



Performance Review of Machine Learning Techniques for the Detection of Humpback Whale Song

¹O. O. Ogundile, ²M. O. Oloyede, ³K. W. Giwa and ⁴O. A. Osanaiye

¹Department of Computer Science, Tai Solarin University of Education, Ijagun, Ogun State, Nigeria.

²Department of Information and Communication Science, University of Ilorin, Ilorin, Kwara State, Nigeria.

³Department of Physics, Tai Solarin University of Education, Ijagun, Ogun State, Nigeria.

⁴Computer Engineering Department, Nile University of Nigeria, Abuja, Nigeria.

Email: ogundileoo@tasued.edu.ng; oloyede.om@unilorin.edu.ng; giwakw@tasued.edu.ng; okeyemi.osanaiye@nileuniversity.edu.ng.

Abstract

In cetacean vocalisations detection, the detection and classification accuracy of Machine Learning (ML) techniques play a vital role for biologist or ecologist to adopt or frequently use the ML technique. The higher the detection and classification accuracy of the ML technique, the more insightful is the decision made by marine ecologist. Accordingly, different ML techniques such as K-nearest neighbour, Gaussian mixture model (GMM), support vector machines (SVMs), hidden Markov models (HMMs), and so on, have been used in the literature to proffer solutions to the detection and classification of cetacean vocalisations. Thus, this paper presents a detailed performance comparison of two different ML techniques (GMM and HMM) that can be employed in the detection and classification of cetacean vocalisations. In addition, these two ML techniques are combined (GMM-HMM), where the result of the combined GMM-HMM technique is analysed in the paper. The performance of these ML techniques is theoretically demonstrated using the acoustic dataset of the continuous recoding of a Humpback whale song gathered off the north coast of Kauai, Hawaii, USA. The simulation results show that the combined GMM-HMM approach exhibited superior detection/classification accuracy in comparison to the GMM and HMM approaches for various experimental conditions. These results will help biologist or ecologist to make discerning selection based on their experimental requirements.

Keywords: Detection accuracy, GMM, HMM, Humpback whale song, GMM-HMM, MFCC, ML,

1. Introduction

Machine learning (ML) approaches to solving signal processing problems have recently gained attention of researchers. ML techniques are widely used in signal processing applications because it can handle multi-dimensional and multi-variety data, and can easily identify different signal trends and patterns without human interventions (automated) (Wuest *et al.*, 2016). Countless ML techniques have been proposed in the literature to solve different signal processing applications. Some examples of these ML techniques include; K-nearest neighbour, Bayesian and statistical methods, support vector machines (SVMs), Gaussian mixture model (GMM), hidden Markov models (HMMs), and so on (Ogundile *et al.*, 2020b; Putland *et al.*, 2018). The performance of these ML techniques depend on the application requirements and formulations. Nevertheless, the performance of the ML techniques can be improved depending on the signal processing application requirements, which is a key advantage of the use of ML techniques. With respect to cetacean vocalisations, ML techniques have been adopted in different forms to analyse and detect cetacean vocal repertoires. These cetacean sounds can be in the form of clicks, pulse calls, songs, and whistles (Brown & J O Miller, 2007). These sounds are gathered over a long duration in a process called passive acoustic monitoring (PAM) (Gibb *et al.*, 2019). As such, the accumulated datasets are too large and impractical to detect and analyse manually. Machine learning tools are thus employed to automatically detect and classify these cetacean vocalisations. In fact, these have motivated the increase in the

design of different automated detection and classification ML techniques for cetacean vocalisation. These ML techniques differ depending on the cetacean species, call type and duration, and the range of frequency (Ogundile *et al.*, 2020a; Babalola *et al.*, 2021). Fig. 1 depicts a simple ML detection/classification system for cetaceans. As shown from the figure, the recorded sound is first preprocessed, then annotations are carried out on the sound signal after the signal must have been recognised by a biologist or ecologist. Reliable features are subsequently extracted from the recognised sound signal before the detection/ classification steps can take place.

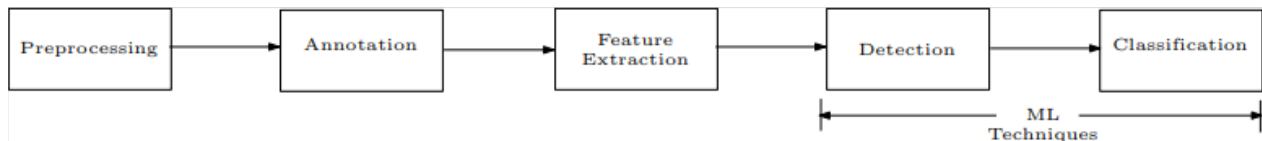


Figure 1: Block diagram of a ML detection system for cetacean vocalisations.

Therefore, this paper review two different ML techniques (GMM, and HMM), that can be used to detect and classify cetacean vocalisations. More so, a detailed performance comparison of these aforementioned ML techniques is presented in the paper. Additionally, these two ML techniques are combined (GMM-HMM) in this paper, where the result of the combined GMM-HMM ML technique is also analysed. The performance of the GMM, HMM and GMM-HMM ML approaches were verified on the acoustic dataset of the continuous recording of a Humpback whale song sampled at 4000Hz, collected off the north coast of Kauai, Hawaii, USA. The contribution of this paper is relevant because the automatic detection of cetacean vocalisations is an emerging research field due to their economic importance and values. Particularly, the whale watching business generates over US\$2 billion yearly. As such, it engages thousand of people and enhances the revenue of the government (Usman *et al.*, 2020; Smith *et al.*, 2019). Therefore, this paper will assist researchers and practitioners in this field to make intuitive decisions based on their design requirements. The rest of this paper is structured as follows. In Section 2, the dataset and problem is explained in details. Section 3 explains the ML techniques used in detecting the humpback whale songs. The experimental set-up is described in Section 4, where the feature extraction (FC) technique adopted in the paper is briefly discussed. Besides, this section describes the combined GMM-HMM ML technique in details. Section 5 presents the simulation test results and discussion. Finally, the paper is summarised in Section 6.

2. Dataset: Humpback Whale Song

Humpback whale, also referred to as *Megaptera novaeangliae* is one of the biggest members of the *Mysticetes* (suborder of cetaceans) species. Humpback whale can be recognised by their physical appearance unlike other *Mysticetes*. They have distinctive shapes, which includes a pectoral fin and a knobbly head. They produce sounds that have the characteristics of a ‘song’ arranged in complex periodic sequences. The song consists of both tonal and pulsed sounds which can change from time to time. Humpback whales are sighted in all oceans; as such, it is important for marine scientist to keep track the information on their population structure. Consequently, the vocal repertoire of Humpback whales is collected across different location in recent years. This has encouraged the design of different ML automated techniques to detect and classify Humpback whale songs. Therefore, this paper examines the detection performance of three ML techniques (GMM, HMM and GMM-HMM) to proffer solution to this problem. The humpback whale song recording documented in mobysound.org, off the north coast of Kauai, Hawaii, USA is used to verify the performance of these ML techniques. Fig. 2 depicts the waveform of the Humpback whale song as viewed from a Sonic Visualiser.

3. ML Techniques

3.1 Gaussian Mixture Model

The GMM is a ML classifier that uses the estimate of the probability density function (PDF) to analyse different type of signals. It has been adopted in different signal processing applications such as image processing, speech processing, and so on. Similarly, GMM has be used in designing different automated detectors and classifiers for

marine mammal sounds. GMM has the capability to randomly model cetacean data (signal) sequence by restyling the parameters and number of Gaussian PDFs (Usman et al., 2020).

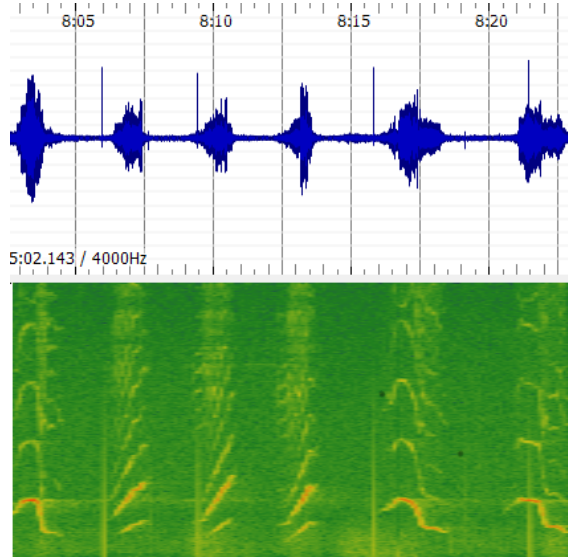


Figure 2: Humpback whale song - Time series and spectrogram representation (Ogundile et al., 2021).

For example, the n Gaussian distribution of a feature vector sequence $\phi = (\phi^1, \phi^2, \dots, \phi^k)^T$ can be represented as (Usman et al., 2020):

$$Pr(y|\mu, \Sigma) = \sum_{i=1}^n \rho_i \frac{1}{(2\pi)^{\frac{1}{2}} |\Sigma_i|^{\frac{1}{2}}} e^{-\frac{1}{2}(\phi - \mu_i)^T \Sigma^{-1} (\phi - \mu_i)} \quad (1)$$

The mean is denoted as μ , k is the dimension of the feature vector, ρ is the mixture weight, Σ is the covariance matrix, Σ^{-1} is the inverse of Σ , $|\Sigma|$ is the determinant of Σ , and ^{T} is the transpose operator. The expectation maximisation algorithm (EM) (Bilmes, 1998) is used to obtain the Gaussian parameters; that is, the μ , Σ and ρ . These Gaussian parameters assumes a flat or random value at the beginning of the EM process. Thus, the K -means clustering (K -MC) algorithm (Forgy, 1965) is used to initialise the Gaussian parameters in to improve the performance of the GMM.

3.2 Hidden Markov Model

The HMM is a ML tool that is widely employed for sound analysis, detection, and classification (Rabiner & Shafer, 1978; Rabiner, 1989). It is flexible, and can be seamlessly used to model, detect, and classify the spectral and temporal features of sound signals. HMM can be perceived as a stochastic state machine, where every change in the hidden states result in an emission of a symbol. The running of the HMM can be broken down into two sequential steps: training step and detection step. During training, the HMM calculates three key parameters from the obtained feature vector of the sound signal; namely, start probability, α , transition matrix, β , and Gaussian parameters, $\gamma = \{\mu, \Sigma, \rho\}$. The β represents the feature extracted from the sound signal as a sequence of states across time. This matrix contains probability values which are used to determine the switch between states. The β can be computed by employing maximum-likelihood estimations such as the Baum-Welch algorithm (BM-alg)

(Baum *et al.*, 1970). The BM-alg is modelled using the Gaussian emission distribution parameters, $\gamma = \{\mu, \Sigma, \rho\}$. Also note that these Gaussian parameters usually assume random or flat values at the start of the BM-alg process. Likewise, the K -MC algorithm is used to initialise the emission distribution parameters to improve the performance of the HMM. In the detection step, the trained HMM parameter (α , β and γ) are used to match the feature vector of the unknown sound signal to be detected. This process is achieved using the Viterbi algorithm (V-alg) (Viterbi, 1967). The V-alg optimally produces the most closely matched hidden sequence of the model.

4. Experimental Setup

Fig. 3 represents the design structure of the GMM, HMM and GMM-HMM detection system. From the figure, features are extracted from the recognised signal and the unknown signal. The feature vectors of the recognised signal are fed to K -MC block to initialise the Gaussian parameters. This is subsequently fed to the GMM and HMM blocks. The output of the GMM/HMM is then used to refine the feature vector from the unknown signal to form a refined feature vector, ϕ_r . This ϕ_r is then matched using the V-alg, which eventually recognises the sound as the humpback whale song or the unwanted signal. As shown in the figure, if the HMM block with the dotted arrow pointing towards the refined feature vector block are removed, the GMM blocks will be pointing towards the refined feature vector and sound matching blocks. Hence, the figure can be viewed as a GMM detector system. Also, if the GMM block with the dotted arrow pointing towards the refined feature vector block are removed from the figure, then the K -MC block will be pointing towards the HMM block. Thus, the figure can be viewed as an HMM detector system. Otherwise, if the dotted arrow from the GMM block pointing towards the refined feature vector block is removed, the figure can be viewed as a GMM-HMM detector system.

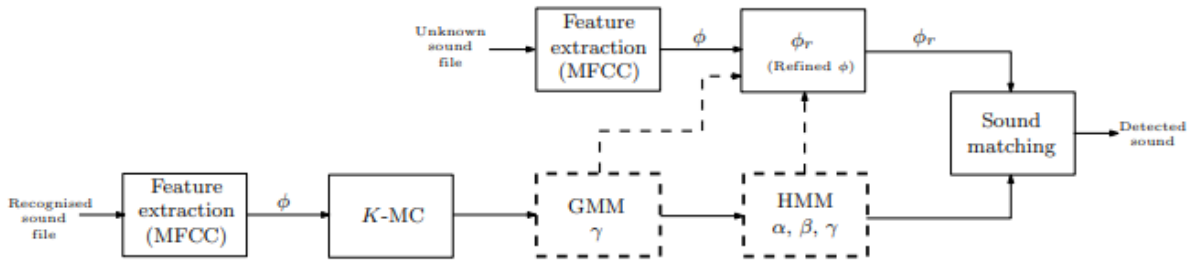


Figure 3: Humpback whale song detection system (a) Without the HMM block (GMM detector system) (b) Without the GMM block (HMM detector system) (c) With both the GMM and HMM blocks (GMM-HMM detector system).

4.1 Feature extraction: MFCC

The dataset which contains the humpback whale song and other unwanted signal is firstly pre-processed. That is, it is filtered in MATLAB, R2019a with a 3rd order Butterworth bandpass filter. This removes frequencies under 50Hz and above 5000Hz, to reduce the background and DC component disturbances. Afterwards, distinctive features are extracted from this dataset. In this paper, the Mel-scale frequency cepstral coefficients (MFCCs) (Majeed *et al.*, 2015) is used to extract features from the dataset. Other FC technique that can be used includes Linear predictive coefficients, empirical mode decomposition, and so on. However, the MFCC method is widely used for FC in many signal processing applications because it exhibits a low computational complexity (Usman *et al.*, 2020); thus, it is adopted in this paper. To extract features, the MFCC uses two kind of filters (linear spaced filters and logarithm spaced filters) to transform the signal from the time domain into the Mel-scale frequency. In summary, FC using the MFCC involves seven steps which are performed sequentially; (1) Pre-emphasis (2) Framing (3) Windowing (4) Fast Fourier transform (5) Mel-scale filter bank (6) Logarithm operation, and (7) Discrete cosine transform (Majeed *et al.*, 2015). Therefore, given a signal, the MFCC estimates the cepstral coefficient as (Ogundile *et al.*, 2020a):

$$\hat{h} = \sum_{g=1}^{\theta} \delta_g \cos \left[h(g - 0.5) \frac{\pi}{\theta} \right], \quad h = 1, 2, \dots, \theta \quad (2)$$

The logarithmic energy of the g^{th} Mel-spectrum band is denoted as δ_g and the number of cepstral coefficients is defined as θ , which usually varies between 10-14 depending on the signal processing application. In this paper, $\theta = 12$ is chosen for the detection of the Humpback whale song.

4.2 Sampling Training

The sound recordings for the Humpback whale songs dataset were visually and aurally analysed with the aid of Sonic Visualiser. The recordings are divided into two uneven sections, such that the small section (less than 30% of the whole sound file) is used to train the detectors. Subsequently, the training samples are annotated in two categories; namely, (1) Humpback whale songs, and (2) Noise (any signal that is not that of the Humpback). It should be noted that the large sound sections are also annotated to evaluate the performance of the detectors. Afterwards, a total of two GMMs or HMMs were formed, where one is used to model the humpback whale sound and the other is used to model the unwanted sound. The number of test sample, t is varied between 18 to 30 (that is $t = 18, t = 24, t = 30$) for both the Humpback song and noise. In addition, the training samples are divided into l number of sampling point, where $l = 500, l = 300$ and $l = 100$ are assumed in this paper. The dimension of the detectors is determined by the size of the feature vector; thus, the detectors are 12-dimensional. In the case of the HMM and GMM-HMM detectors, an ergodic type HMM with four states and two mixture is used. The Humpback whale sound and the noise are trained separately with the four states and two mixture HMM. The Humpback whale sound and the noise HMMs are combined after training to form eight states and four mixture HMM. The first 1-4 states can represent the humpback whale sound while the remaining 5-8 states can represent the noise with equal switching probabilities. This output of the HMM and the processed feature vector from the unknown sound signal are fed into the V-alg which predicts the sequence of state, thereby classifying the unknown signal as either the humpback whale sound or noise. Whereas, for the GMM detector, the V-alg is used to interpret the Gaussian parameters coupled with the processed feature vector from the unknown sound signal, which eventually classifies the unknown signal as either the humpback whale sound or noise.

5. Results and Discussion

5.1. Comparative Metric and Result

The performance of the ML techniques is compared based on two existing standard means of measurement in marine mammal sound processing (Yack *et al.*, 2009).

1. Sensitivity: The detection accuracy or sensitivity, which is defined in Eqn. (3) (Yack *et al.*, 2009), is the capability of the developed automated detector to correctly detect the sound.

$$S = \frac{TP}{TP + FN}. \quad (3)$$

In eqn. (3), TP (true positive) is the number of times the output of automated detector matches with the manually detected sound and FN (false negative) is the number of times the automated detector missed the manually detected sound.

2. False Discovery Rate: The false discovery rate defined by Eqn. (4) can be interpreted as a measure for the reliability of the automated detector.

$$F = \frac{FP}{TP + FP}, \quad (4)$$

where FP (false positive) is number of sounds detected incorrectly by the automated detector.

5.2 Discussion

Tables 1, 2 and 3 depict the results of the three detectors. As shown in the table, the GMM-HMM detectors exhibit superior sensitivity and false discovery rate performances in comparison to the GMM and HMM detector.

For instance, in Table 3, at $l = 100$ and $t = 30$, the GMM-HMM ML technique offered a sensitivity (S) performance gain of 10.16% and 9.93% in comparison to the GMM and HMM ML techniques, respectively. Likewise, for the same experimental conditions, the false discovery rate (F) of the GMM-HMM ML technique reduces by 14.14% and 4.9% as compared to the GMM and HMM ML techniques, respectively. Whereas the S and F performances of the HMM detector is better as compared to the GMM detector. In addition, note that the performance of the three detectors improve as the number of sampling points, l reduce from 500 to 100. For example, comparing Table 1 ($l = 500$) to Table 2 ($l = 300$), the S performance of the GMM-HMM ML technique improves by 0.96%, 2.05% and 3.72% for $t = 18$, $t = 24$, and $t = 30$ respectively. Similarly, comparing Table 2 ($l = 300$) to Table 1 ($l = 100$), the S performance of the GMM-HMM ML technique improves by 4.68%, 2.02% and 1.27% for $t = 18$, $t = 24$, and $t = 30$ respectively. This performance gain can also be noticed in a similar manner for the false discovery rate (F) as we move from $l = 500$ to $l = 100$. Lastly, observe that as the number of the test samples, t increase, the performance of all the detectors improves. Thus, it is envisaged that the obtained S and F performances can be improved if the number of test samples, t increase from 30. But this performance increase will start to saturate at a particular value of t .

Table 1: Performance comparison of three ML detection techniques for Humpback whale songs, $l = 500$

t	$S(\%)$			$F(\%)$		
	GMM	HMM	GMM-HMM	GMM	HMM	GMM-HMM
18	46.02	59.77	71.06	27.77	20.11	6.23
24	52.14	63.02	73.04	25.02	17.88	5.34
30	64.98	67.11	75.08	20.67	15.98	4.28

Table 2: Performance comparison of three ML detection techniques for Humpback whale songs, $l = 300$

t	$S(\%)$			$F(\%)$		
	GMM	HMM	GMM-HMM	GMM	HMM	GMM-HMM
18	48.98	61.88	72.02	25.34	17.82	5.01
24	54.22	66.22	75.09	23.77	14.13	3.66
30	67.14	69.99	78.80	18.69	12.56	2.49

Table 3: Performance comparison of three ML detection techniques for Humpback whale songs, $l = 100$

t	$S(\%)$			$F(\%)$		
	GMM	HMM	GMM-HMM	GMM	HMM	GMM-HMM
18	52.05	65.89	76.70	21.01	9.66	3.02
24	57.00	68.11	77.11	19.88	8.76	2.77
30	69.91	70.14	80.07	16.25	7.01	2.11

To emphasize these results, Fig. 4 and Fig. 5 depict the sensitivity and false discovery rate performances of the three ML detection techniques with respect to the number of test samples t . As explained earlier, the GMM detector offers the worst S and F performances, while the GMM-HMM detector offers superior performance in comparison to the GMM and HMM detectors. Regardless, the result has shown that ML tools/approaches are suitable for the real-time detection of cetacean vocalization.

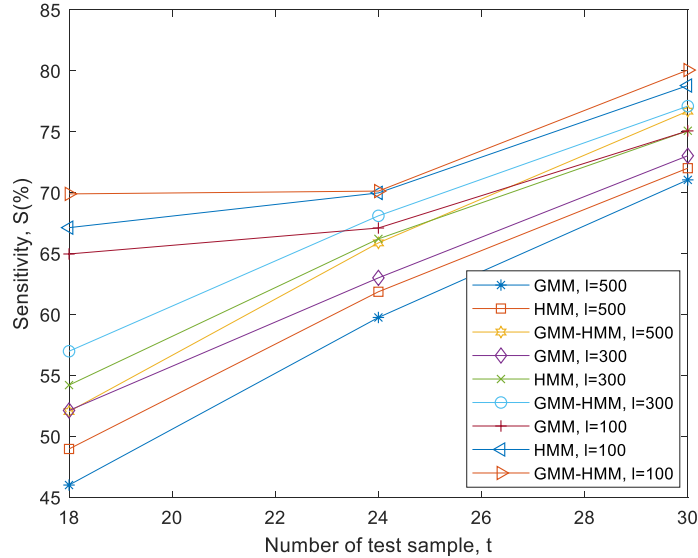


Figure 4: Sensitivity performance comparison of three ML detection techniques

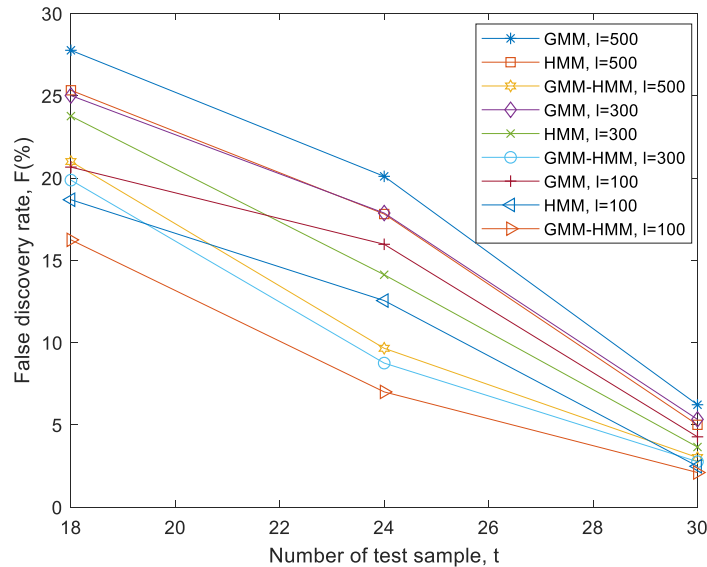


Figure 5: False discovery rate performance comparison of three ML detection techniques

6. CONCLUSION

This paper has reviewed the performance of three different ML techniques (GMM, HMM, GMM-HMM) for the detection of Humpback whale songs. As presented, the combined GMM-HMM ML detection approach offered the best detection accuracy in comparison to the GMM approach and the HMM approach. For instance, at $l =$

100 and $t = 30$, the GMM-HMM ML technique offered a sensitivity (S) performance gain of 10.16% and 9.93% in comparison to the GMM and HMM ML techniques, respectively. Likewise, for the same experimental conditions, the false discovery rate (F) of the GMM-HMM ML technique reduces by 14.14% and 4.9% as compared to the GMM and HMM ML techniques, respectively. Although this performance gain comes with increased computational time complexity, the GMM-HMM approach is suitable for real-time application. Also, this approach can be extended to other cetacean that produce characteristics vocalisation.

7. References

- Babalola, O. P., Usman, A. M., Ogundile, O. O., & Versfeld, D. J. J. (2021). Detection of Bryde's whale short pulse calls using time domain feature with hidden Markov models. *SAIEE Africa Research Journal*, 112(1), 15-23.
- Baum, L. E., Petrie, T., Soules, G., & Weiss, N. (1970). A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. *Ann. Math. Statist.*, 41, 164-171.
- Bilmes, J. A. (1998). A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. *Int. Comput. Sci. Inst.*, 4(510), 126.
- Brown, J., & J O Miller, P. (2007). Automatic classification of killer whale vocalizations using dynamic time warping. *The Journal of the Acoustical Society of America*, 122, 1201-1207.
- Forgy, E. W. (1965). Cluster analysis of multivariate data: Efficiency versus interpretability of classification. *Biometrics*, 21(3), 768-779.
- Gibb R., Browning E., Glover-Kapfer P., & Jones K. E. (2019). Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods Ecol. Evol.*, 10(2), 169-185.
- Majeed, S., Husain, H., Abdul Samad, S., & Idbeaa, T. (2015). Mel frequency cepstral coefficients (MFCC) feature extraction enhancement in the application of speech recognition: A comparison study. *Journal of Theoretical and Applied Information Technology*, 79, 38-56.
- Ogundile, O. O., Usman, A. M., Babalola, O. P., & Versfeld, D. J. J. (2021). A Dynamic mode decomposition: A feature extraction technique based hidden Markov model for detection of Mysticetes' vocalisations. *Ecological Informatics*, <https://doi.org/10.1016/j.ecoinf.2021.101306>.
- Ogundile, O. O., Usman, A. M., Babalola, O. P., & Versfeld, D. J. J. (2020). A hidden Markov model with selective time domain feature extraction to detect inshore Bryde's whale short pulse calls. *Ecological Informatics*, 57, 1-7, Art. no. 101087.
- Ogundile, O. O., Usman A. M. & Versfeld, D. J. J. (2020). An empirical mode decomposition based hidden Markov model approach for detection of Bryde's whale pulse call. *Journal of Acoustic Society of America*, 147(2), EL125-EL131.
- Putland, R., Ranjard, L., Constantine, R., and Radford, C. (2018). A hidden Markov model approach to indicate Bryde's whale acoustics. *Ecological Indicator*, 84, 479-487.
- Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77, 257-286.
- Rabiner, L. R., & Shafer, R. (1978). Digital processing of speech signals. London: Prentice-Hall International Inc.
- Smith, J. S., Hill, L. K., & Gonzalez, R. M. (2019). Whale watching and preservation of the environment in central Baja California, Mexico. *FOCUS Geography*, 62, 1.
- Usman, A. M., Ogundile, O. O., & Versfeld, D. J. J. (2020). Review of automatic detection and classification techniques for Cetacean vocalization. *IEEE Access*, 8, 105181-105206.
- Viterbi, A. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13, 260-269.
- Wuest, T., Weimer, D., Irgens, C., & Thoben, K-D (2016). Machine learning in manufacturing: advantages, challenges, and applications. *Production & Manufacturing Research*, 4(1), 23-45.
- Yack, T., Barlow, J., Rankin, S., & Gillespie, D. (2009). Testing and Validation of Automated Whistle and Click Detectors Using PamGuard 1.0 NOAA Technical Memorandum NMFS. U. S.