0.010	0.812 ± 0.045	0.968 ± 0.008
	PCA-NNET	0.88	0.645 ± 0.056	0.959 ± 0.008	0.666 ± 0.057	0.959 ± 0.008
	DFA	0.74	0.536 ± 0.058	0.933 ± 0.019	0.541 ± 0.742	0.933 ± 0.018
SW bats	RF	**0.96**	**0.940 ± 0.021**	**0.991 ± 0.002**	**0.933 ± 0.037**	**0.992 ± 0.002**
Genus-only	SVM	0.92	0.917 ± 0.037	0.985 ± 0.006	0.878 ± 0.057	0.984 ± 0.006
	NNET	0.91	0.770 ± 0.064	0.961 ± 0.010	0.812 ± 0.038	0.962 ± 0.010
	PCA-NNET	0.92	0.762 ± 0.046	0.964 ± 0.009	0.792 ± 0.055	0.962 ± 0.008
	DFA	0.87	0.647 ± 0.072	0.944 ± 0.012	0.674 ± 0.045	0.955 ± 0.010
SW bats	RF	0.79	0.834 ± 0.045	0.924 ± 0.023	0.837 ± 0.045	0.924 ± 0.023
Myotis-only	SVM	0.83	**0.875 ± 0.047**	**0.937 ± 0.018**	**0.859 ± 0.034**	**0.937 ± 0.021**
	NNET	**0.85**	0.827 ± 0.046	0.927 ± 0.022	0.822 ± 0.045	0.926 ± 0.023
	PCA-NNET	0.79	0.773 ± 0.042	0.905 ± 0.030	0.773 ± 0.055	0.904 ± 0.030
	DFA	0.82	0.551 ± 0.082	0.856 ± 0.024	0.539 ± 0.064	0.866 ± 0.032
FL bats	RF	**0.85**	**0.827 ± 0.078**	**0.969 ± 0.012**	**0.845 ± 0.062**	**0.971 ± 0.010**
All species	SVM	0.7	0.711 ± 0.074	0.939 ± 0.014	0.722 ± 0.053	0.943 ± 0.013
	NNET	0.67	0.427 ± 0.098	0.896 ± 0.047	0.565 ± 0.126	0.904 ± 0.017
	PCA-NNET	0.75	0.451 ± 0.107	0.904 ± 0.020	0.521 ± 0.107	0.908 ± 0.016
	DFA	0.77	0.444 ± 0.084	0.901 ± 0.020	0.816 ± 0.059	0.904 ± 0.010

Bolded values indicate the highest values (± SE) of each metric for each classification task.

computationally-intensive back-propagation process, during which feature weights are modified according to an iterative algorithm. None of the other classification models use such an algorithm, and instead rely on either binary decision rules (RF's splitting nodes) or mathematical separation of feature sets (SVM's kernel trick and DFA's linear/quadratic functions). However, this property of neural networks can hinder the interpretation of the relationships between the weighted nodes in the final model and their effects on input features. Transformation of the feature set into a series of principal components effectively eliminates problems of multicollinearity and would decrease the computational complexity of very large feature sets (Diamantaras and Kung, 1996). In this study, however, the feature set

was small enough that the difference in the time to train an ANN on principal components or the entire feature set was negligible. The classification accuracies of this study compared favorably to others that used zero-crossings data (e.g., Lance et al. 1996; Herr et al., 1997; Britzke, 2003; Corcoran, 2007). It is clear, though, that the information retained in time-expanded data are important in species discrimination, which is why this study had lower average accuracy rates than those using time-expanded calls (e.g., Biscardi et al., 2004; Fukui et al., 2004; Redgwell et al., 2009).

Support vector machines showed high training accuracies even after the elimination of highly correlated features. The effects of multicollinearity among predictors are unclear, especially after
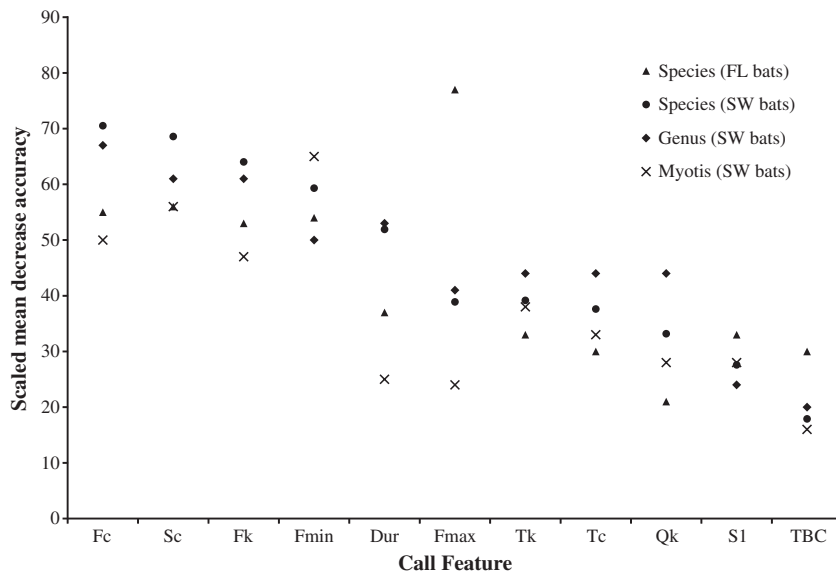
Fig. 2. Scaled mean decrease accuracy of predictor features for random forests applied to four classification tasks. Higher values indicate greater contribution to the predictive accuracy of final models.

nonlinear transformations of the data. With relaxed assumptions, SVM classification may have benefitted from retaining correlated features, as it has in other studies (Drake et al., 2006). In this case, *a posteriori* training of SVMs on the entire feature class led to models incapable of correctly predicting classes better than class assignments based on chance alone. Due to their ease of implementation and their ability to cope with large numbers of predictor variables, support vector machines have seen a recent increase in use in ecology, especially within the fields of landscape ecology and bioacoustics (Meyer et al., 2003; Drake et al., 2006; Fagerlund, in press; Acevedo et al., 2009; Huang et al., 2009). As far as we are aware, this is the second test of support vector machines in the classification of bat calls. The earlier study arrived at results very similar to ours (genus-level: 93% vs. 91%; species-level: 87% vs. 82%; *Myotis*-only: 89% vs. 82%) despite the entirely different species pool (UK bats) and use of time-expanded call recordings rather than frequency-division (Redgwell et al., 2009).

Random forests produced the best outcome in terms of classification accuracy, sensitivity, specificity, positive predictive power and negative predictive power for all test cases. They are also the second-fastest (behind DFA) to train and implement, since they only have one tuning parameter (*mtry*) that occupies a limited range of positive integers. A high-end laptop computer took approximately 2 min to train a random forest model, 10 min for support vector machines, and from 30 min to several hours for artificial neural networks. Larger grid searches for ANN's and high-degree kernel functions for SVM's dramatically increase computation time. Another advantage of random forests is their ability to easily model correlated features and account for those correlations when calculating variable importance measures (Archer and Kimes, 2007). The strength of this approach lies in the "out-of-bag" samples left behind after the random bootstrap procedure builds each tree. These OOB samples are used to evaluate performance and calculate the "loss of accuracy" of a tree when the values of a feature are completely randomized.

4.2. Relative importance of discriminating features

RF predictor importance measures differed between classification tasks, but frequency information (Fmax, Fmin, Fk, and Fc) consistently performed better than most other variables associated with timing. The duration (Dur) and characteristic slope (Sc) were also important features in accurate classification. Other studies ranking variable importance have arrived at similar conclusions (Vaughan et al., 1997; Parsons and Jones, 2000; Russo and Jones, 2002; Redgwell et al., 2009). We found that the characteristic frequency of the call (Fc) was on average the most important variable for bat call classification across all classification tasks. It occurs at the end of the flattest part of the echolocation call and shows very low intra-specific variation. Fc is somewhat analogous to Fp (the peak energy frequency in time-expanded calls), which is recognized as being very useful in discriminating among species (Corben, 2004). Discrimination among *Myotis* bats has traditionally been complicated by low variation among features. *Myotis* species tend to produce short, wideband, high frequency calls, which explains the poor classification performance of total call duration (Dur), time between calls (TBC), and maximum frequency (Fmax) for task 4, which attempted to discriminate among the *Myotis* species in the SW bats library. Fmax was far more helpful in discriminating among genera in task 3, which attempted to classify the genera of bats in the SW bats library. Our results suggest that the minimum (Fmin), knee (Fk), and characteristic frequencies (Fc) are the most important metric for discriminating among *Myotis* species. Parsons and Jones (2000) similarly identified the start frequency (F-start), end frequency (F-end), and the peak energy frequency (Fp) as the most important features for distinguishing *Myotis* spp., *Nyctalus* spp., and *Pipistrellus* spp. Studies by Vaughan et al. (1997) and by Russo and Jones (2002) achieved similar results when discriminating among a number of FM bat species (including

Myotis spp., *Plecotus* spp., and *Barbastella* spp.). Among *Myotis* spp. of northern Italy, F-end, Fmin, and call duration were identified as the most important discriminating features (Preatoni et al., 2005).

Our results indicate the existence of geographic differences in the relative contribution of some call features to the predictive accuracy of RF models. Eight of the 11 features measured provided greater discriminatory contributions among species in the SW bats library than among species in the FL bats library (Fig. 2). However, maximum frequency (Fmax), initial slope (S1), and time between calls (TBC) contributed greater discriminatory abilities among species in the FL bats library than in the SW bats library. This is probably because the Florida bats library contained fewer species with extremely similar call structures (i.e., fewer *Myotis* species).

4.3. Conclusions

The acoustic monitoring of bats compliments capture techniques and should always be considered when assessing species diversity and habitat use by bats (Weller and Lee, 2007). Currently, the major challenge in echolocation monitoring studies is the accurate classification of unknown calls to a particular genus or species. In many cases, this task is aided by the use of machine learning algorithms. We suggest a two-stage solution to bat call classification problems whereby unknown calls are analyzed by a hierarchical series of machine learners (ANN, SVM, and RF) trained to first separate groups by genus, and then by species (or species groups sharing similar call characteristics). We recommend random forests because they were the most accurate with regards to sensitivity, specificity, positive predictive power and negative predictive power; were quick to train; and provide useful measures of variable importance. In passive and laboratory bioacoustics studies, the computational complexity, and thus the memory and time requirements of these algorithms are trivial, and researchers can now easily dedicate multiple processors working in parallel to compute multilayer neural networks and high-dimensional kernel functions. However, with the advent of server-side cloud computing and the increasing roles of active acoustic monitoring in ecological surveys, it stands to reason that there may someday be need for near-immediate acoustic classification under field conditions, which may favor the use of computationally-simpler techniques.

Depending on the objectives of the study (e.g., to evaluate species or genus richness vs. to detect the presence of an endangered species), learners should be trained to a minimum classification accuracy and selected for maximal sensitivity, specificity, PPP (positive predictive powers) and NPP (negative predictive powers) values from independent training data. It has been suggested that PPP is the most important metric to maximize in call ID models, since it measures the probability that an unknown call will be correctly identified (Jennings et al., 2008). Many models can incorporate cost arguments that select models which maximize one or all of these measurements for a particular group (Fielding, 1999). Ideally, the results from the model will be independently verified by a trained observer, though oftentimes this is a very time-consuming task. In general, when higher classification accuracy is required, time-expanded recordings should be used in conjunction with either a hierarchical set of classifiers such as random forests, hidden Markov models, or artificial neural networks. Nevertheless, analysis of zero-crossings data using machine learning tools results in acceptable classification accuracies even with the loss of harmonic and energy information.

According to Thomas Kuhn, the route to scientific discovery often demands the development of novel technology to aid in testing theoretical prediction (Kuhn, 1970). As such, the trend towards small, mobile computers running externally-hosted services via the internet represents a novel (and currently untapped) resource that ecologists could be employing to measure, analyze, and report data. Researchers in the United Kingdom have successfully launched an initiative dubbed "BeeID," which makes use of social networking and photo-

sharing websites to collaboratively identify and geo-tag user-submitted photographs of bees. These data will then be used by entomologists in mapping bee diversity and distributions (Kirkhope et al., 2010). So-called "smartphones" are equipped with high-speed processors, GPS, and microphones. A number of programs for these phones already allow users to record fragments of music which are uploaded to a server and identified within 30 s (e.g., http://www.shazam.com). Such systems have already been outlined and implemented with success (Aanensen et al., 2009). This technology could be expanded to animal vocalizations, whereby a sound is recorded, its relevant features extracted; and is then fed through an online series of classification tools. The results could be returned to the user along with pages displaying information on the species. The phone's built-in GPS could also upload the species and location information to a database such as eBird (http://www.ebird.org), which already makes use of smartphone technology in reporting the results of bird counts. The benefit to an automated approach would be that anybody possessing the technology could contribute to the monitoring effort. Obviously, a system such as this is still far off, and the difficulties involved in collecting and digitizing voucher calls, automatically extracting their features, and programming a quick online classification tool are huge.

Acknowledgements

References

Aanensen, D.M., Huntley, D.M., Feil, E.J., al-Own, F., Spratt, B.G., 2009. EpiCollect: linking smartphones to web applications for epidemiology, ecology and community data collection. PLoS One 4 (9), e6968. doi:10.1371/journal.pone.0006968.

Abe, S., 2005. Support Vector Machines for Pattern Classification. Springer-Verlag, London, UK.

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. Ecological Informatics 4, 206–214.

Andersen, B.B., Miller, L.A., 1977. Portable ultrasonic-detection system for recording bat cries in the field. Journal of Mammalogy 58, 226–229.

Anguita, D., Ridella, S., Sterpi, D., 2004. A new method for multiclass support vector machines. Proceedings of International Joint Conference on Neural Networks (IJCNN 2004). Budapest Hungary 1, 407–412.

Archer, K.J., Kimes, R.V., 2007. Empirical characterization of random forest variable importance measures. Computational Statistics and Data Analysis 52, 2249–2260.

Armitage, D.W., 2010. The effects of prescribed fire on bat activity in the longleaf pine sandhills ecosystem. M.S. thesis, University of Florida, Gainesville, FL, USA.

Bao, L., Cui, Y., 2005. Prediction of the phenotypic effects of non-synonymous single nucleotide polymorphisms using structural and evolutionary information. Bioinformatics 21, 2185–2190.

Barclay, R.M.R., 1999. Bats are not birds—a cautionary note on using echolocation calls to identify bats: a comment. Journal of Mammalogy 80, 290–296.

Best, T., Harvey, M., Altenbach, J., 2000. Batcall: Accoustic library and species accounts (on-line). Accessed July 28, 2010 at: http://www.msb.unm.edu/mammals/batcall/html/speciesaccounts.html.

Betts, B.J., 1998. Effects of interindividual variation in echolocation calls on identification of big brown and silver-haired bats. Journal of Wildlife Management 62, 1003–1010.

Biscardi, S., Orprecio, J., Fenton, M.B., Tsoar, A., Ratcliffe, J.M., 2004. Data, sample sizes, and statistics affect the recognition of species of bats by their echolocation calls. Acta Chiropterologica 6, 347–363.

Block, B.A., Dewar, H., Farwell, C., Prince, E.D., 1998. A new satellite technology for tracking the movements of Atlantic bluefin tuna. Proceedings of the National Academy of Sciences 95, 9384–9389.

Breiman, L., 2001. Random forests. Machine Learning 45, 5–32.

Britzke, E.R., 2003. Use of ultrasonic detectors for acoustic identification and study of bat ecology in the eastern United States. Ph.D. dissertation, Tennessee Technological University, Cookeville, TN, USA.

Britzke, E.R., Murray, K.L., 2000. A quantitative method for selection of identifiable search-phase calls using the Anabat system. Bat Research News 41, 33–36.

Caputo, B., Sim, K., Fursejo, F., Smola, A., 2002. Appearance-based object recognition using SVMs: which kernel should I use? Proceedings of NIPS workshop on Statistical Methods for Computational Experiments in Visual Processing and Computer Vision, Whistler.

Corben, C., 2004. Zero-crossings analysis for bat identification: an overview. In: Brigham, R.M., Kalko, E.K.V., Jones, G., Parsons, S., Limpens, H.J.G.A. (Eds.), Bat Echolocation Research: Tools, Techniques, Analysis. Bat Conservation International, Austin, TX, USA, pp. 95–106.

Corben, C., Fellers, G.M., 2001. Choosing the 'correct' bat detector—a reply. Acta Chiropterologica 3, 253–256.

Corcoran, A.J., 2007. Automated acoustic identification of nine bat species of the eastern United States. M.S. thesis, Humboldt State University, Arcata, CA, USA.

Cristianini, N., Shawe-Taylor, J., 2000. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press, Cambridge, UK.

Cutler, D.R., Edwards Jr., T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. Ecology 88, 2783–2792.

Diamantaras, K.I., Kung, S.Y., 1996. Principal Components Neural Networks: Theory and Applications. John Wiley and Sons, inc., New York, USA.

Drake, J.M., Randin, C., Guisan, A., 2006. Modelling ecological niches with support vector machines. Journal of Applied Ecology 43, 424–432.

Erickson, J.L., West, S.D., 2003. Associations of bats with local structure and landscape features of forested stands in western Oregon and Washington. Biological Conservation 109, 95–102.

Fagerlund, S., in press. Bird species recognition using support vector machines. EURASIP Journal on Advances in Signal Processing. Article ID 38637, doi:10.1155/2007/38637.

Fenton, M.B., 1970. A technique for monitoring bat activity with results obtained from different environments in southern Ontario. Canadian Journal of Zoology 48, 847–851.

Fenton, M.B., 2000. Choosing the 'correct' bat detector. Acta Chiropterologica 2, 215–224.

Fenton, M.B., Bell, G.P., 1981. Recognition of species of insectivorous bats by their echolocation calls. Journal of Mammalogy 62, 233–243.

Fielding, A.H., 1999. How should accuracy be measured? In: Fielding, A.H. (Ed.), Machine Learning Methods for Ecological Applications. Kluwer Academic Publishers, Boston, MA, USA, pp. 209–224.

Fukui, D., Agetsuma, N., Hill, D.A., 2004. Acoustic identification of eight species of bat (Mammalia: Chiroptera) inhabiting forests of southern Hokkaido, Japan: potential for conservation monitoring. Zoological Science 21, 947–955.

Gannon, W.L., O'Farrell, M.J., Corben, C., Bedrick, E.J., 2004. Call character lexicon and analysis of field recorded bat echolocation calls. In: Thomas, J., Moss, C., Vater, M. (Eds.), Echolocation In Bats and Dolphins. University of Chicago Press, Chicago, Illinois, pp. 478–484.

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., Tatham, R.L., 2006. Multivariate Data Analysis. Pearson Education Inc., Upper Saddle River, NJ, USA.

Hayes, J.P., Ober, H.K., Sherwin, R.E., 2009. Survey and monitoring of bats, In: Kunz, T.H., Parsons, S. (Eds.), Ecological and Behavioral Methods for the Study of Bats, second ed. Johns Hopkins University Press, Baltimore, MD, USA, pp. 112–129.

Herr, A., Klomp, N.I., Atkinson, J.S., 1997. Identification of bat echolocation calls using a decision tree classification system. Complexity International 4, 1–14.

Huang, C., Yang, Y., Yang, D., Chen, Y., 2009. Frog classification using machine learning techniques. Expert Systems with Applications 36, 3737–3743.

Humes, M.L., Hayes, J.P., Collopy, M.W., 1999. Bat activity in thinned, unthinned, and old-growth forests in western Oregon. Journal of Wildlife Management 63, 533–561.

Hurlbert, S.H., 1984. Pseudoreplication and the design of ecological field experiments. Ecological Monographs 54, 187–211.

Jennings, N., Parsons, S., Pocock, M.J.O., 2008. Human vs. machine: identification of bat species from their echolocation calls by humans and by artificial neural networks. Canadian Journal of Zoology 86, 371–377.

Kalcounis, M.C., Hobson, K.A., Brigham, R.M., Hecker, K.R., 1999. Bat activity in the boreal forest: importance of stand type and vertical strata. Journal of Mammalogy 80, 673–682.

Karanth, K.U., Nichols, J.D., 1998. Estimation of tiger densities in India using photographic captures and recaptures. Ecology 79, 2852–2862.

Kazial, K.A., Burnett, S.C., Masters, W.M., 2001. Individual and group variation in echolocation calls of big brown bats (Eptesicus fuscus) (Chiroptera: Verspertilionidae). Journal of Mammalogy 82, 339–351.

Kirkhope, C.L., Williams, R.L., Catlin-Groves, C.L., Rees, S.G., Montesanti, C., Jowers, J., Stubbs, H., Newberry, J., Hart, A.G., Goodenough, A.E., Stafford, R., 2010. Social networking for biodiversity: the BeeID project. International Conference on Information Society, June 28–30, 2010, London, U.K, pp. 637–638.

Krusic, R.A., Neefus, C.D., 1996. Habitat associations of bat species in the White Mountain National Forest. In: Barclay, R.M.R., Brigham, R.M. (Eds.), Bats and forests symposium, October 19–21, 1995. Victoria, British Columbia, Canada, pp. 185–198.

Kuhn, T.S., 1970. The Structure of Scientific Revolutions, second edition. The University of Chicago Press, Chicago, IL, USA.

Kuhn, M., 2008. Building predictive models in R using the caret package. Journal of Statistical Software 28, 1–26.

Kuhn, M., 2009. caret: Classification and Regression Training. R package version 4.30 http://cran.r-project.org/web/packages/caret/index.html.

Laiolo, P., 2010. The emerging significance of bioacoustics in animal species conservation. Biological Conservation 143, 1635–1645.

Lance, R.F., Bollich, B., Callahan, C.L., Leberg, P.L., 1996. Surveying forest-bat communities with Anabat detectors. In: Barclay, R.M.R., Brigham, R.M. (Eds.), Bats and forests symposium, October 19–21, 1995. Victoria, British Columbia, Canada, pp. 175–184.

Lek, S., Giraudel, J.L., Guégan, J.F., 2000. Neuronal networks: algorithms and architectures for ecologists and evolutionary ecologists. In: Lek, S., Guégan, J.F. (Eds.), Artificial Neuronal Networks: Application to Ecology and Evolution. Springer-Verlag, Berlin, Germany, pp. 3–27.

Limpens, H.J.G.A., 2004. Choosing a bat detector: theoretical and practical aspects. In: Brigman, R.M., Kalko, E.K.V., Jones, G., Parsons, S., Limpens, H.J.G.A. (Eds.), Bat Echolocation Research: Tools, Techniques, Analysis. Bat Conservation International, Austin, TX, USA, pp. 28–37.

Marks, C.S., Marks, G.E., 2006. Bats of Florida. University Press of Florida, Gainesville, FL, USA.

Menzel, J.M., Menzel Jr., M.A., Kligo, J.C., Ford, W.M., Edwards, J.W., McCracken, G.F., 2005. Effect of habitat and foraging height on bat activity in the coastal plain of South Carolina. Journal of Wildlife Management 69, 235–245.

Meyer, D., Leisch, F., Hornik, K., 2003. The support vector machine under test. Neurocomputing 55, 169–186.

Mukherjee, S., 2006. Classifying Microarray Data Using Support Vector Machines, Understanding And Using Microarray Analysis Techniques: A Practical Guide. Kluwer Academic Publishers, Boston, MA, USA.

Mundry, R., Sommer, C., 2007. Discriminant function analysis with nonindependent data: consequences and an alternative. Animal Behaviour 74, 965–976.

Nicodemus, K.K., Malley, J.D., 2009. Predictor correlation impacts machine learning algorithms: implications for genomic studies. Bioinformatics 25, 1884–1890.

O'Farrell, M.J., Miller, B.W., Gannon, W.L., 1999a. Qualitative identification of free-flying bats using the Anabat detector. Journal of Mammalogy 80, 11–23.

O'Farrell, M.J., Corben, C., Miller, B.W., Gannon, W.L., 1999b. Confronting the dogma: a reply. Journal of Mammalogy 80, 297–302.

O'Farrell, M.J., Corben, C., Gannon, W.L., 2000. Geographic variation in the echolocation calls of the hoary bat (Lasiurus cinereus). Acta Chiropterologica 2, 185–196.

Obrist, M.K., Boesch, R., Flückiger, P.F., 2004. Variability in echolocation call design of 26 Swiss bat species: consequences, limits, and options for automated field identification with a synergetic pattern recognition approach. Mammalia 68, 307–322.

Parsons, S., Jones, G., 2000. Acoustic identification of twelve species of echolocating bat by discriminant function analysis and artificial neural networks. Journal of Experimental Biology 203, 2641–2656.

Parsons, S., Szewczak, J.M., 2009. Detecting, recording, and analyzing the vocalizations of bats. In: Kunz, T.H., Parsons, S. (Eds.), Ecological and Behavioral Methods for the Study of Bats, second ed. Johns Hopkins University Press, Baltimore, MD, USA, pp. 91–111.

Preatoni, D.G., Nodari, M., Chirichella, R., Tosi, G., Wauters, L.A., Martinoli, A., 2005. Identifying bats from time-expanded recordings of search calls: comparing classification methods. Journal of Wildlife Management 69, 1601–1614.

R Development Core Team, 2009. R: A language and environment for statistical computing http://www.r-project.org/]. R Foundation for Statistical Computing, Vienna, Austria.

Redgwell, R.D., Szewczak, J.M., Jones, G., Parsons, S., 2009. Classification of echolocation calls from 14 species of bat by support vector machines and ensembles of neural networks. Algorithms 2, 907–924.

Ripley, B.D., 1996. Pattern Recognition and Neural Networks. Cambridge University Press, Cambridge, UK.

Russo, D., Jones, G., 2002. Identification of twenty-two bat species (Mammalia: Chiroptera) from Italy by analysis of time-expanded recordings of echolocation calls. Journal of the Zoological Society of London 258, 91–103.

Samarasinghe, S., 2007. Neural Networks for Applied Sciences and Engineering: From Fundamentals to Complex Pattern Recognition. Taylor and Francis Group, LLC, Boca Raton, FL, USA.

Skowronski, M.D., Fenton, M.B., 2009. Quantifying bat call detection performance of humans and machines. Journal of the Acoustical Society of America 125, 513–521.

Skowronski, M.D., Harris, J.G., 2006. Acoustic detection and classification of micro-chiroptera using machine learning: lessons learned from automatic speech recognition. Journal of the Acoustic Society of America 119, 1817–1833.

Stocker, R.C., 1998. The role of artificial neural networks in the analysis of ultrasonic bat calls: a case study. Complexity International 5, 1–15.

Vaughan, N., Jones, G., Harris, S., 1997. Identification of British bat species by multivariate analysis of echolocation call parameters. Bioacoustics 7, 189–207.

Veelenturf, L.P.J., 1995. Analysis and Applications of Artificial Neural Networks. Prentice Hall International, Ltd., Hertfordshire, UK.

Weller, T.J., Lee, D.C., 2007. Mist net effort required to inventory a forest bat assemblage. Journal of Wildlife Management 71, 251–257.

White, H., 1992. Artificial Neural Networks: Approximation and Learning Theory. Blackwell, Cambridge, MA, USA.

Zhu, J., Zou, H., 2007. Variable selection for the linear support vector machine. In: Chen, K., Wang, L. (Eds.), Trends in Neural Computation, Studies in Computation Intelligence 35. Springer-Verlag, Berlin, Germany, pp. 35–60.