www.czasopisma.pan.pl www.journals.pan.pl

ARCHIVES OF ACOUSTICS
Vol. **45**, No. 4, pp. 721–731 (2020)
DOI: 10.24425/aoa.2020.135278

Research Paper

Marine Mammals Classification using Acoustic Binary Patterns

Maheen NADIR[(1)], Syed M. ADNAN[(1)], Sumair AZIZ[(2)]*, Muhammad Umar KHAN[(2)]

[(1)] *Department of Computer Sciences*
University of Engineering and Technology
Taxila, Pakistan

[(2)] *Department of Electronics Engineering*
University of Engineering and Technology
Taxila, Pakistan

*Corresponding Author e-mail: sumair.aziz@uettaxila.edu.pk

(*received July 13, 2019; accepted June 10, 2020*)

Marine mammal identification and classification for passive acoustic monitoring remain a challenging task. Mainly the interspecific and intraspecific variations in calls within species and among different individuals of single species make it more challenging. Varieties of species along with geographical diversity induce more complications towards an accurate analysis of marine mammal classification using acoustic signatures. Prior methods for classification focused on spectral features which result in increasing bias for contour base classifiers in automatic detection algorithms. In this study, acoustic marine mammal classification is performed through the fusion of 1D Local Binary Pattern (1D-LBP) and Mel Frequency Cepstral Coefficient (MFCC) based features. Multi-class Support Vector Machines (SVM) classifier is employed to identify different classes of mammal sounds. Classification of six species named *Tursiops truncatus*, *Delphinus delphis*, *Peponocephala electra*, *Grampus griseus*, *Stenella longirostris*, and *Stenella attenuate* are targeted in this research. The proposed model achieved 90.4% accuracy on 70–30% training testing and 89.6% on 5-fold cross-validation experiments.

Keywords: marine mammals; 1D Local Binary Patterns; Mel frequency cepstral coefficients; feature extraction; passive acoustic monitoring.

1. Introduction

Wildlife management and conservation require extensive and robust information about animals' behaviour. Cetologists are interested to identify the animal's population structure, anatomy, physiology, genetics, parasites, diseases, behaviour and sensory abilities, evolutionary relationships, ecology, and conservation. Cetacean (whales, dolphins, and porpoises) plays a vital role in the oceanic ecosystem by taking control of the lower trophic level population (Bowen, 1997). Dolphins usually live in beach areas and use the same seafood which humans consume and hence become a great sign of pollution and health for the ecosystem. Whales are responsible to balance ocean ecosystems through establishing proper food chain and population, i.e. blue whales consume 40 million krill per day and ensure that certain species do not overpopulate the ecosystem (Sakthivel *et al.*, 2014). Cetacean's (whale) poop helps in phytoplankton's growth which takes out the carbon from the environment and provides a healthier and clean atmosphere for all the land and sea animals making cetaceans more important (Ojala *et al.*, 1996). Now, spectating activities of dolphins introduce them as a source of economic growth for many countries (Amin, Thomas, 1996).

In dark water, acoustical detection of mines and other dangerous objects remains a difficult target to achieve. Navy and other federal organizations require dolphins (Binder, Hines, 2014) and sea lions because of their excellent sensory and diving competencies which help to protect lives and naval assets (Kaniklides, 2014) as they can detect and locate enemy swimmers and mines which threaten human lives, military and civilian ships. Adaptability, trainability and hearty nature of bottlenose dolphins and white whales make them well-known species for navy tasks (Nalavade, Meshram, 2012). Blue whales are good consumers of certain species like krill. Sperm whale helps to offset the carbon from the atmosphere.

Thus, the classification of marine mammals becomes an emerging subject for researchers. Instead of visual methods such as films and photos, audio provides better means of tracking animals from a great distance. This kind of tracking involves information about the environmental circumstances. Some parameters like the time of day and night, inaccessibility of some area, costs, time spans animals use to spend at that location definitely affect the visual observations. Acoustic observations provide confident results as sounds are not affected by the weather.

PAM (Passive Acoustic Monitoring) (Lin *et al.*, 2014) is an enhanced way of monitoring marine mammals. Acoustic identification emerges as a beneficial source because sound can transmit much faster than light. Vocal of whales can travel at a great distance which helps to identify them from tens of kilometres and can be used to classify different species/individuals based on their unique characteristics. Traditionally, long-term monitoring has employed archival instruments from which data are accessed only when the recording instrument is retrieved (Baumgartner *et al.*, 2018). Some current PAM systems implement real-time processing like hydrophone arrays where processing is done on-board or on a discrete system (a type of PAM system) which records all sounds for post-processing such as noise loggers.

Identification and classification of marine mammals using their calls remain a challenging task in PAM (Lin *et al.*, 2014). Cetacean uses calls and whistles to communicate with each other. High variation exists in whistles within species which makes their classification problematic. Along with the variation in calls within species similar calls are produced by different species which is another challenge. To meet these challenges, Lin and Chou (2015) and Lin *et al.* (2013) introduce a Local-max detector to identify frequencies of calls of marine mammals to classify them. Recordings of seven different species were included in this research achieved from MobySound.org (Mellinger, Clark, 2006). Three acoustical parameters were measured for the distribution of frequencies. Using the statistical analysis this study achieved a 70.3% correct classification rate.

González-Hernández *et al.* (2017) applied 1/6-octave analysis for feature extraction and a combination of four parallel feed-forward neural networks with a 90% classification rate. Teager-Kaiser Energy Operator based method was used for clicks detection (Luo *et al.*, 2017). Naive Bayesian, K-nearest neighbors, artificial neural networks, support vector machines and hidden Markov models were used for classification after preprocessing and feature extraction. Feroze *et al.* (2018) demonstrated that a single classification approach is not enough to classify signals with very high accuracy. With an emphasis on non-stationary noise sources, many DCL algorithms were enhanced. To detect the nearest neighbour approach along with 3D localization with multiple arrivals was established by applying time-difference-of-arrival (TDOA) methods, recalling TDOAs a few times more than normal three detections while associations among given phone's detections with the nearest neighbors were used (Mellinger *et al.*, 2017). The network-based classification method proves effective for unsupervised and rapid marine call classification using large datasets that contain clicks types that may not be recognized as *a priori* (Frasier *et al.*, 2017).

To classify blue whale calls, wavelet packet transforms and short-time Fourier are proposed and to compute and construct energies and vocalization, characteristics feature vectors are constructed (Bahoura, Simard, 2010). 86.25% classification accuracy is achieved by using multilevel perceptron on the tested dataset. Rankin *et al.* (2017) introduced BANTER – a classification method that includes data of all call types. For echolocation clicks, whistles and burst pulses, an individual classifier was created by using PAM Guard with an 84% accuracy rate for all classes. Feature of sound produced by similar groups was extracted by using four types of individual features and three classifiers were utilized to classify introduced species. Sparse classifier and Mel Frequency Cepstral coefficients (MFCCs) identify species with an accuracy rate of 82.7% (Ibrahim, *et al.* 2018).

PAM often shows inaccurate results which overburdened acoustic analyst. To overcome this problem automatic recognition methods were proposed by (Binder, Hines, 2014). The whole process is divided into two steps. Firstly, automatic detection (Bougher *et al.*, 2012) is applied to a dataset of four species of cetacean bowhead (Mellinger, Clark, 2000), humpback (Payne, McVay, 1971), North Atlantic right (Bort *et al.*, 2015) and sperm whales (Thode *et al.*, 2002). Secondly, automatic classifier is used to accurately distinguish between species which shows 85% accuracy but aural classifier was applied on a limited dataset (Young, Hines, 2007).

High intra-specific variation makes species classification a challenging task (Lin *et al.* 2014). To observe seasonal changes in species diversity automatic classification and detection techniques (Guisan *et al.*, 2002, Seavy *et al.*, 2005) were applied on acoustic features of marine by using marine cable hosted Observatory(MACHO) (Hsu *et al.*, 2007) with twelve features vector and to classify four cetacean species discriminant function analysis was used which results in 72.2% classification rate (Ramayah *et al.*, 2010). Local Binary Patterns is widely used approach for 2D image processing such as facial recognition and texture identification. This research aims to develop a 1D Local Binary Patterns for feature extraction directly from acoustic features rather than converting acoustic signals into images/histograms.

2. Materials and methods

2.1. Data collection

WMMS-DB (The Watkins Marine Mammals Sound Database, Woods Hole Oceanographic Institution) is a standard dataset, explicitly designed for personal and academic use. The dataset contains about 2000 recordings of more than 60 species approximately. AX-58, Tyack Suction cup, Ithaco 602M108 hydrophones were used to detect the sound waves underwater along with Magnemite and Magnecorder tape recorders with the frequency response of 70–8000 Hz. Table 1 describes the dataset used in the study.

Table 1. Summary of sound recordings examined in this study.

S#	Species	No. of files
1	Bottlenose dolphins	24
2	Common dolphins	52
3	Melon headed whales	63
4	Risso's dolphins	67
5	Spinner dolphins	114
6	Spotted dolphins	66

2.2. Bottlenose dolphins

Bottlenose dolphins got the name because of their short stubby beak which looks like a bottle. Scientifically they are known as *Tursiops Truncatus*. They can usually weigh up to 1000 pounds and can grow up to 12 feet long. Bottlenose dolphins are frequently met in coastal and oceanic water, generally in moderate water among 45°N and 45°S (Baumann-Pickering *et al.*, 2010). Three different types of vocal learning are present in bottlenose dolphins like comprehension, production, and usage. These learning capabilities are utilized to recognize the unique signature clicks of individuals (Janik, 2013).

2.3. Common dolphins

Based on genetic and morphological differences, common species are categorized into two species "short-beaked common dolphin" (*Delphinus delphis*) and "long-beaked common dolphins" (*Delphinus capensis*). Delphinus dolphins mostly live in deeper water up to 590 feet (López *et al.*, 2013). Short beak common dolphins use four types of clicks and whistles to communicate, e.g. pure tones, visual cues, tactile cues, and non-vocal acoustic cues. These dolphins mostly use pure clicks to express happiness, excitement and even panic. Their clicks/whistles have basics from 3 to 24 kHz and last 0.01–4 s (Erbe *et al.*, 2017).

2.4. Melon-headed whales

Melon-headed whales (*Peponocephalaelectra*) are found worldwide in oceanic water from 20°N to 20°S and frequently in Palmyra Atoll (northern line island). They usually grow up to 2.78 m long. Female whales are often smaller than males. Melon-headed whales use acoustic signals like clicks and whistles to prey and communicate among groups. They have a dominant frequency range of whistles between 8 kHz to 29.7 kHz (Frankel, Yin, 2010). Frequency of their calls depends on their motions, they usually produce low-frequency calls among 8–12 kHz while they rest or swim but when they feel excited or frightened, their clicks can reach up to 20–40 kHz.

2.5. Risso's dolphins

Risso's dolphins (*Grampusgriseus*) were named after French naturalist Antoine Risso. They are also known as "a big fish" because they are mostly 4 m long and weigh more than 500 kg. Risso's dolphins live where the continental slope is near to shore and they are widespread in the Mediterranean Sea (Bearzi *et al.*, 2011). Risso has a life span of 20–30 years where carnal development starts at the age of 8–10 and 10–12 years for females and males respectively. At a trophic level, reduction in Risso's dolphins population as prey items will affect the energy budgets of consumer species.

2.6. Spotted dolphins

Atlantic spotted dolphins are often misidentified species because as they grow they develop special spots on their bodies, unlike young spotted dolphins. Male dolphins are much longer than females as they mature with a size of 7½ feet and with a weight of 240 to 360 kilograms. A significant number of spotted dolphins are in the Bahamas ocean but they also live in different locations, e.g. Africa, Europe, the United States and the Gulf of Mexico. This species has a broad frequency range of whistles and clicks from 1.15 to 23.44 kHz.

2.7. Spinner dolphins

Spinner dolphins (*Stenellalongirostris*) are well known "acrobats" of the ocean. Spinners named due to their ability to spin their bodies multiple times. They are approximately 6½ feet long with dark gray backs and white bellies. They mostly live in warm water 30 to 40°C around the world, e.g. Thailand, Hawaiian Island and Pacific Ocean of America (Thorne *et al.*, 2012). Spinner dolphins usually travel in groups called "schools". Their schools contain other species such as humpback whales, tunas and other species of dolphins. They have a special hearing sense which enables them to determine the size, movement, and position of a particular object with the utilization of sounds (Benoit-Bird, Au, 2009).

The features of the species mentioned above are shown in Table 2.

Table 2. Summary of sound recordings examined in this study.

Species	Scientific name	Frequency range [kHz]	Weight/length
Bottlenose dolphins	Tursiops Truncatus	60–140	1000 pounds/12 feet long
Common dolphins	*Delphinus delphis/Delphinus capensis*	3–24	440 pounds/9 feet long
Melon-headed Whales	*Peponocephalaelectra*	8–29.7	2.78 m long
Risso's dolphins	*Grampusgriseus*	4–22	500 kg/4 m long
Spotted dolphins	*Stenella frontalis*	1.15–23.44	240–360 kg/7½ feet long
Spinner dolphins	*Stenellalongirostris*	32.3–65	6½ feet long

3. Feature extraction and classification approach

3.1. Mel Frequency Cepstral Coefficient (MFCC)

To extract spectral features MFCC becomes the most dominant feature extraction method which extracts audio features by extracting parameters of the speech while de-emphasizing all other data (Chaudhari *et al.*, 2015; Aziz *et al.*, 2019a). As an acoustic descriptor MFCCs have been widely used in the research using audio signals (Valero, Alías, 2012; Bhalke *et al.*, 2016; Reljin, Pokrajac, 2017; Aziz *et al.*, 2019d). The base of MFCC is the frequency domain while using the Mel scale (Das, Parekh, 2012). Mel-scale can be defined as a scale on which each tone of a sound signal with frequency t is measured in hertz (Dash *et al.*, 2012).

After the conversion of an analog signal into a digital signal, speech samples go through the procedure of *frame blocking* in which they are segmented into small frames with the maximum length of 40 ms. Voice signals are divided into N-sample's frames. Neighbouring samples are isolated by some value M, where $M > N$ and mostly M has value 100 when N is 256 (Aida-Zade *et al.*, 2006). Hamming window is used for windowing steps in the feature extraction process that assimilates those frequency lines which are the closest. Now time domain to frequency domain conversion is performed for frames of N samples in Fast Fourier Transform (FFT). The convolution of vocal tract impulse response $H[n]$ and a pulse $U[n]$ is converted into the time domain using Fourier transforms. The linear scale is not followed by voice signal because of this; filters are applied according to Mel scale in the fourth step Mel-frequency wrapping of this chain. For the centre frequency of two contiguous filters linearly is declined to zero by using each filter (Tiwari, 2010). For the given frequency f Eq. (1) is applied to extract the Mel frequency

$$f_{\text{Mel}} = 25951 \log_{10}\left(1 + \frac{f}{700}\right). \tag{1}$$

To extract the Mel Frequency Cepstral Coefficient, the log Mel spectrum is converted into the time domain using a Discrete Cosine Transform (DCT) in the last step of MFCC (Dash *et al.*, 2012). The resultant set of Mel Frequency Cepstral Coefficient is called acoustic vectors. In this study 13 MFCCs were used to form an acoustic vector of a certain sample.

3.2. Local Binary Pattern (LBP)

LBP is a widely used gray-scale texture operator in many computer vision applications because of its computation simplicity (Morales *et al.*, 2017). The LBP operator works on a local approach that defines the relationship between the centre pixel and its neighbour pixels. Generally, LBP operator compares the gray values of 8 neighborhood pixels with respect to their centre pixel with the rule that each neighbour pixel will assign the value 1 if they have greater or equal value than a centre pixel, otherwise they will be given value 0 (Tang *et al.*, 2015). To form a binary number, each neighbour is assigned a binary number acquired clockwise.

LBP is commonly used for feature extraction from images. For one-dimensional signal processing, an adaption from LBP was introduced that produced histograms from data generated from 1D-LBP codes (Chatlani, Soraghan, 2010; Irtaza *et al.*, 2017). 1D-LBP is computed by comparing the neighbouring samples with $X[n]$ (McCool *et al.*, 2012)

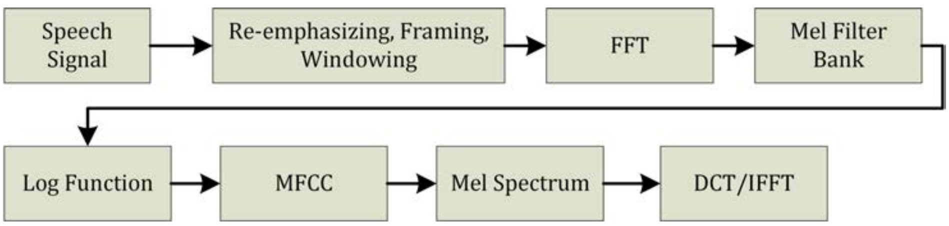

Fig. 1. MFCC Derivation.

$$\mathrm{LBP}_P(x[n]) = \sum_{n=0}^{\frac{P}{2}-1} \left\{ S\left[x\left[n+r-\frac{P}{2}\right]-x[n]\right]2^r + S[x[n+r+1]-x[n]]2^{r+\frac{P}{2}} \right\}, \tag{2}$$

where P is the value of the current neighbour, in this case it is 8, r is the radius and $S[\cdot]$ is the sign function defined by

$$S[x] = \begin{cases} 1 \text{ if } & x \geq 0, \\ 0 \text{ otherwise.} \end{cases} \tag{3}$$

Possible local binary patterns extracted from Eq. (2), there is 2^P distribution of LBP codes within the windowed portion of a signal is termed as LBP histograms that can be defined as Eq. (4)

$$H_b = \sum_{\frac{P}{2} \leq n \leq N-\frac{P}{2}} \delta(\mathrm{LBP}_P(x[n]), b), \tag{4}$$

where N denotes the length of the window or complete signal and the number of histogram bins is shown by B where $b = 1, ..., B$. Histograms can be uniform that has bins with two transitions as 0 to 1 and vice versa while non-uniform histograms which are placed in the same bin (McCool *et al.*, 2012). In this research, 20 LBP values were used to form a texture feature vector of a certain sample.

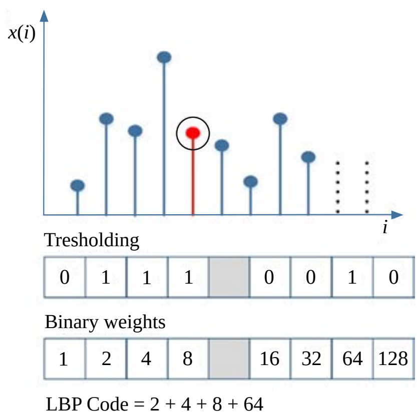

Fig. 2. Local Binary Patterns.

3.3. Classification – Support Vector Machines (SVM)

Statistical learning is the base theory of SVM which increase the performance of various fields because of its high accuracy such as pattern recognition (Aziz *et al.*, 2019b; 2019c), financial forecasting, regression estimation, text categorization, medical diagnosis (Khan *et al.*, 2019a; 2019b), face detection, marketing estimation, and handwritten digit recognition etc. (Shin *et al.*, 2005). SVM has been recognized as a base classifier in different research problems related to classification as the final result sets (Bhalke *et al.*, 2017; Qian *et al.*, 2018). SVM works on the principle of risk minimization while minimizing upper bound on expected risk using structural risk minimization (SRM) (Sugumaran *et al.*, 2007). In two-class pattern recognition, problem binary classifier is produced by SVM that linearly divides the classes and selects a decision boundary which minimizes the generalization error to a great extent (Pal, Mather, 2005). To separate the classes linearly SVM creates optimal hyperplanes that isolate the data among hyperplanes with maximum distance and closest training points with hyperplanes are known as "support vectors". If not the case, SVM tries to trade-off among classification error controlled by user constant and margin by locating hyperplanes which maximize margin while minimizing the classification error. SVM can also handle nonlinear decision problems (Boser *et al.*, 1992) with technique to project the input data onto a high dimensional features space using kernel function and formulating a linear classification problem in that feature space. In linear separable problems of support vector machine, optimal hyperplanes are estimated. But in the case of multiple classes, problems cannot be classified in linear manners, hence fail to find optimal hyperplanes.

In such cases well define kernel function $k(x, y)$ is used that changes the inner product (x, y) of the input vector. The defined kernel function projects the inner product into a new high dimensional space where the optimal hyperplane can be defined, and data can be separable linearly.

Kernel functions can be defined by Eq. (5)

$$K(x, y) = \phi(x)^{\mathrm{T}} \phi(y)^{\mathrm{T}}. \tag{5}$$

Possible kernel functions in Support Vector Machine for non-linear classification problems are elaborated in Table 3. The strategy used in this paper compares one class with all other classes taken as one class.

Table 3. Kernel Support Vector Machine Functions.

Kernel	Function	Parameters
Gaussian Radial Basis	$K(x, y) = \exp\left(-\frac{\lVert x-y \rVert^2}{2\sigma^2}\right)$	σ – width of Gaussian function
Polynomial	$K(x, y) = (x.y + 1)^d$	d – degree of polynomial
Multi-layer Perception	$K(x, y) = \tanh(k(x.y) - \mu)$	k – scale, μ – offset
Linear Kernel	$K(x, y) = x^{\mathrm{T}} \cdot y$	
Sigmoid Kernel	$K(x, y) = \tanh(\gamma x^{\mathrm{T}} y + r)$	r – Kernal parameter

For n classes, n classifiers are generated using this method. The highest margin class becomes the final output.

4. Results

Recordings of 386 sound files of the Cetacean dataset described in Table 1 are used to analyze and present the results produced by the classifier. Five-fold cross-validation and 70–30% training/testing validation are applied to isolate training and testing data. The classifier learns on training data and features are extracted using Acoustic LBP and MFCC.

Table 4 describes the classification accuracy in terms of percentage using MFCC and Acoustic LBP along with 70–30% training/testing validation for the Cetacean dataset. Our proposed method successfully recognizes three classes of Bottlenose dolphins, Risso's dolphins and Spinner dolphins with 100% accuracy rate while for Spotted dolphins it shows 65% accuracy, hence 90.4% classification accuracy is achieved by applying proposed method using acoustic features of the Cetacean dataset which is much greater than shown in literature cited.

The classifier achieved 89.6% accuracy when 5-fold cross-validation was applied on the given dataset. Classification accuracy in terms of percentage for each class is given in Table 5. This elaborates that the proposed method attains 100% and 74% maximum and minimum accuracy for Bottlenose dolphin and Spotted dolphins classes, respectively.

Figure 3 provides details of the above-cited results using 70–30% training testing data and 5-fold cross-validation on a given dataset which briefly explains that the accuracy rate of 70–30% training/testing data is higher than 5-fold cross-validation. By using the true positive, true negative, false positive and false negative values that occurred during computation, performance parameters, such as precision, recall, F-1 score, accuracy and error rate, were calculated for a given dataset. Results of classification performance for 5-fold cross-validation and 70–30% train test evaluation are given in Tables 6 and 7. Figures 4 and 5 present the classification performance for 5-fold cross-validation and 70–30% train test evaluation in terms of accuracy, precision, recall, error rate, and F1 score. The output quality of a classifier is evaluated through the Precision-Recall metric. Precision in other terms Positive Predictive Values (PPV) can be defined as a fraction of

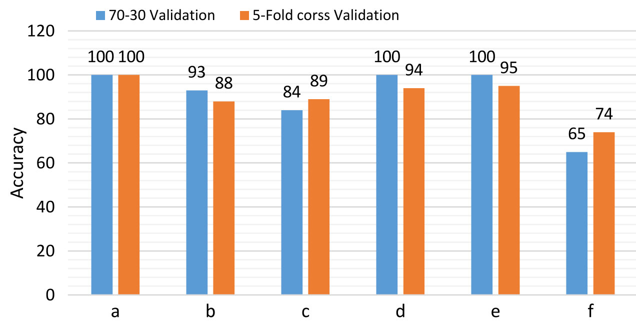

Fig. 3. Class-wise accuracy comparison for 5-fold cross validation and 70–30% train test experiments.

Table 4. Confusion matrix with 70–30% training-testing.

Species	True class	Predicted class					
		a	b	c	d	e	f
Bottlenose dolphins	a	100	–	–	–	–	–
Common dolphins	b	–	93	–	–	7	–
Melon Headed Whales	c	–	–	84	16	–	–
Risso's dolphins	d	–	–	–	100	–	–
Spotted dolphins	e	–	–	–	–	100	–
Spinner dolphins	f	–	–	–	–	35	65

Table 5. Confusion Matrix with 5-fold cross validation.

Species	True class	Predicted class					
		a	b	c	d	e	f
Bottlenose dolphins	a	100	–	–	–	–	–
Common dolphins	b	–	88	6	–	4	2
Melon Headed Whales	c	–	–	89	5	3	3
Risso's dolphins	d	–	–	4	94	1	–
Spotted dolphins	e	–	–	1	–	95	4
Spinner dolphins	f	–	2	–	5	20	74

Table 6. Classification performance with 70–30% training-testing.

Species	a	b	c	d	e	f
Precision	1	0.93	0.84	1	1	0.65
Recall	1	1	1	0.86	0.80	1
F-1 score	1	0.96	0.91	0.92	0.88	0.78
Accuracy	1	0.98	0.97	0.97	0.93	0.94
Error rate	0	0.01	0.02	0.02	0.06	0.05

Table 7. Classification performance with 5-fold cross validation.

Accuracy/Species	a	b	c	d	e	f
Precision	1	0.88	0.89	0.94	0.95	0.73
Recall	1	0.97	0.89	0.90	0.77	0.89
F-1 score	1	0.91	0.88	0.91	0.84	0.78
Accuracy	1	0.97	0.97	0.98	0.95	0.94
Error rate	0	0.02	0.03	0.02	0.05	0.06

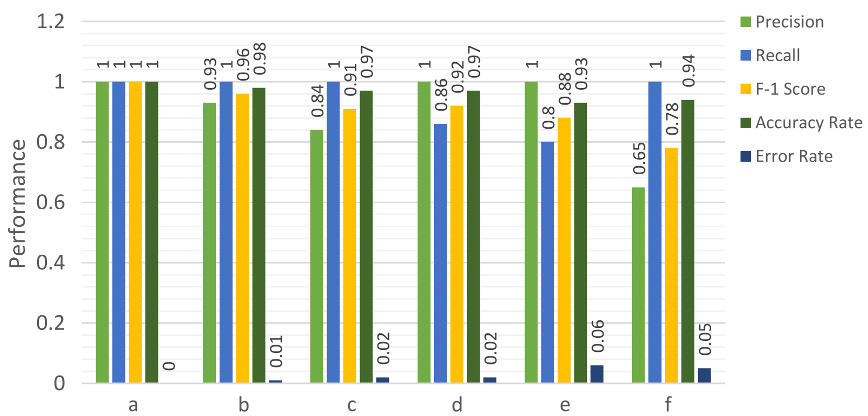

Fig. 4. Performance evaluation with 70-30% training-testing.

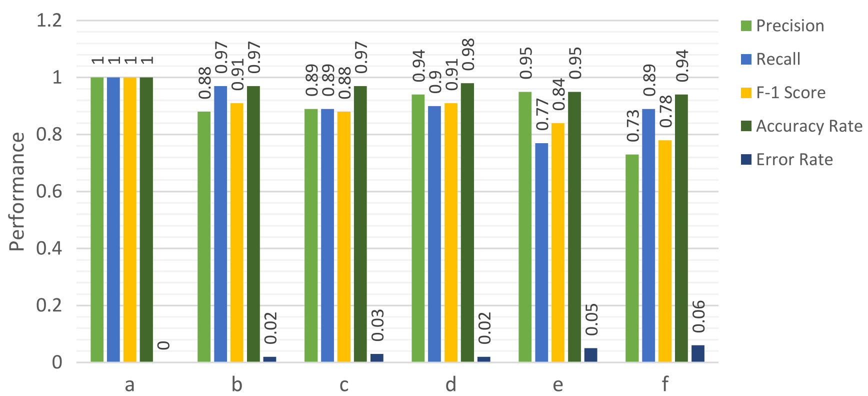

Fig. 5. Performance evaluation with 5-fold cross validation.

relevant elements between extracted values, whereas recall measures the number of true relevant results returned in any information retrieval system, in other words, defines the sensitivity of the system. Therefore, measurement of relevance is the base key of Precision and recall which in terms defines the output quality of the classifier.

5. Discussions

Cetaceans play a vital role in the oceanic ecosystem. Nowadays marine mammals like dolphins and sea lions are widely trained in navy and other federal organizations because of their excellent sensory and diving competencies which help to protect lives

Table 8. Comparison with some state of the art methods.

References	Dataset	Methods	Limitation	Results [%]
Lin, Chou (2015), Lin *et al.* (2013)	MobySound (7 classes)	Local-max detector	Frequency-based features	70.3
Ibrahim *et al.* (2018)		Sparse classifier and MFCC	Intraspecies classification	82.7
Binder, Hines (2014)	4 cetacean species	Aural classifier	Limited dataset	85
Ramayah *et al.* (2010)	4 cetacean species	Discriminant function analysis	Limited dataset	72.2

and naval assets. As the role of marine mammals has been increasing in humans' lives the accurate identification of marine mammals becomes the need of the day. Marine mammals use calls and whistles to communicate with each other. High variation exists in whistles within species which makes their classification difficult. Along with the variation in calls within species similar calls are produced by different species which poses another challenge. For the identification and classification of marine mammals using acoustic features becomes a challenging task. Thus, different classification techniques were developed to classify marine mammals using their acoustic features. To accurately identify and classify marine mammals, we used the WMMS dataset. In this research, we constructed a feature vector by fusion of 1D LBP and MFCC features for marine mammal classification. One dimensional local binary pattern (1D-LBPs) aims to capture the distribution of audio structure and is complementary to conventional Mel-scaled cepstral coefficients. A combination of MFCC and 1D-LBP adds weight to the benefit of capturing complex acoustic structure. Thus, a unique feature set is extracted using two techniques MFCC and 1D-LBP. The proposed system successfully achieved higher recognition rate with linear SVM (85.2%, 83.9%, 83.2%), cubic SVM (90.4%, 88.9%, 90.7%) and Quadratic SVM (87.8%, 89.6%, 88.8%) for given dataset with 70-30% training testing and 5-fold cross-validation techniques which is higher than found in literature. This high recognition accuracy indicates the correctness and compactness of the unique features set we extracted using techniques like MFCC and 1D LBP. The main purpose of this study is to address the issue of accurately identifying and classifying marine mammals based on their acoustic features. Hence, the proposed classifier achieves higher accuracy than reported in the literature. An accuracy of 90.4% is achieved when the feature set is classified through SVM which is higher than the existing results found in literature work.

The work in (Lin, Chou, 2015; Lin *et al.*, 2013) presents a frequency-based feature extraction method for cetacean classification. A local maximum detector was applied to seven classes and achieved a 70.3% recognition rate. Features were extracted based on the frequency of whistles generated by marine mammals which limits to differentiate between inter species. Discriminant function analysis and an aural classifier were applied to classify four classes of cetacean species by Ramayah *et al.* (2010) and Binder and Hines (2014) respectively which gain 72.2 % and 85% accuracy rate on a limited dataset. For intraspecies classification, Ibrahim *et al.* (2018) classified grouper vocalization from ambient sounds recorded by fixed hydrophones by applying weighted MFCC and sparse classifier which achieved an 82.7% accuracy rate. To overcome all the limitations of existing approaches this study aims to classify six classes of cetaceans based on their acoustic features by applying local binary patterns and MFCC to extract features and support vector machine to classify among classes which successfully achieved a 90.7% accuracy rate which is better than existing approaches.

6. Conclusion

In this article, feature fusion based methodology for accurate classification of marine mammals using acoustic signatures is presented. The proposed method uses the fusion of one-dimensional texture patterns and MFCC features with multiclass SVM for classification. Experimental results validate the robustness of proposed features for acoustic monitoring and classification of marine mammals. The proposed method achieved satisfactory classification performance on a small dataset containing few acoustic samples for some classes. The research aims to develop a 1D Local Binary Patterns for feature extraction directly from acoustic features rather than converting acoustic signals into images/histograms. As possible improvements, the combination of this classifier with versions of local patterns will be explored in the future works also an automatic system that permits to efficiently identify and classify between multiple numbers of classes will be evaluated in the classification system.

Acknowledgments

We are thankful for Woods Hole Oceanographic Institution for providing a sound recording of marine mammals for this research article.

References

1. Aida-Zade K., Ardil C., Rustamov S. (2006), Investigation of combined use of MFCC and LPC features in speech recognition systems, *World Academy of Science, Engineering and Technology*, **19**: 74–80.
2. Amin A., Thomas D. (1996), The negotiated economy: state and civic institutions in Denmark, *Economy and Society*, **25**: 255–281.
3. Aziz S., Awais M., Akram T., Khan U., Alhussein M., Aurangzeb K. (2019a), Automatic scene recognition through acoustic classification for behavioral robotics, *Electronics*, **8**(5): 483; doi: 10.3390/electronics8050483.
4. Aziz S., Khan M.U., Aamir F., Javid M.A. (2019b), Electromyography (EMG), data-driven load classification using empirical mode decomposition and feature analysis, [in:] *2019 International Conference on Frontiers of Information Technology (FIT)*, pp. 272–2725, doi: 10.1109/FIT47737.2019.00058.
5. Aziz S., Khan M.U., Choudhry Z.A., Aymin A., Usman A. (2019c), ECG-based Biometric authentication using empirical mode decomposition and support vector machines, [in:] *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pp. 0906–0912, doi: 10.1109/IEMCON.2019.8936174.
6. Aziz S., Khan M.U., Shakeel M., Mushtaq Z., Khan A.Z. (2019d), An automated system towards diagnosis of pneumonia using pulmonary auscultations, [in:] *2019 13th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS)*, pp. 1–7, doi: 10.1109/MACS48846.2019.9024789.
7. Bahoura M., Simard Y. (2010), Blue whale calls classification using short-time Fourier and wavelet packet transforms and artificial neural network, *Digital Signal Processing*, **20**(4): 1256–1263, doi: 10.1016/j.dsp.2009.10.024.
8. Baumann-Pickering S., Wiggins S.M., Hildebrand J.A., Roch M.A., Schnitzler H.-U. (2010), Discriminating features of echolocation clicks of melon-headed whales (Peponocephala electra), bottlenose dolphins (Tursiops truncatus), and Gray's spinner dolphins (Stenella longirostris longirostris), *The Journal of the Acoustical Society of America*, **128**(4): 2212–2224, doi: 10.1121/1.3479549.
9. Baumgartner M.F., Stafford K.M., Latha G. (2018), Near real-time underwater passive acoustic monitoring of natural and anthropogenic sounds, [in:] Venkatesan R., Tandon A., D'Asaro E., Atmanand M. [Eds], *Observing the Oceans in Real Time. Springer Oceanography*, pp. 203–226, Springer, Cham, doi: 10.1007/978-3-319-66493-4_10.
10. Bearzi G., Reeves R.R., Remonato E., Pierantonio N., Airoldi S. (2011), Risso's dolphin Grampus griseus in the Mediterranean Sea, *Mammalian Biology – Zeitschrift für Säugetierkunde*, **76**(4): 385–400, doi: 10.1016/j.mambio.2010.06.003.
11. Benoit-Bird K.J., Au W.W. (2009), Phonation behavior of cooperatively foraging spinner dolphins, *The Journal of the Acoustical Society of America*, **125**(1): 539–546, doi: 10.1121/1.2967477.
12. Bhalke D.G., Rajesh B., Bormane D.S. (2017), Automatic genre classification using fractional fourier transform based mel frequency cepstral coefficient and timbral features, *Archives of Acoustics*, **42**(2): 213–222, doi: 10.1515/aoa-2017-0024.
13. Bhalke D.G., Rao C.R., Bormane D. (2016), Hybridisation of mel frequency cepstral coefficient and higher order spectral features for musical instruments classification, *Archives of Acoustics*, **41**(3): 427–436, doi: 10.1515/aoa-2016-0042.
14. Binder C.M., Hines P.C. (2014), Automated aural classification used for inter-species discrimination of cetaceans, *The Journal of the Acoustical Society of America*, **135**(4): 2113–2125, doi: 10.1121/1.4868378.
15. Bort J., Van Parijs S.M., Stevick P.T., Summers E., Todd S. (2015), North Atlantic right whale Eubalaena glacialis vocalization patterns in the central Gulf of Maine from October 2009 through October 2010, *Endangered Species Research*, **26**(3): 271–280, doi: 10.3354/esr00650.
16. Boser B.E., Guyon I.M., Vapnik V.N. (1992), A training algorithm for optimal margin classifiers, [in:] *Proceedings of the Fifth Annual Workshop On Computational Learning Theory*, pp. 144–152, ACM, doi: 10.1145/130385.130401.
17. Bougher B.B., Hood J., Theriault J., Moors H. (2012), Generalized marine mammal detection based on improved band-limited processing, *Proceedings of Meetings on Acoustics ECUA2012*, **17**(1): 070067, doi: 10.1121/1.4773596.
18. Bowen W. (1997), Role of marine mammals in aquatic ecosystems, *Marine Ecology Progress Series*, **158**: 267–274.
19. Chatlani N., Soraghan J.J. (2010), Local binary patterns for 1-D signal processing, [in:] *18th European Signal Processing Conference*, pp. 95–99.
20. Chaudhari R.H., Waghmare K., Gawali B.W. (2015), Accent recognition using MFCC and LPC with acoustic features, *International Journal of Innovative Research in Computer and Communication Engineering*, **3**(3): 2128–2134, doi: 10.15680/ijircce.2015.0303078.
21. Das B.P., Parekh R. (2012), Recognition of isolated words using features based on LPC, MFCC, ZCR and STE, with neural network classifiers, *International Journal of Modern Engineering Research*, **2**(3): 854–858.
22. Dash K., Padhi D., Panda B., Mohanty S. (2012), Speaker identification using Mel frequency cepstral coefficient and BPNN, *International Journal of Advanced Research in Computer Science and Software Engineering Research Paper*, **2**(4): 326–332.
23. Erbe C. *et al.* (2017), Review of underwater and in-air sounds emitted by Australian and Antarctic marine mammals, *Acoustics Australia*, **45**(2): 179–241, doi: 10.1007/s40857-017-0101-z.

24. Feroze K., Sultan S., Shahid S., Mahmood F. (2018), Classification of underwater acoustic signals using multi-classifiers, [in:] *15th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, pp. 723–728, doi: 10.1109/IBCAST.2018.8312302.

25. Frankel A.S., Yin S. (2010), A description of sounds recorded from melon-headed whales (Peponocephala electra), off Hawai'i, *The Journal of the Acoustical Society of America*, **127**(5): 3248–3255, doi: 10.1121/1.3365259.

26. Frasier K.E., Roch M.A., Soldevilla M.S., Wiggins S.M., Garrison L.P., Hildebrand J.A. (2017), Automated classification of dolphin echolocation click types from the Gulf of Mexico, *PLoS Computational Biology*, **13**(12): e1005823, doi: 10.1371/journal.pcbi.1005823.

27. González-Hernández F.R., Sánchez-Fernández L.P., Suárez-Guerra S., Sánchez-Pérez L.A. (2017), Marine mammal sound classification based on a parallel recognition model and octave analysis, *Applied Acoustics*, **119**: 17–28, doi: 10.1016/j.apacoust.2016.11.016.

28. Guisan A., Edwards T.C., Hastie T. (2002), Generalized linear and generalized additive models in studies of species distributions: setting the scene, *Ecological modelling*, **157**(2–3): 89–100, doi: 10.1016/S0304-3800(02)00204-1.

29. Hsu S.-K. *et al.* (2007), Marine Cable Hosted Observatory (MACHO) project in Taiwan, [in:] *2007 Symposium on Underwater Technology and Workshop on Scientific Use of Submarine Cables and Related Technologies*, pp. 305–307, doi: 10.1109/UT.2007.370808.

30. Ibrahim A.K. *et al.* (2018), An approach for automatic classification of grouper vocalizations with passive acoustic monitoring, *The Journal of the Acoustical Society of America*, **143**(2): 666–676, doi: 10.1121/1.5022281.

31. Irtaza A., Adnan S.M., Aziz S., Javed A., Ullah M.O., Mahmood M.T. (2017), A framework for fall detection of elderly people by analyzing environmental sounds through acoustic local ternary patterns, [in:] *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 1558–1563, doi: 10.1109/SMC.2017.8122836.

32. Janik V.M. (2013), Cognitive skills in bottlenose dolphin communication, *Trends in Cognitive Sciences*, **17**(4): 157–159, doi: 10.1016/j.tics.2013.02.005.

33. Kaniklides S. (2014), *Effects of Volcanic Tsunamis on Marine Mammals*, Thesis for: PhD, doi: 10.13140/RG.2.1.4696.1687.

34. Khan M.U., Aziz S., Amjad F., Mohsin M. (2019a), Detection of dilated cardiomyopathy using pulse plethysmographic signal analysis, [in:] *2019 22nd International Multitopic Conference (INMIC)*, pp. 1–5, doi: 10.1109/INMIC48123.2019.9022734.

35. Khan M.U., Aziz S., Ibraheem S., Butt A., Shahid H. (2019b), Characterization of term and preterm deliveries using electrohysterograms signatures, [in:] *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pp. 0899–0905, doi: 10.1109/IEMCON.2019.8936292.

36. Lin T.-H., Chou L.-S. (2015), Automatic classification of delphinids based on the representative frequencies of whistles, *The Journal of the Acoustical Society of America*, **138**(2): 1003–1011, doi: 10.1121/1.4927695.

37. Lin T.-H., Chou L.-S., Akamatsu T., Chan H.-C., Chen C.-F. (2013), An automatic detection algorithm for extracting the representative frequency of cetacean tonal sounds, *The Journal of the Acoustical Society of America*, **134**(3): 2477–2485, doi: 10.1121/1.4816572.

38. Lin T.-H., Yu H.-Y., Chou L.-S., Chen C.-F. (2014), Passive acoustic monitoring on the seasonal species composition of cetaceans from the marine cable hosted observatory, [in:] *OCEANS 2014-TAIPEI*, pp. 1–6, doi: 10.1109/OCEANS-TAIPEI.2014.6964392.

39. López B.D., Addis A., Fabiano F. (2013), Ecology of common bottlenose dolphins along the Northwestern Sardinian coastal waters (Italy), *Thalassas: An International Journal of Marine Sciences*, **29**(2): 35–44.

40. Luo W., Yang W., Song Z., Zhang Y. (2017), Automatic species recognition using echolocation clicks from odontocetes, [in:] *2017 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, pp. 1–5, doi: 10.1109/ICSPCC.2017.8242503.

41. McCool P., Chatlani N., Petropoulakis L., Soraghan J.J., Menon R., Lakany H. (2012), 1-D local binary patterns for onset detection of myoelectric signals, [In:] *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, pp. 499–503.

42. Mellinger D.K., Clark C.W. (2000), Recognizing transient low-frequency whale sounds by spectrogram correlation, *The Journal of the Acoustical Society of America*, **107**(6): 3518–3529, doi: 10.1121/1.429434.

43. Mellinger D.K., Clark C.W. (2006), MobySound: A reference archive for studying automatic recognition of marine mammal sounds, *Applied Acoustics*, **67**(11–12): 1226–1242, doi: 10.1016/j.apacoust.2006.06.002.

44. Mellinger D.K. *et al.* (2017), Advanced methods for passive acoustic detection, classification, and localization of marine mammals, *The Journal of the Acoustical Society of America*, **141**(5): 3604–3604, doi: 10.1121/1.4987711.

45. Morales S., Engan K., Naranjo V., Colomer A. (2017), Retinal disease screening through local binary patterns, *IEEE Journal of Biomedical and Health Informatics*, **21**(1): 184–192, doi: 10.1109/JBHI.2015.2490798.

46. Nalavade K., Meshram B. (2012), Data classification using support vector machine, [in:] *National Conference on Emerging Trends in Engineering & Technology (VNCET)*, pp. 181–184.

47. Ojala T., Pietikäinen M., Harwood D. (1996), A comparative study of texture measures with classification based on featured distributions, *Pattern Recognition*, **29**(1): 51–59, doi: 10.1016/0031-3203(95)00067-4.

48. Pal M., Mather P. (2005), Support vector machines for classification in remote sensing, *International Journal of Remote Sensing*, **26**(5): 1007–1011, doi: 10.1080/01431160512331314083.

49. Payne R.S., McVay S. (1971), Songs of humpback whales, *Science*, **173**(3997): 585–597, doi: 10.1126/science.173.3997.585.

50. Qian K. *et al.* (2018), Teaching machines on snoring: A benchmark on computer audition for snore sound excitation localisation, *Archives of Acoustics*, **43**(3): 465–475, doi: 10.24425/123918.

51. Ramayah T., Ahmad N.H., Halim H.A., May-Chiun S.R.M.Z. (2010), Discriminant analysis: an illustrated example, *African Journal of Business Management*, **4**(9): 1654–1667, doi: 10.5897/AJBM.9000211.

52. Rankin S. *et al.* (2017), Acoustic classification of dolphins in the California Current using whistles, echolocation clicks, and burst pulses, *Marine Mammal Science*, **33**(2): 520–540, doi: 10.1111/mms.12381.

53. Reljin N., Pokrajac D. (2017), Music performers classification by using multifractal features: a case study, *Archives of Acoustics*, **42**(2): 223–233, doi: 10.1515/aoa-2017-0025.

54. Sakthivel M., Gopakumar G., Ramkumar S., Mhatre V., Jose S.T. (2014), *Development of protocols for stranding/beaching and post-mortem analysis of cetaceans for the capacity building of officials and local people of Sindhudurg district of Maharashtra*, Project Report. Central Marine Fisheries Research Institute, Mandapam, http://eprints.cmfri.org.in/id/eprint/10468.

55. Seavy N.E., Quader S., Alexander J.D., Ralph C.J. (2005), Generalized linear models and point count data: statistical considerations for the design and analysis of monitoring studies, [in:] Ralph C.J., Rich T.D. [Eds], *Bird Conservation Implementation and Integration in the Americas: Proceedings of the Third International Partners in Flight Conference*, March 20–24, 2002; Asilomar, California, Vol. 2, General Technical Report PSW-GTR-191, Albany, CA: U.S. Dept. of Agriculture, Forest Service, Pacific Southwest Research Station, p. 744–753.

56. Shin K.-S., Lee T.S., Kim H.-J. (2005), An application of support vector machines in bankruptcy prediction model, *Expert Systems with Applications*, **28**(1): 127–135, doi: 10.1016/j.eswa.2004.08.009.

57. Sugumaran V., Muralidharan V., Ramachandran K. (2007), Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing. *Mechanical Systems and Signal Processing*, **21**(2): 930–942, doi: 10.1016/j.ymssp.2006.05.004.

58. Tang Z., Su Y., Er M.J., Qi F., Zhang L., Zhou J. (2015), A local binary pattern based texture descriptors for classification of tea leaves, *Neurocomputing*, **168**: 1011–1023, doi: 10.1016/j.neucom.2015.05.024.

59. Thode A., Mellinger D.K., Stienessen S., Martinez A., Mullin K. (2002), Depth-dependent acoustic features of diving sperm whales (*Physeter macrocephalus*) in the Gulf of Mexico, *The Journal of the Acoustical Society of America*, **112**(1): 308–321, doi: 10.1121/1.1482077.

60. Thorne L.H. *et al.* (2012), Predictive modeling of spinner dolphin (*Stenella longirostris*), resting habitat in the main Hawaiian Islands, *PLoS One*, **7**(8): e43167, doi: 10.1371/journal.pone.0043167.

61. Tiwari V. (2010), MFCC and its applications in speaker recognition, *International Journal on Emerging Technologies*, **1**(1): 19–22.

62. Valero X., Alías F. (2012), Hierarchical classification of environmental noise sources considering the acoustic signature of vehicle pass-bys, *Archives of Acoustics*, **37**(4): 423–434.

63. Young V.W., Hines P.C. (2007), Perception-based automatic classification of impulsive-source active sonar echoes, *The Journal of the Acoustical Society of America*, **122**(3): 1502–1517, 10.1121/1.2767001.