Classification of red hind grouper call types using random ensemble of stacked autoencoders

Ali K. Ibrahim, Hanqi Zhuang, Laurent M. Chérubin, Michelle T. Schärer Umpierre, Ali Muhamed Ali, Richard S. Nemeth, and Nurgun Erdol

Citation: The Journal of the Acoustical Society of America **146**, 2155 (2019); doi: 10.1121/1.5126861
View online: https://doi.org/10.1121/1.5126861
View Table of Contents: https://asa.scitation.org/toc/jas/146/4
Published by the Acoustical Society of America

Classification of red hind grouper call types using random ensemble of stacked autoencoders

Ali K. Ibrahim,[1,a)] Hanqi Zhuang,[1] Laurent M. Chérubin,[2] Michelle T. Schärer Umpierre,[3] Ali Muhamed Ali,[1] Richard S. Nemeth,[4] and Nurgun Erdol[1]
[1]*Department Computer and Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, Florida 33431, USA*
[2]*Harbor Branch Oceanographic Institute, Florida Atlantic University, 5600 US1 North, Fort Pierce, Florida 34946, USA*
[3]*Department of Marine Sciences, University of Puerto Rico, Mayagüez 00681, Puerto Rico*
[4]*Center for Marine and Environmental Studies, University of Virgin Islands, Saint Thomas, United States Virgin Islands*

(Received 23 April 2019; revised 30 August 2019; accepted 3 September 2019; published online 3 October 2019)

In this paper, a method is introduced for the classification of call types of red hind grouper, an important fishery resource in the Caribbean that produces sounds associated with reproductive behaviors during yearly spawning aggregations. For the undertaken task, two distinct call types of red hind are analyzed. An ensemble of stacked autoencoders (SAEs) is then designed by randomly selecting the hyperparameters of SAEs in the network. These hyperparameters include a number of hidden layers in each SAE and a number of nodes in each hidden layer. Spectrograms of red hind calls are used to train this randomly generated ensemble of SAEs one at a time. Once all individual SAEs are trained, this ensemble is used as a whole to classify call types of red hind. More specifically, the outputs of individual SAEs are combined with a fusion mechanism to produce a final decision on the call type of the input red hind sound. Experimental results show that the innovative approach produces superior results in comparison with those obtained by non-ensemble methods. The algorithm reliably classified red hind call types with over 90% accuracy and successfully detected some calls missed by human observers. © *2019 Acoustical Society of America*.
https://doi.org/10.1121/1.5126861

[JAS] Pages: 2155–2162

I. INTRODUCTION

Fisheries of the Caribbean and tropical Atlantic Ocean depend heavily on hinds and groupers of the family Epinephelidae. The population of red hind (*Epinephelus guttatus*) in the United States (U.S.) Caribbean supports one of the most valuable fisheries today after Nassau grouper (*Epinephelus striatus*) declined significantly (Cummings *et al.*, 1997; Sadovy *et al.*, 1994). Similar to the Nassau grouper, red hind exhibit characteristics that make them vulnerable to overfishing, such as slow growth, long lived, protogynous sexual strategy, and they form transient spawning aggregations (Colin *et al.*, 1987; Manooch, 1987; Sadovy, 1992; Shapiro *et al.*, 1993). Historically, red hind were caught most frequently during spawning aggregations (Colin *et al.*, 1987; Shapiro *et al.*, 1993). Fish spawning aggregations (FSAs) share common features such as strong site fidelity exhibited by spawning fish, and geomorphological attributes (i.e., shelf-break, capes) that make them highly predictable in space and time (Kobara *et al.*, 2013; Starr *et al.*, 2007). Fishing during FSAs has caused declines in the number of fish and a female bias in the sex ratio due to the capture of larger-sized males (Beets and Friedlander, 1999). Some protective measures, such as establishing marine protected areas and imposing fishery bans during the red hind's previously documented spawning season (December–February), have been implemented by fishery managers with mixed results in the region (Beets and Friedlander, 1999; Marshak and Appeldoorn, 2007; Nemeth, 2005). Areas closed to fishing year-round in the U.S. Virgin Islands (USVI) have resulted in an increase in the size of the red hind as well as a decrease in the female bias of the sex ratio (Beets and Friedlander, 1999; Nemeth, 2005). Seasonal marine protected areas in western Puerto Rico suggested an initial increase in catch per unit effort, length frequency, and sex ratios, but the long-term effects have not been evidenced (Marshak and Appeldoorn, 2007). In order for seasonal bans to be effective and help meet fishery management goals, the timing should include the full extent of the reproductive activity when red hind are aggregated to spawn.

Fishes are known to produce sound in a variety of behavioral contexts including courtship, threats, defending territory, and other reproductive behaviors (Lobel *et al.*, 2010). Passive acoustic sensors deployed underwater record low frequency bands of ambient sounds from which fish sound signals can be identified. These signals have been associated with some species during specific behaviors by means of synchronous video recordings, which have revealed that fish species can produce different sound types. The analysis of underwater ambient sound recordings during FSAs of species with known reproductive activity patterns (Mann *et al.*, 2010; Rowell *et al.*, 2012; Rowell *et al.*, 2015; Schärer *et al.*, 2012a; Schärer *et al.*, 2014) has become a useful approach in addition to underwater visual observations to

a)Electronic mail: Aibrahim2014@fau.edu

monitor FSAs documenting fish presence, reproductive activity, and residence time. This approach has also been applied to search areas where FSAs likely exist and detect the species undergoing behaviors associated with reproduction (Rowell *et al.*, 2011). This combination of species, sounds, and sites can help detect additional aggregation sites as well as study the status of known FSAs (Ladich, 2004). The most important application of this type of research is to measure the spatio-temporal variability of grouper reproductive behaviors and assess the timing of FSAs with the timing of seasonal protective measures.

Red hind is a well-known sound-producing species with an acoustic repertoire consisting of multiple sound types (Mann *et al.*, 2010; Zayas-Santiago *et al.*, 2019), with most calls consist of a combination of pulses, pulse trains, and tones. Most studies reported in the literature (Aalbers and Drawbridge, 2008; Mann *et al.*, 2010; Radford *et al.*, 2015; Schärer *et al.*, 2012b; Schärer *et al.*, 2012a;Schärer *et al.*, 2014; Thorson and Fine, 2002; Tricas and Boyle, 2014) investigated peak frequency and duration of either full calls or a selected portion of each call. These measurements may not give a complete description of red hind calls, and they lack important characteristics that differentiate red hind sounds from those of other species. Therefore, a more detailed analysis of red hind calls is needed for the detection and accurate classification of red hind sound types.

Passive acoustics has long been used to detect and classify underwater sounds. In Ibrahim *et al.* (2018a), time-frequency features, such as Mel frequency coefficients and multiresolution acoustic features, have been used to classify grouper sounds. Recently, a deep learning approach was used by Ibrahim *et al.* (2018b) to improve the performance of grouper sound classification. Researchers have also applied unsupervised classification methods to determine the underlying representation in the input data without labeled data. One method in this solution category is stacked autoencoders (SAEs; Liu *et al.*, 2016; Sun *et al.*, 2016). To learn features using autoencoders (AEs), one must first encode the input to a vector of lower dimension and then decode this vector to an approximation of the original input. In contrast to other generative models, such as hidden Markov models (HMMs), which represent the sequential structure of sound signals, AEs learn latent representation from the input data. Moreover, HMMs are statically inefficient for modeling data that lie on a nonlinear manifold in the feature space (Baker *et al.*, 2009), which limits their performance. On the other hand, methods based on deep learning, including AEs, normally need more training data samples to achieve superior classification performances.

In this paper, a random ensemble of stacked autoencoders (RESAE) is designed to classify two red hind grouper sound types. In a RESAE, a number of SAEs are randomly generated to form an ensemble of classifiers. The outputs of these models are then fed to a fusion mechanism to decide the label of a given input. The main contributions of this work are as follows: (i) It provides more detailed analysis of red hind grouper call types. (ii) Due to the fact that the hyperparameters of each SAE are randomly chosen, the algorithm is easy to implement. (iii) The performance of this ensemble of models is superior over that of any individual models in the ensemble. The paper is organized as follows. In Sec. II, a detailed analysis of red hind grouper sounds is given. The concept of AEs is presented in Sec. III. The structure of SAEs is presented in Sec. IV. Experimental results with RESAEs are presented and discussed in Sec. IV. Concluding remarks are given in Sec. V.

II. RED HIND SOUND TYPES

Red hind grouper produce at least four distinct types of sound that are most commonly heard during the FSA. They consist of a combination of what we labeled RH1 and RH2 or either one of them or part of one of them. Type 1 of red hind calls (RH1) consists of pulses produced at a variable pulse rate [Fig. 1(a)]. While the number of pulses is around 25 per second, the period between pulses increases exponentially, which leads to frequency decrease, as seen Fig. 1(a), in the frequency domain signal. The duration of each pulse is around 0.01 s and the peak frequency is around 180—200 Hz, as shown in Figs. 1(b) and 1(c). The frequency distribution of a single pulse varies between 40 and 280 Hz with a 200 Hz peak frequency as shown in Fig. 1(c).

Type 2 red hind calls (RH2) consist of tonal calls. From Fig. 2, one can see there are two peak frequencies: the first one is around 100 Hz, and the second is around 180 Hz. A classification algorithm can be designed to explore the distinct temporal-frequency characteristics of the call types.

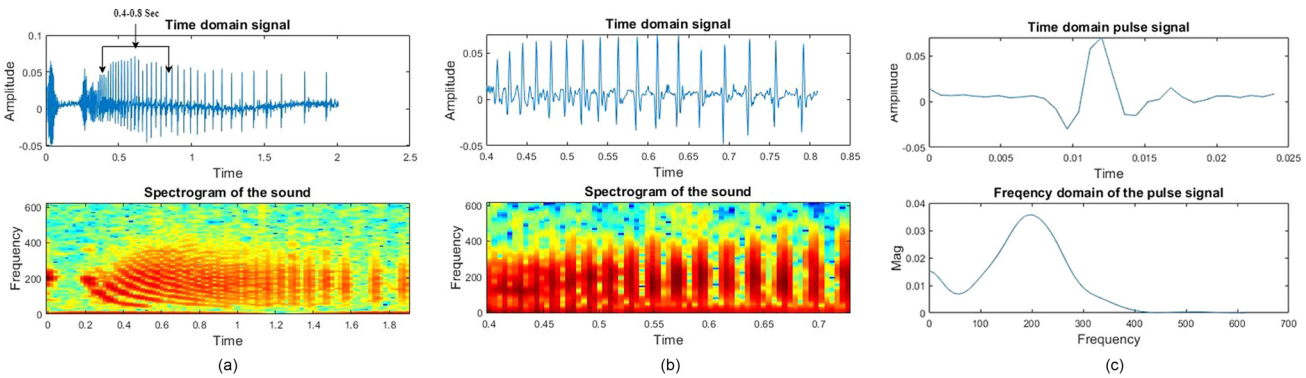

FIG. 1. (Color online) (a) Two-second oscillogram (top) and spectrogram (bottom) of RH1 type calls. Arrows on oscillogram show the 0.4-s oscillogram (top) and spectrogram (bottom) windows shown in (b). (c) Magnitude of one pulse in the time domain (top) and frequency domain (bottom).

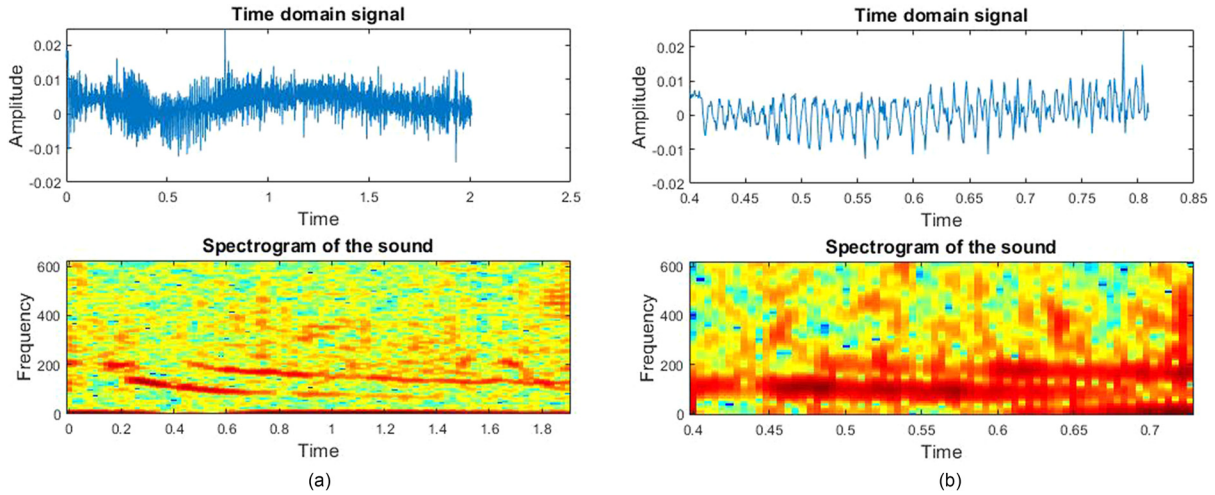

FIG. 2. (Color online) (a) Two-second oscillogram (top) and spectrogram (bottom) of RH2 type calls. Arrows on oscillogram show the 0.4-s oscillogram (top) and spectrogram (bottom) windows shown in (b).

A. Sparse auto encoder

An AE is an unsupervised neural network with three layers, namely, the input layer that takes an input signal, the hidden layer that represents learned features, and the output layer, which has same dimension as that of the input layer, that reconstruct the input signal. The input and hidden layers that form the encoder network are responsible for transforming input signals into features, whereas the output layer forms the decoder network responsible for reconstructing the original input signals. Figure 3 shows the architecture of an AE as described in Tricas and Boyle (2014).

For each input vector x^d from datasets $\{x^d\}_{d=1}^M$, the representation vector h^d and the reconstructed vector $\hat{x}^d$ can be defined as

$$h^d = f\left(W^{(1)}x^d + b^{(1)}\right), \tag{1}$$

$$\hat{x}^d = f\left(W^{(2)}h^d + b^{(2)}\right), \tag{2}$$

where $W^{(1)}$ and $W^{(2)}$ are the weight matrices, $b^{(1)}$ and $b^{(2)}$ are the bias vectors, and f is the activation function. The sigmoid function f is used in this study. The reconstruction error $L(x^d, \hat{x}^d)$ between x^d and $\hat{x}^d$ is defined as

$$L(x^d, \hat{x}^d) = \frac{1}{2}\|x^d - \hat{x}^d\|^2. \tag{3}$$

The overall cost function of the M samples can be defined as

$$J(W,b) = \left[\frac{1}{M}\sum_{d=1}^{M} L(x^d, \hat{x}^d)\right] + \frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{S_l}\sum_{j=1}^{S_{l+1}}\left(W_{ij}^l\right)^2. \tag{4}$$

The first term of Eq. (4) denotes the reconstruction error of the whole dataset, and the second term is the regularization weight penalty term, which aims to prevent over-fitting by restraining the weights magnitude. λ is the weight decay parameter, l is the layer number of the network, S_l denotes the neuron number in layer l, and W_{ij}^l is the connecting weight between neuron i in layer $l+1$ and neuron j in layer l.

Clearly, the output of an AE simply copies its input; that is, although the learned feature representations may perfectly reconstruct the original input, the features may be redundant. A way to solve the problem is to add a sparsity penalty term to the cost function of AEs, resulting in a sparse AE, which has a great potential to learn more abstract and representative features of the input. Equation (5) shows the overall cost function of the AE, where $J(W,b)$ is shown before as Eq. (4), and the second term is the sparsity penalty term, where $\hat{\rho}_g(g = 1, 2, \ldots, e)$ is the average activation value of the hidden unit g, which is defined in Eq. (6), ρ is the sparsity parameter of a small value, and β is the sparsity penalty term parameter used to control the relative importance between the first reconstruction term and the second penalty term.

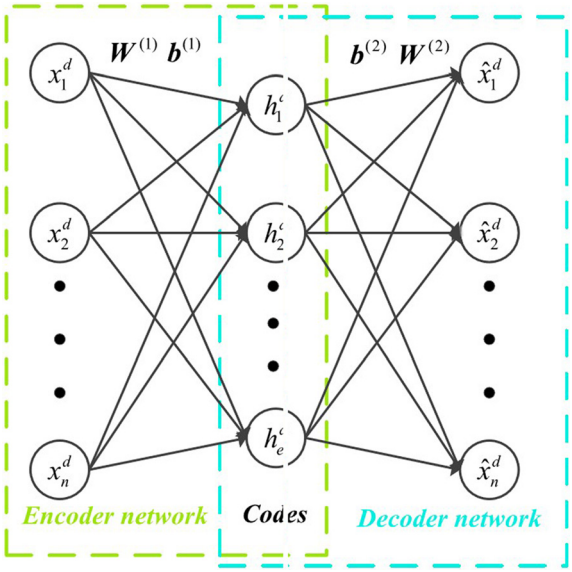

FIG. 3. (Color online) Architecture of AE. AE consist of three layers: input (x), hidden (h), and output layer ($\hat{x}$) The input and hidden layers encode the input data to extract the features, and the hidden and output layers decode the features to reconstruct input data. x^d is the input vector from datasets $\{x^d\}_{d=1}^M$, h^d is the representation vector, and $\hat{x}^d$ is the reconstructed vector. $W^{(1)}$ and $W^{(2)}$ are the weight matrices, $b^{(1)}$ and $b^{(2)}$ are the bias vectors.

$$J_{sp}(W,b) = J(W,b) + \beta \sum_{g=1}^{e} \left(\rho \log \frac{\rho}{\hat{\rho}} + (1-\rho) \log \frac{1-\rho}{1-\hat{\rho}_g} \right), \tag{5}$$

$$\hat{\rho}_g = \frac{1}{M} \sum_{d=1}^{M} h_g^d. \tag{6}$$

The sparse AE aims to learn more representative and sparse features, which can extract more information from input signals. The network is trained by minimizing the cost function $J_{sp}(W,b)$ using the back propagation (BP) algorithm. The optimal parameter sets $W^{(1)}$, $W^{(2)}$, $b^{(1)}$, and $b^{(2)}$ can be learned simultaneously.

III. SYSTEM IMPLEMENTATION

A RESAE is a new technique that can be used for classification of grouper sound types. The structure of a RESAE is shown in Fig. 4, which contains multiple SAEs. The number of SAEs depends on a particular application.

For each SAE, one specifies a range in which two hyperparameters, number of AEs in each SAE (L), and number of neurons in each hidden layer are randomly generated. The design of a randomized SAE is done as follows. One first generates L random numbers, each representing the number of neurons for one AE. To ensure that the dimension of the captured features is decreasing after each layer, these L random numbers are sorted in a descending order before being used as the number of hidden neurons of individual AEs. After all individual AEs are generated, they are stacked together to form a SAE. Given N, the number of SAEs to be generated, and the hyperparameter parameter ranges, the algorithmic steps to train the RESAE are given as follows:

(1) Generate all randomly selected hyperparameters from the respective parameter ranges.

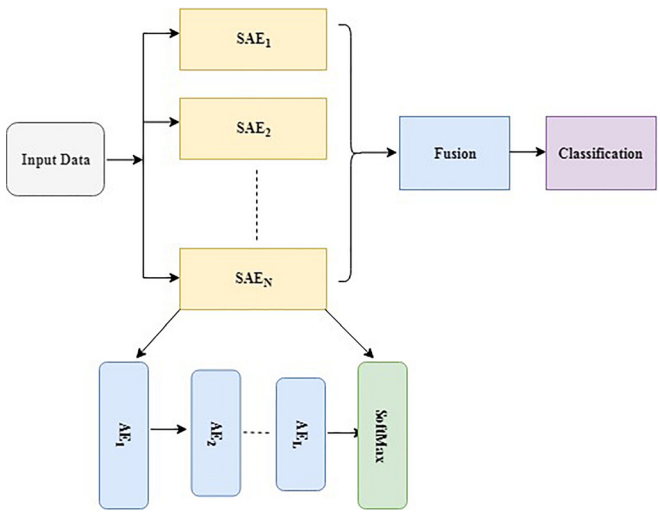

FIG. 4. (Color online) Classification with a random ensemble of SAEs. The system consists of N number of SAEs with a soft max layer. The output of each SAE is fused using fusion methods such as majority voting, unweighted, and weighted averaging.

(2) Generate N SAEs with the randomized hyperparameters.
(3) Train these SAEs one by one with the same training dataset until all SAEs are trained.

In the test phase, the following steps are used:

(1) For each input, feed the input vector simultaneously to all SAEs.
(2) In each SAE, extract its predicted labels together with the output of the softMax layer corresponding to the input.
(3) Pipe the outputs of all SAEs into a fusion mechanism to make a final decision on the label of the input.

A simple fusion mechanism taking the averaging values of the outputs of all SAEs for binary classification 0 or 1 of an input is described in the following equation:

$$M(y_{i1}, y_{i2}, \ldots, y_{iN}) = \left[\frac{1}{2} + \frac{\sum_{j=1}^{N} y_{ij} - \frac{1}{2}}{N} \right], \tag{7}$$

where y_{ij} is the output prediction for data point i by the jth SAE. The equation can be easily modified for other fusion methods such as weighted averaging and majority vote.

The resulting ensemble of SAEs is both computationally less expensive and structurally simpler than a very deep SAE. In addition, the proposed structure creates a generalized architecture, which is able to adapt to data, making it less likely to overfit. By relying on a wide range of hyperparameters and randomization, our system effectively learns the best network architecture. Furthermore, our methodology does not simply default to the best chosen model; instead the outputs of individual models are combined to make a final decision through a fusion mechanism. This ensures that the system takes full advantage of the strengths of both good models and outliers.

IV. EXPERMENTAL RESULTS

A. Dataset

Two types of data are used in this study. The first one was obtained from bottom mounted hydrophones deployed from December 2013 to June 2014 on the west coast of Puerto Rico at Abrir La Sierra (ALS), which is located off the west coast of Puerto Rico, in the Mona Passage (Fig. 5; Rowell *et al.*, 2011). Ambient sounds were recorded every 5 min for 20 s at a sampling rate of 10 kHz. The files were, with presence of red hind courtship associated sounds (CAS), detected manually by visualizing spectrogram and listening with noise canceling headphones to 12 files recorded during one hour, between 22:00 and 23:00 GMT [18:00–19:00 Alaska standard time (AST)] for ALS. The second dataset was recorded from a surface moving platform, the SV3 wave glider (WG), which is equipped with a passive acoustic monitoring (PAM) system that records and classifies fish sounds in real time (Chérubin *et al.*, 2017). The WG was deployed at two known spawning aggregation sites, including ALS and south of St. Thomas in the USVI (Fig. 5). The fish sounds were recorded on a hydrophone trailed 10 m behind the WG and 8–15 m below the surface.

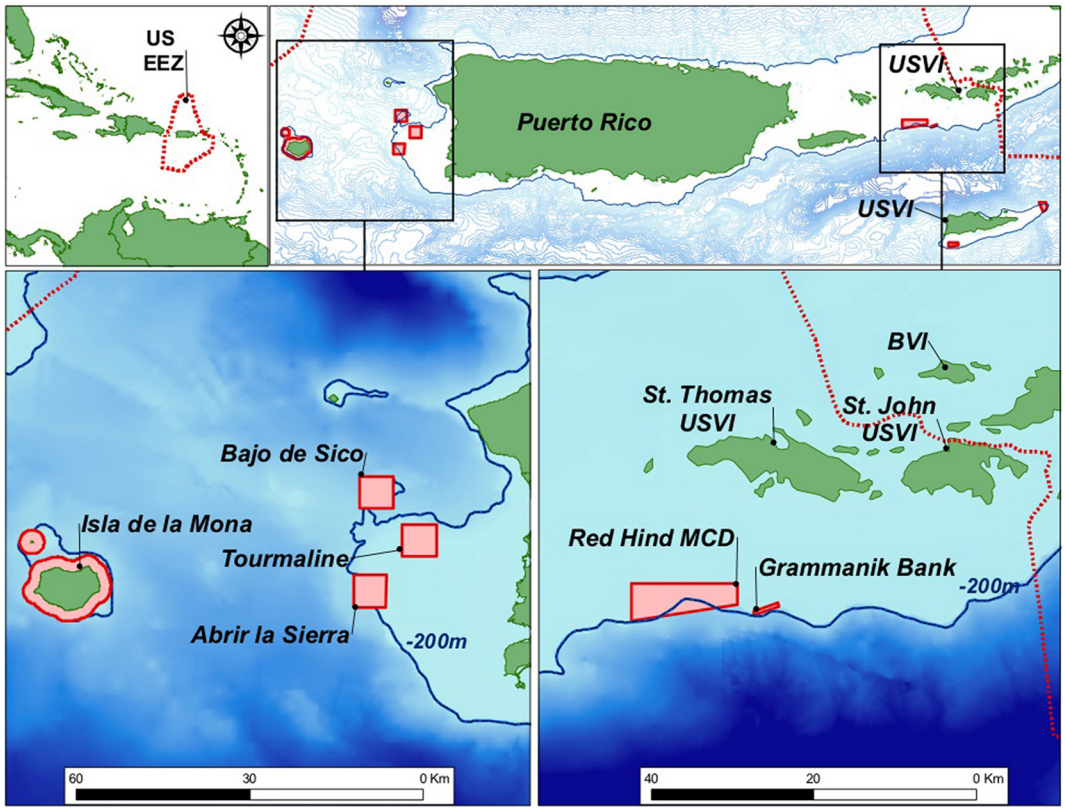

FIG. 5. (Color online) Maps identifying the locations where the acoustic data were collected, namely ALS on the Puerto Rican western shelf and along the edge of the southern shelf in the USVI both inside and outside the red hind Marine Conservation District (MCD) and the Grammanik Bank.

The PAM system records 10-s audio files every 30 s at a sampling rate of 10 kHz. Figure 6 shows the spectrograms of red hind sounds collected in the USVI and Puerto Rico from the mobile platform.

B. Results

The results of the human detection analysis were totaled by day by adding the files with red hind call types within each calendar day. These were then compared to the algorithm detection results of red hind call types. Table I summarizes the number of files detected manually with red hind call types per month.

Two important measures for the evaluation are detection rate and false-positive rate. Detection rate (also known as sensitivity or recall) is defined as the ratio between the number of true-positive events, i.e., the number of red hind calls correctly identified, and the total number of red hind calls in the recording set [true positives (TP) and false negatives]. False-positive (FP or false-alarm) rate is defined as the ratio

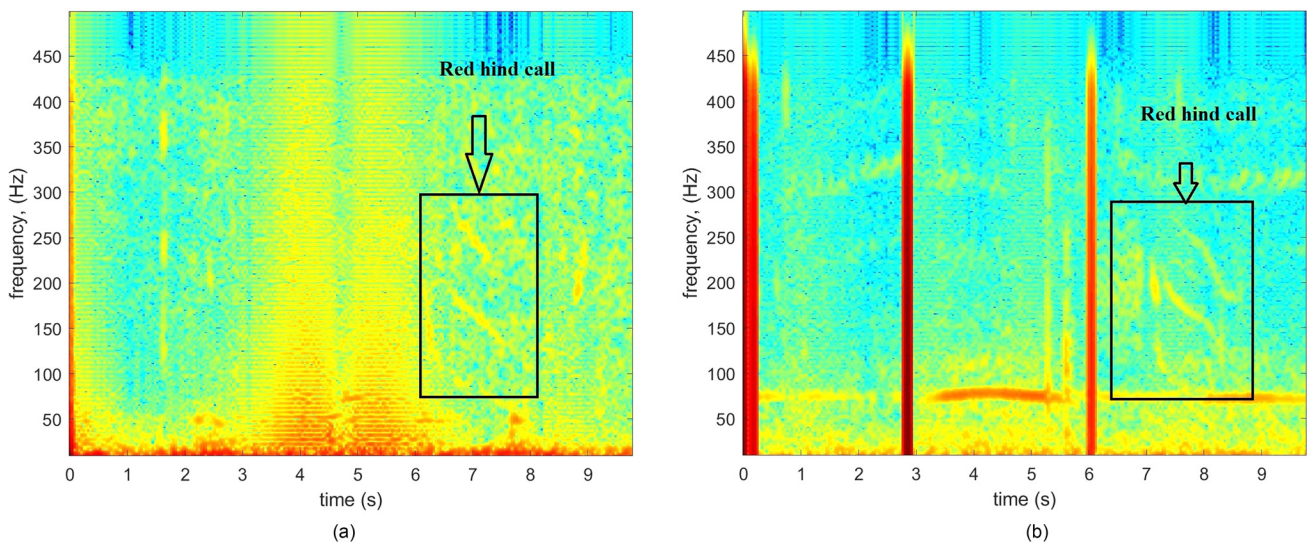

FIG. 6. (Color online) Examples of RH2 type spectrograms collected from the WG in the USVI (a) and Puerto Rico (b).

TABLE I. Number of files with calls identified by human.

Month	Number of files with calls
December	56
January	175
February	57
March	1
April	4
May	2

between the number of false positives (non-red-hind calls identified erroneously as red hind calls) and the total number of non-red-hind calls in the recording set (including true negatives). Thus, if TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively, then the detection rate is TP/(TP + FN), and the false-positive rate is FP/(FP + TN). The importance of obtaining a high detection rate is obvious. However, a low false-positive rate is perhaps even more important in order to avoid the contamination of data with non-related signals, which may alter the interpretation of results. In the experiment, we first converted the raw sound segments into spectrograms using a 0.1-s frame length (1000 samples) with 70% overlapping. After framing, we applied first a Hanning window and then fast Fourier transform (FFT) with the number of points being 16 384 to obtain spectrograms with a high frequency resolution. For each spectrogram, we selected a spectrum with the frequency range from 10 to 400 Hz, and then applied log 10 to the spectrum to obtain a log spectrogram.

In the experiment, we used 9900 red hind calls, and 29 500 segments of noise and other sounds. In addition, we used a ratio of 7:1.5:1.5 to divide the data into training, validation, and testing. Figure 7(a) shows the performance of SAEs using two AEs in terms of the mean square error (MSE) for training, validation, and testing. Figure 7(b) shows the performance of SAEs using three AEs.

The evaluation metrics for SAE classifiers are summarized in Table II. In Table II, the first column lists the number of AEs cascaded in the SAE. From Table II, the SAE with two cascading AEs achieved higher detection and lower false alarm rates overall than the other SAE configurations.

To investigate the effectiveness of RESAEs, a wide range of hyperparameters were tested, such as the number of SAEs, number of layers in each SAE, number of neurons in each hidden layer, and batch size. Table III shows the range of hyperparameters used to generate these networks. Majority voting and unweighted averaging have been used as a fusion method. Table IV shows the comparison between the two fusion methods in terms of detection of red hind sounds. Note that the number of SAEs was five for both cases.

We also compared classification results of RH1 and RH2 by using 5, 10, and 15 random multi AEs. The training was done with 400 sounds for each class and 200 sounds for testing. We used different range of hyperparameters (number of autoencodes range [1–5], number of features range [250–1500], and number of epochs [200–400]). Table V provides the results using unweighted average as fusion strategy.

The comparison between human detection and automated detection using a dataset obtained in the USVI is shown in Fig. 8. In this case, the model was trained by the ALS dataset and tested with the USVI dataset; the latter was quite different from the former as described in Sec. IV A. Again, the human detection was done by checking each file's spectrogram and listening to the sounds in each file. The total number of red hind calls was 323, and the number of TP was 281, and number of FP was 51.

V. CONCLUSION

In this paper, two types of red hind sounds, RH1 and RH2, were analyzed. A new call detection and classification algorithm for red hind calls based on the RESAE network is proposed. The algorithm can also reliably classify red hind call types. In addition, since each of the SAEs in the

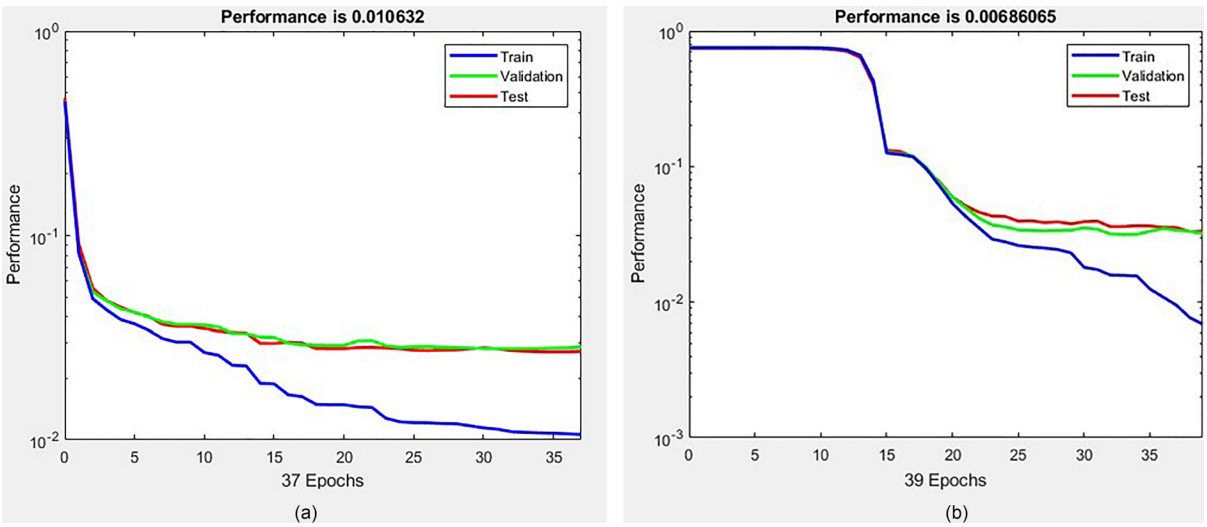

FIG. 7. (Color online) Performance of SAEs. (a) Two AEs were used, and (b) three AEs were used. Note that the horizontal axis denotes the number of iterations in an epoch (each epoch is one iteration of training using the entire training dataset), and the vertical axis denotes mean square error (MSE; 1 means 100% error).

TABLE II. Results obtained by individual SAEs.

Number of AEs	Detection rate	False alarm rate
1	74.81%	9.3%
2	87.46%	5.6%
3	82.16%	7.12%
4	80.4%	8.32%
5	76.17%	9.26%

TABLE III. Ranges of hyperparameters.

Hyperparameter	Range
Number of layers	1–5
Number of neurons	100–1500
Mini = batchsize	<256

TABLE IV. Results of five models RESAE.

Ensemble method	ACC
Majority voting	92%
Unweighted average	94%

TABLE V. Results of the 5, 10, 15 models RESAE.

Number of models	ACC
5	86%
10	93%
15	95.2%

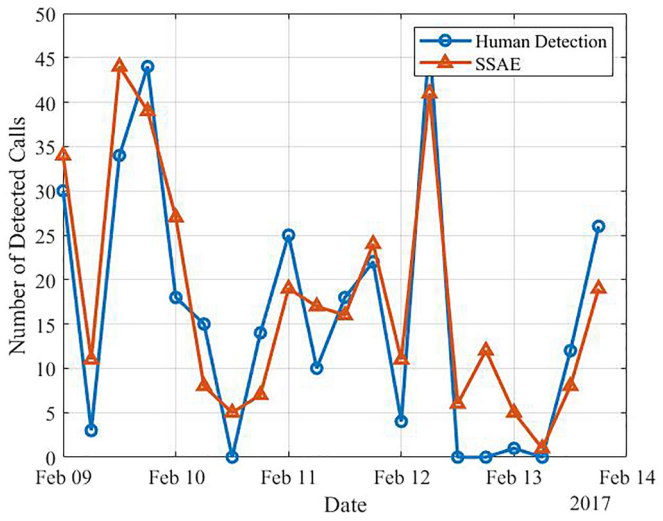

FIG. 8. (Color online) Comparison between human detection and automated detection in the U.S. Virigin Islands datasets.

ensemble is shallow, the RESAE is easy to train and needs few data points to train in comparison to a single deep network. In comparison to a single shallow SAE, the RESAE provides better performance in terms of detection rate and false alarm rate with a slightly higher computational cost, although, as expected, its performance degrades when the test dataset is different from the training dataset. In the proposed method, sparse AEs are used to learn features from sound spectrograms. On the other hand, many existing approaches for fish sound classification relied on physical or statistical features (Malfante *et al.*, 2016; Putland *et al.*, 2018; Ranjard *et al.*, 2017). The method proposed in this paper can be also used for other marine sound detection and classification problems. Because of its small footprint, the RESAE network can be more readily implemented in an embedded device for real-time detection and classification of fish sounds.

ACKNOWLEDGMENTS

The authors acknowledge the Harbor Branch Oceanographic Institute Foundation for supporting part of this research. A.I., L.M.C., M.T.S.U., and R.N. were also supported in part by The National Oceanic and Atmospheric Administration (NOAA) Saltonstall-Kennedy Grant No. NA15NMF4270329. Passive acoustic data were collected with the aid of the University of Puerto Rico, Mayagez campus, the Caribbean Fishery Management Council funding for research, the Caribbean SEAMAP program, and permits provided by the Department of Natural and Environmental Resources (No. 2014-IC-040). We thank the crew of Orca Too, as well as the volunteer divers and students, primarily Tim Rowell, Kimberly Clouse, and Carlos Zayas, who analyzed passive acoustic data. This is contribution number 207 of the University of the Virgin Islands Center for Marine and Environmental Studies.

Aalbers, S. A., and Drawbridge, M. A. (**2008**). "White seabass spawning behavior and sound production," Trans. Am. Fish. Soc. **137**(2), 542–550.

Baker, J., Deng, L., Glass, J., Khudanpur, S., Lee, C.-h., Morgan, N., and O'Shaughnessy, D. (**2009**). "Developments and directions in speech recognition and understanding, Part 1," IEEE Signal Processing Mag. **26**(3), 75–80 [DSP education].

Beets, J., and Friedlander, A. (**1999**). "Evaluation of a conservation strategy: A spawning aggregation closure for red hind, *Epinephelus guttatus*, in the US Virgin Islands," Environ. Biol. Fishes **55**(1-2), 91–98.

Chérubin, L. M., Dalgleish, F., Ibrahim, A., Schärer-Umpierre, M., Nemeth, R., Appeldoorn, R., Ouyang, B., and Dalgleish, A. (**2017**). "Implementation of a passive acoustic monitoring system on a sv3 wave glider and applications," in *Proceedings of the 70th Gulf Caribbean Fisheries Institute*, Vol. 70.

Colin P. L., Shapiro D. Y., and Weiler, D. (**1987**). "Aspects of the reproduction of two groupers, *Epinephelus guttatus* and *E. straitus* in the West Indies," Bull. Mar. Sci. **40**, 220–230.

Cummings, N., Parrack, M., and Zweifel, J. (**1997**). "The status of red hind and coney in the U.S. Virgin Islands between 1974 and 1992," in *Proceedings of the 49th Gulf Caribbean Fisheries Institute*, Vol. 49, pp. 354–397.

Ibrahim, A. K., Chérubin, L. M., Zhuang, H., Schärer Umpierre, M. T., Dalgleish, F., Erdol, N., Ouyang, B., and Dalgleish, A. (**2018a**). "An approach for automatic classification of grouper vocalizations with passive acoustic monitoring," J. Acoust. Soc. Am. **143**(2), 666–676.

Ibrahim, A. K., Zhuang, H.,. Chérubin, L. M., Schärer-Umpierre, M. T., and Erdol, N. (**2018b**). "Automatic classification of grouper species by their

sounds using deep neural networks," J. Acoust. Soc. Am. **144**(3), EL196–EL202.
Kobara, S., Heyman, W. D., Pittman, S., and Nemeth, R. S. (**2013**). "Biogeography of transient fish spawning aggregations in the Caribbean: A synthesis for future research and management," Oceanogr. Mar. Biol. Annu. Rev. **51**, 281–326.
Ladich, F. (**2004**). "Sound production and acoustic communication," in *The Senses of Fish* (Springer, New Delhi, India), pp. 210–230.
Liu, Y., Feng, X., and Zhou, Z. (**2016**). "Multimodal video classification with stacked contractive autoencoders," Signal Process. **120**, 761–766.
Lobel, P. S., Kaatz, I. M., and Rice, A. N. (**2010**). "Acoustical behavior of coral reef fishes," in *Reproduction and Sexuality in Marine Fishes: Patterns and Processes*, edited by C. Kole (University of California Press, California), pp. 307–386.
Malfante, M., Dalla Mura, M., Mars, J. I., and Gervaise, C. (**2016**). "Automatic fish sounds classifcation," J. Acoust. Soc. Am. **139**(4), 2115–2116.
Mann, D., Locascio, J., Schärer, M., Nemeth, M., and Appeldoorn, R. (**2010**). "Sound production by red hind epinephelus guttatus in spatially segregated spawning aggregations," Aquat. Biol. **10**(2), 149–154.
Manooch, C. (**1987**). *Age and Growth of Snappers and Groupers* (Westview, Boulder, CO), pp. 329–373.
Marshak, A., and Appeldoorn, R. S. (**2007**). "Evaluation of seasonal closures of red hind, *Epinephelus guttatus*, spawning aggregations to fishing off the west coast of Puerto Rico using fishery-dependent and independent time series data," in *Proceedings of the 60th Annual Gulf and Caribbean Fisheries Institute*, Vol. 60, pp. 566–572.
Nemeth, R. S. (**2005**). "Population characteristics of a recovering US Virgin Islands red hind spawning aggregation following protection," Mar. Ecol.: Prog. Ser. **286**, 81–97.
Putland, R., Mackiewicz, A., and Mensinger, A. F. (**2018**). "Using passive acoustics to localize vocalizing oyster toadfish (*Opsanus tau*)," J. Acoust. Soc. Am. **144**(3), 1692–1692.
Radford, C. A., Ghazali, S., Jeffs, A. G., and Montgomery, J. C. (**2015**). "Vocalisations of the bigeye, *Pempheris adspersa*: Characteristics, source level, and active space," J. Exp. Biol. **218**, 940–948.
Ranjard, L., Reed, B. S., Landers, T. J., Rayner, M. J., Friesen, M. R., Sagar, R. L., and Dunphy, B. J. (**2017**). "Matlabhtk: A simple interface for bioacoustic analyses using hidden Markov models," Meth. Ecol. Evol. **8**(5), 615–621.
Rowell, T., Appeldoorn, R., Rivera, J., Mann, D., Kellison, T., Nemeth, M., and Schärer Umpierre, M. (**2011**). "Use of passive acoustics to map grouper spawning aggregations, with emphasis on red hind, *Epinephelus guttatus*, off western Puerto Rico," Proc. Gulf Caribb. Fisheries Inst. **63**, 139142.
Rowell, T., Nemeth, R. S., Scharer, M. T., and Appeldoorn, R. S. (**2015**). "Fish sound production and acoustic telemetry reveal behaviors and spatial patterns associated with spawning aggregations of two Caribbean groupers," Mar. Ecol.: Prog. Ser. **518**, 239–254.
Rowell, T., Schärer, M. T., Appeldoorn, R. S., Nemeth, M. I., Mann, D. A., and Rivera, J. A. (**2012**). "Sound production as an indicator of red hind density at a spawning aggregation," Mar. Ecol.: Prog. Ser. **462**, 241–250.
Sadovy, Y., Figuerola, M., and Roman, A. (**1992**). "Age and growth of red hind, *Epinephelus guttatus* in Puerto Rico and St. Thomas," Fish. Bull. Fish. Wildl. Serv. US **90**, 516–528.
Sadovy, Y., Rosario, A., and Roman, A. (**1994**). "Reproduction in an aggregating grouper, the red hind, *Epinephelus guttatus*," in *Women in Ichthyology: An Anthology in Honour of ET, Ro and Genie* (Springer, Netherlands), pp. 269–286.
Schärer, M. T., Nemeth, M. I., Mann, D., Locascio, J., Appeldoorn, R. S., and Rowell, T. J. (**2012a**). "Sound production and reproductive behavior of yellowfin grouper, *Mycteroperca venenosa* (serranidae) at a spawning aggregation," Copeia **2012**(1), 135–144.
Schärer, M. T., Nemeth, M. I., Rowell, T. J., and Appeldoorn, R. S. (**2014**). "Sounds associated with the reproductive behavior of the black grouper (*Mycteroperca bonaci*)," Mar. Biol. **161**(1), 141–147.
Schärer, M., Rowell, T., Nemeth, M., and Appeldoorn, R. (**2012b**). "Sound production associated with reproductive behavior of Nassau grouper *Epinephelus striatus* at spawning aggregations," Endang. Spec. Res. **19**(1), 29–38.
Shapiro, D. Y., Sadovy, Y., and McGehee, M. A. (**1993**). "Size, composition, and spatial structure of the annual spawning aggregation of the red hind, *Epinephelus guttatus* (Pisces: Serranidae)," Copeia **1993**, 399–406.
Starr, R. M., Sala, E., Ballesteros, E., and Zabala, M. (**2007**). "Spatial dynamics of the Nassau grouper *Epinephelus striatus* in a Caribbean atoll," Mar. Ecol.: Prog. Ser. **343**, 239–249.
Sun, W., Shao, S., Zhao, R., Yan, R., Zhang, X., and Chen, X. (**2016**). "A sparse autoencoder-based deep neural network approach for induction motor faults classification," Measurement **89**, 171–178.
Thorson, R. F., and Fine, M. L. (**2002**). "Crepuscular changes in emission rate and parameters of the boatwhistle advertisement call of the gulf toadfish, *Opsanus beta*," Environ. Biol. Fishes **63**(3), 321–331.
Tricas, T. C., and Boyle, K. S. (**2014**). "Acoustic behaviors in Hawaiian coral reef fish communities," Mar. Ecol.: Prog. Ser. **511**, 1–16.
Zayas-Santiago, C. (**2019**). "Red hind *Epinephelus guttatus* vocal repertoire characterization, temporal patterns and call detection with micro accelerometers," MS thesis, University of Puerto Rico, Mayagüez, 68 pp.