IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 1 August 2024, accepted 30 August 2024, date of publication 5 September 2024, date of current version 18 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3454815

RESEARCH ARTICLE

Convolutional Neural Networks for Enhancing Detection of Dolphin Whistles in a Dense Acoustic Environment

DAVID SCARADOZZI[1,2,3], (Senior Member, IEEE), ROCCO DE MARCO[4], DANIEL LI VELI[4], ALESSANDRO LUCCHETTI[3,3], LAURA SCREPANTI[1], (Member, IEEE), AND FRANCESCO DI NARDO[1]

[1]Dipartimento di Ingegneria dell'Informazione, Università Politecnica delle Marche, Ancona, 60131 Ancona, Italy
[2]ANcybernetics, Università Politecnica delle Marche, 60131 Ancona, Italy
[3]National Biodiversity Future Center, 90133 Palermo, Italy
[4]Institute of Biological Resources and Marine Biotechnology (IRBIM), National Research Council (CNR), 60125 Ancona, Italy

Corresponding author: Francesco Di Nardo (f.dinardo@staff.univpm.it)

This work was supported in part by the National Recovery and Resilience Plan (NRRP), Mission Four Component Two Investment 1.4 (Call for tender No. 3138 of December 2021, rectified by Decree n.3175 of December 2021 of Italian Ministry of University and Research funded by the European Union) NextGenerationEU; and in part by the Financial Instrument for the Environment [L'Instrument Financier pour l'Environnement (LIFE)] Financial Instrument of the European Community, Life Delfi Project–Dolphin Experience: Lowering Fishing Interactions under Grant LIFE18NAT/IT/000942.

ABSTRACT Latest developments in acoustic research suggest that using surveying methods based on artificial intelligence (AI) could improve the effectiveness of underwater monitoring. Passive acoustic monitoring (PAM) has proven to be a cost-effective approach for gathering information about the acoustic behavior of dolphins and plays a crucial role in studying their vocalizations, particularly whistles. This study investigates the efficiency of a binary convolutional neural network (CNN) in detecting dolphin whistles amidst high-density vocalizations in an aquatic environment. Specifically, this analysis intends to determine whether a properly trained CNN can recognize a single whistle even in challenging condition, including situations where multiple dolphins vocalize simultaneously, resulting in overlapping whistles that may have different shapes and durations. To this aim, experimental trials were conducted at Oltremare marine park, Riccione, Italy, where underwater recordings of seven-dolphin vocalizations were collected over 22 consecutive hours. The CNN was trained on labeled whistle spectrograms. The model, comprising three convolutional layers followed by max pooling layers and rectified linear unit (ReLU) activation functions, was evaluated using a 10-fold cross-validation approach. Confusion matrix and performance metrics indicate that the proposed approach achieves results comparable to those reported in the literature, despite the more challenging working conditions. The study supports the potential of AI models in enhancing passive acoustic monitoring techniques.

INDEX TERMS Artificial intelligence, deep learning, dolphin vocalization, passive acoustic monitoring, whistles.

I. INTRODUCTION

Common bottlenose dolphins (*Tursiops truncatus*), referred to hereafter simply as bottlenose dolphins, are renowned for their sophisticated vocalizations, which play a crucial role in their communication, navigation, and social interactions [1]. Dolphins use three main types of acoustic signals: echolocation clicks, multiple burst pulse signals, and frequency-modulated whistles. Echolocation clicks are short, broadband pulses with frequencies that can extend up to 140 kHz. These clicks are crucial for navigation and foraging, enabling dolphins to construct detailed auditory images of their surroundings [2], [3]. Burst pulse sounds are rapid sequences of clicks or pulses, characterized by their high

The associate editor coordinating the review of this manuscript and approving it for publication was Jiajia Jiang.

repetition rate and variable frequency content. They play a social role, often associated with aggressive interactions, such as those observed during depredation, and may be used to resolve rank conflicts and reduce competition among group members [4]. Whistles are frequency-modulated narrow-band sounds used predominantly for social communication. Moreover, whistles can contain unique signature patterns that can identify individual dolphins, highlighting their importance in maintaining social bonds and facilitating individual recognition within pods. These vocalizations are characterized by frequencies typically ranging from 1 up to 25 kHz and durations varying from 0.1 to some seconds [1].

Over the years, passive acoustic monitoring (PAM) has emerged as a non-invasive and cost-effective method for acquiring long-term insights into the presence of dolphin populations, their behavior, social structures, and habitats [5]. It has proved especially beneficial for gathering data during nighttime and adverse weather conditions [6]. Whistles are frequently used to indicate the presence of dolphins across various marine environments. These sounds are especially valuable in passive acoustic monitoring due to their distinct acoustic features, which facilitate detection and analysis [7], [8]. Whistles are typically analyzed using spectrograms, which are visual representations of the frequency spectrum of a signal as it varies with time [9]. A spectrogram provides a detailed view of the frequency content of a signal, allowing to observe the temporal patterns and frequency modulations of dolphin whistles. On spectrograms, dolphin whistles appear as continuous, curved lines that vary in frequency over time.

Recent advancements in artificial intelligence (AI) have significantly enhanced the capabilities of automated whistle detection. AI approaches, particularly deep learning methods, have shown great promise in identifying and classifying dolphin whistles with high accuracy. The specific shape of the whistle in the spectrogram, indeed, appears to be particularly suitable for automatic identification by an AI architecture. These methods typically involve training neural networks on large datasets of spectrograms, enabling the networks to learn the distinctive patterns and features of dolphin whistles. Numerous studies in literature highlighted the potential of AI models for monitoring cetaceans and identifying dolphin presence through whistles. It has been reported, indeed, that whistle-detection performance could be improved over classical approaches adopting convolutional neural networks [10], [11], even in presence of relevant environmental noise [12]. A further study highlighted that the semantic segmentation of whistle by neural networks could contribute to this virtuous process [13]. Furthermore, deep learning techniques have also been successfully applied to the traditional task of classifying whistles into various classes [14], [15].

Despite these promising studies, several challenges must be still addressed. Different factors should be taken into account, including variability in environmental conditions, complexity of dolphin vocalizations, and differences in datasets. Marine environments are highly dynamic, with factors such as water depth, temperature, salinity, sea currents, and background noise varying significantly across different locations and times. These variations can affect the quality and characteristics of recorded whistle sounds, making the detection task challenging for AI models [16]. A further significant challenge is posed by the complexity of this dolphin vocalization. Dolphin whistles, indeed, can vary not only between species but also among individuals of the same species. Dolphins can also modify their whistles in response to social interactions, environmental changes, and other stimuli. Furthermore, it is well-known that dolphins often move in pods, so it frequently happens that recordings contain overlapping whistles from different dolphins [1], [7], [9]. Moreover, differences in datasets used for training and testing the model can lead to inconsistent results. The availability of large, standardized, and diverse datasets is crucial for developing robust AI models capable of performing reliably in different environments.

The current study is designed to test the performance of a convolutional neural network (CNN) in detecting bottlenose dolphin whistles in underwater recordings with a high density of whistles over time. Specifically, this analysis intends to determine whether a properly trained CNN can recognize a single whistle even when multiple dolphins are vocalizing simultaneously, emitting whistles that may have different meanings and purposes. To this aim, underwater sound recording procedure was performed in a series of interconnected pools of a dolphin park where seven free-to-swim bottlenose dolphins were engaged in their daily activities, which included playing, eating, and exercising.

II. MATERIALS AND METHODS

A. SIGNAL RECORDING AND PROCESSING

Recording trials were performed at the Oltremare thematic marine park in Riccione, Italy. Acquisition started at 10:22 in the morning of November 20, 2021, and stopped after slightly more than 22 hours. Underwater acoustic signals were recorded submerging the recording systems in a series of interconnected pools with seven bottlenose dolphins (*T. truncatus*). The recording system was composed of the SQ26-05 hydrophone (Sensor Technology) associated with the UREC 384K autonomous underwater recorder (Dodotronic and Nauta). The sensitivity of the hydrophone is -193.5 dB re 1 V/μPa @ 20 °C between at least 1 Hz and 28 kHz [17]. Acquisition characteristics are: sampling frequency $=$ 192 kHz and resolution $=$ 16 bit. Signals were stored as wave files lasting 5 minutes each.

Stored signals were min-max normalized, scaling the values to the [0–1] range, and then appropriately processed using a band-pass filter between 3 and 24 kHz. Each 5-minute signal block was analyzed using the spectrogram visualization of the open-source software Audacity. A trained PAM expert reviewed the spectrograms with the aim of visually inspecting and labeling the dolphin whistles. The spectrograms were then segmented into parts of 0.8 seconds and the

whistle was centered within each segment. The spectrograms were converted in jpeg images. The segment length was set 0.8 seconds since this value statistically allows to contain almost all the whistles [18]. If the labeled signal is longer than 0.8 seconds, it was split into 0.8-second segments, with a 50% overlap, until the whole signal is covered. The size of each spectrogram was 300 × 150 pixels. The release of the whistle-labelled dataset is still in preparation. Nonetheless, the data could be already available by contacting the authors of the current study.

B. TRAINING THE MODEL

Convolutional neural network (CNN) has been largely adopted for speech recognition and for audio-related studies, including studies on dolphins [10], [11], [12], [13], [14], [15], [19]. In the current study, CNN is employed [20], consisting of three convolutional layers with 32, 64, and 128 filters, respectively, with kernel size of 6 × 3. General practice for CNN design, indeed, suggests that simple tasks with small datasets might require 2-3 layers with 32-128 filters. The rectified linear unit (ReLU) activation function is applied to every layer. Following the convolutional layer is a max pooling layer with a 2 × 2 window, which decreases the spatial dimensionality of the output, thereby lessening the calculation complexity and contributing to avert the risk of overfitting. The output of the convolutional and pooling layers is flattened into a one-dimensional vector, which is then processed by a dense layer with 128 units and ReLU activation. The final dense layer comprises a single neuron utilizing sigmoid activation for binary classification. The above-described CNN model was trained and tested on spectrograms described in the paragraph II-A. The whole dataset includes 6000 labelled spectrograms. To ensure balance during training, an equal number of spectrograms labeled "1" and "0" were used.

Then, the whole dataset was separated into 10 folds, including the same number of whistles. The spectrograms from 9 out of the 10 folds were utilized to train the model. The spectrograms from the remaining fold were employed to test the model performance. Ten different trainings were performed, following the 10-fold cross-validation depicted in Figure 1.

C. PERFORMANCE MEASUREMENTS

Global confusion matrix and mean values over the ten folds of accuracy, precision, recall, and F1-score were adopted to quantify CNN performances. The computation of confusion matrix is based on the values achieved for true negatives (TN), false negatives (FN) true positives (TP), and false positives (FP). The other performance metrics are defined as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

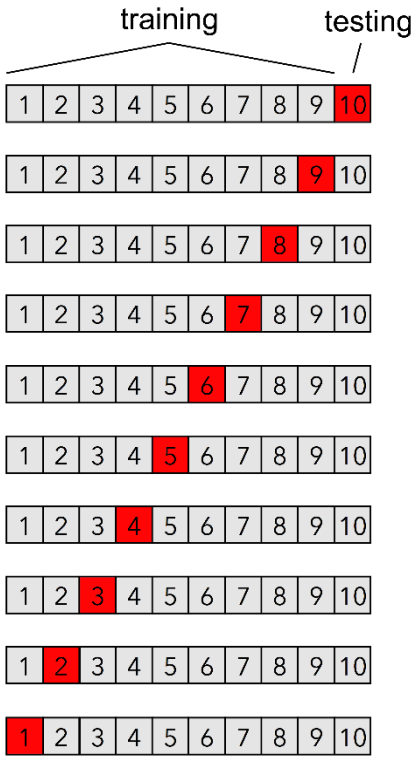


FIGURE 1. The 10-fold cross-validation procedure.

$$\text{F1} - \text{score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

III. RESULTS

Figure 2 depicts an example of a colored spectrogram identified by the PAM expert and included in the dataset used for this study. As evident from Figure 2, the whistle becomes recognizable around 0.2 s and ends around 0.6 s. Approximately 0.3 s after the start of the spectrogram, other underwater sounds overlap with the whistle. These further signals could represent either other types of dolphin vocalizations (such as echolocation clicks or burst pulse sounds) or ambient noise. Both are considered noise for the purposes of this analysis.

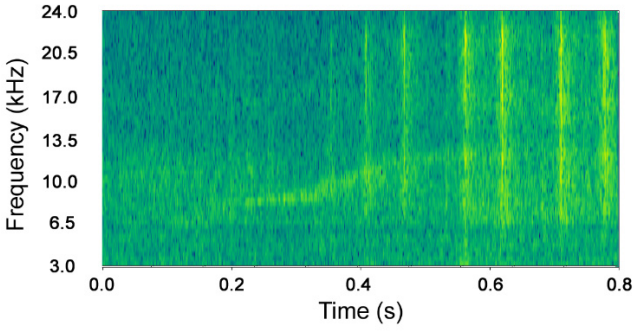


FIGURE 2. An example of 0.8s spectrogram of a representative whistle recorded during the acquisition trials.

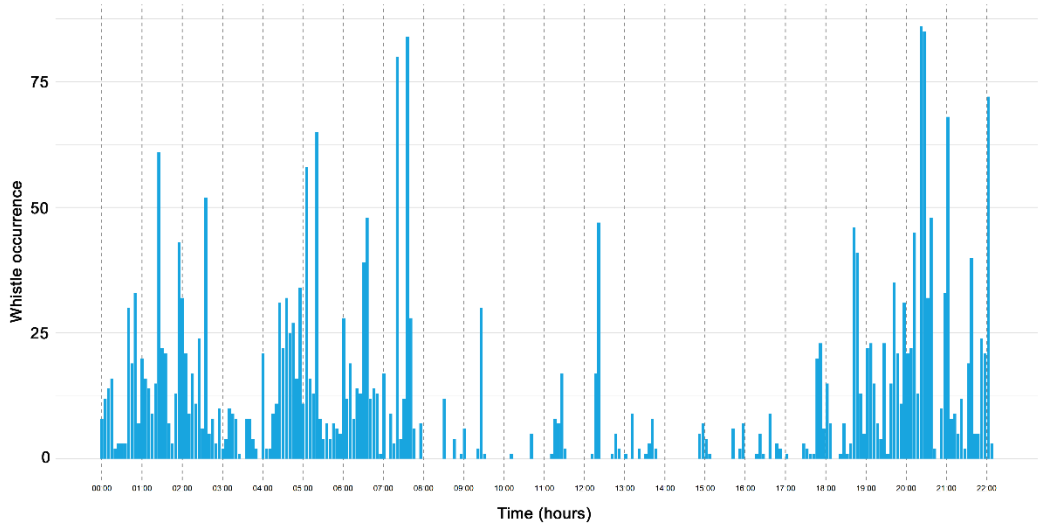


FIGURE 3. Density of whistle occurrences over time. Every blue bar represents the number of whistles detected by the PAM expert in the correspondent 5-minute segment. Time 00:00 on the x-axis indicates the starting event of the 22-hour recording session.

The present dataset contains dolphin whistles of variable duration ranging from a minimum of 0.053 s to a maximum of 4.98 s. In order to quantify the density of whistle occurrences over time, Figure 3 reports, in the y-axis, the number of whistles identified by the PAM expert in each 5-minute interval into which the 22 hours of recording were split. It is worth noting that after 20 hours and 20 minutes of recordings, the graph reported two consecutive peaks of 86 and 85 whistles, respectively. This means that in 10 minutes a total of 171 whistle were detected, indicating a very high density of whistle occurrence in this part of the recording signal.

Confusion matrix is depicted in Figure 4, reporting grey-scale normalized values of TN, FN, TP, and FP. Progressively darker colors indicate progressively higher values, up to 1 that is represented by black color.

The darkest square in the confusion matrix indicates that the CNN correctly identifies 92.7% of the spectrograms that do not contain whistles. Similarly, the bottom-right square shows that the CNN correctly identifies 81.3% of the spectrograms containing whistles. Detailed percentage values of the classification performances in each one of the ten folds are shown in Table 1.

Table 1 shows as each metric presented small variability within ten folds. Indeed, accuracy is between 83.0% and 90.2%; precision is between 86.8% and 95.0%; recall is between 77.8% and 84.8; and F1–score is between 82.1% and 89.6%. The last two rows of the table indicate the average values an SDs over ten folds. Mean accuracy indicates that the model provides 87.0 $\pm$ 2.4% correct predictions (both true positives and true negatives) out of all predictions made. Average precision indicates that the percentage of correctly predicted whistles out of all predicted whistles is 91.7 $\pm$ 2.9%. Average recall shows that the percentage of actual whistles that were correctly identified by the model is

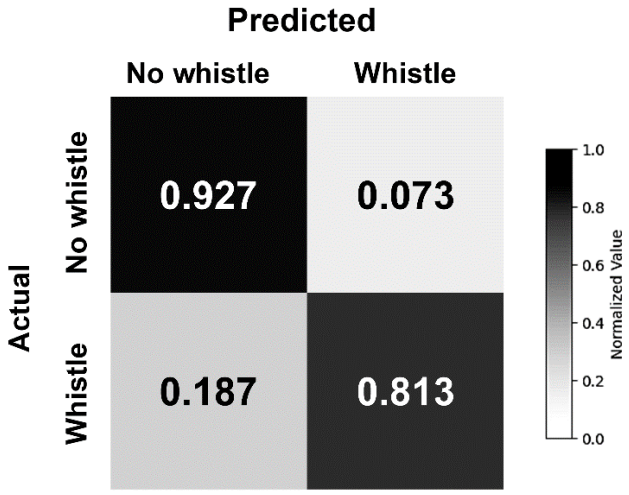


FIGURE 4. A Confusion matrix. Data are reported as mean value over ten folds and normalized between 0 (white) and 1 (black).

81.3 $\pm$ 2.3%. It is worth noting that, despite the complexity of the analyzed dataset, the mean percentage value of the synthetic metric F1-score is still above 86% and detailed values never drop below 82%.

The present outcomes have been achieved using a Vivobook Pro 15 N580GD laptop, equipped with an Intel Core™ i7-8750H processor clocked at 2.2 GHz. The installed RAM is 16 GB, and the operating system is Windows 11 64-bit. The GPU utilized is an NVIDIA GeForce GTX 1050 with refresh rate of 60 KHz.

IV. DISCUSSION

One of the most challenging factors affecting the performance of a neural network trained to identify dolphin whistles from

TABLE 1. Average percentage performance metrics over ten folds.

Fold	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
1	86.0	91.6	79.2	85.0
2	89.0	94.1	83.2	88.3
3	83.0	86.8	77.8	82.1
4	84.8	88.6	79.7	83.9
5	84.9	88.5	80.3	84.2
6	87.8	93.8	81.0	87.0
7	90.2	95.0	84.8	89.6
8	86.4	90.9	80.9	85.6
9	89.9	94.3	84.8	89.3
10	87.8	93.8	81.0	87.0
mean	87.0	91.7	81.3	86.2
SD	2.4	2.9	2.3	2.5

underwater recordings is the overlap of sounds from multiple dolphins vocalizing simultaneously [21]. Additionally, the fact that each individual dolphin is characterized by a different signature whistle in terms of waveform, frequency, and duration [22] and that dolphins can modulate their whistles depending on intentions [23], further complicate the automatic classification of these vocalizations. This study aimed to evaluate the performance of a convolutional neural network on a dataset that, by its nature, presents both aforementioned challenges. To pursue this aim, the experimental recordings were conducted in a controlled environment where seven dolphins shared all the activities that are typically carried out daily in a thematic marine park, such as playing, eating, and exercising. This resulted in the recording of a very high number of whistles with different characteristics, as highlighted in Figure 3.

Each bar in this figure represents the number of whistles identified within a 5-minute interval by a PAM expert trained for this purpose, regardless of which dolphin the whistle came from or what activity the dolphin was engaged in during that period. A very intense vocalization activity was recorded during the first eight hours of recording. As reported in Section II-A, the recording started at 10:22 a.m. on November 20, 2021. Therefore, the intense vocalization activity corresponded to the daytime period when all the seven dolphins engaged in their typical daily activities, frequently interacting with each other and with the trainers. At the end of this period, there is a noticeable decrease in activity between the eighth and nineteenth hours, which corresponds to the nighttime period (06:22 p.m. to around 5 a.m. of the following day). This is followed by a new increase in vocalizations due to the resumption of the dolphin daily activities. Furthermore, it is worth noting that the whistles included in the present dataset exhibit a large variability also in duration, ranging from 0.053 s (very short whistles) to 4.98 s (very long whistles).

As highlighted by the analysis of vocalization timing and duration, the adopted CNN had to work under challenging conditions, characterized by high variability of whistles and frequent interconnections and overlapping among them. Nevertheless, the CNN still managed to identify 93% of the spectrogram with ''no whistle'' labelled by the PAM expert, as reported in the confusion matrix (Figure 4). Lower values, but still above 80%, were achieved in the task of whistle detection where percentage TPs are 81.3%. This lower value is likely due to the overlapping among whistles from different dolphins and to the superimposition with further dolphin vocalization, such as clicks and burst pulse sounds, as for example in Figure 2. The detailed classification performances reported in Table 1 also seem to support what emerged from the confusion matrix. Encouraging average values were, indeed, achieved for both classification accuracy (87.0%) and average F1-score (86.2%). The low standard deviations associated with the mean values of accuracy and F1-score across the 10 folds (approximately 2.5%) indicate that the model performed consistently across different subsets. This indicated that the observed performance is a true reflection of the model capabilities rather than being influenced by random variations in the data, thus suggesting high consistency and reliability of the model performance. These promising results are also supported by comparisons with what, to the best of our knowledge, are the only two other studies in the literature that adopt similar approaches [12], [13]. Average accuracy, precision, recall, and F1-score are, indeed, in line with what was reported in the recent study by Jin et al. [13], who indicated a mean accuracy of 89%, precision of 96%, recall lower than 80%, and an F1-score of 86% achieved in the attempt to use a CNN-based semantic segmentation model for whistle profile extraction. Moreover, the current performances in Table 1 are not far off from those reported by Nur Korkmaz et al. [12], who tested and compared the performance achieved by a vanilla CNN and by a pre-trained CNN based on the VGG16 architecture. In particular, the CNN experimented here shows a higher average precision ($91.7 \pm 2.9\%$ vs. 90.5%), whereas the performances reported in [12] are better in terms of accuracy ($87.0 \pm 2.4\%$ vs. 92.3%) and especially recall ($81.3 \pm 2.3\%$ vs. 89.6%), although the average F1-score was not reported. One of the main reasons for these differences could be ascribed to the fact that Nur Korkmaz et al. systematically excluded from analysis all whistles longer than 0.78 s, whereas in the present study, no long whistle was discarded, as indicated in Section II-A and in the ''Results''. Longer whistles, indeed, are more complex to identify correctly because, exceeding the observation window in duration, they may not be uniquely identifiable and can lead to various false positives that undermine performance. On the other hand, their length increases the risk of overlap with whistles from other dolphins, thereby leading to potential false negatives. Moreover, in [12] raw data were strongly processed to remove sporadic cut-offs and extensive noise periods. In this current study, the decision was made not to preprocess the data in order to preserve the integrity

of the research, which focuses on testing the capabilities of a CNN on a dataset characterized by high variability. This variability arises because the dataset includes vocalizations from multiple dolphins, sometimes vocalizing together, and is associated with various dolphin activities. Preprocessing could potentially distort or simplify this natural variability, which is crucial for accurately evaluating the CNN performance in real-world, complex scenarios involving diverse dolphin behaviors and interactions.

Although this study provided encouraging indications for the employment of artificial intelligence systems to identify dolphin whistles even in complex environmental conditions, some limitations must be discussed. The present model has been trained with data recorded form seven bottlenose dolphins ranging in interconnected pools within a marine park. and validated in this dataset with these specific characteristics of the animals and their surrounding environment. Variations among individual dolphins and different species can affect the model ability to generalize. Additionally, variability in environmental conditions and environmental noise can introduce distortions, leading to potential inaccuracies and deterioration of model performances.

However, further approaches can be explored in the effort to improve and generalize the robustness of the current method. From this perspective, it could be useful to test signal processing techniques capable of highlighting the morphological characteristics of whistles and minimizing the effect of other vocalizations or ambient noise on the whistle identification process, such as edge detection algorithms [24]. As previously noted, whistles are generally studied through spectrograms, which visually represent the frequency spectrum of a signal over time [25]. The spectrogram offers a detailed perspective on signal frequency content, enabling the analysis of temporal patterns and frequency modulation of dolphin whistles. In spectrograms, dolphin whistles display characteristics that resemble edges in images, often appearing as distinct, continuous lines. This resemblance implies that edge detection techniques commonly used in image processing could be advantageous also in enhancing spectrograms for CNN-based detection of dolphin whistles.

Moreover, in the vast majority of cases, dolphin whistles overlap with other vocalizations such as echolocation clicks and burst pulse sounds, making it even more challenging to identify individual vocalizations. Further specific pre-processing of the audio signal or the spectrogram image could help isolate the individual vocalization, improving the ability to uniquely identify it, and consequently enhancing the model performance. Further studies will focus on identifying the most suitable digital signal processing technique, including edge detection, to generalize to different dolphin species and/or in different marine environments the promising findings achieved in the present study. All of this should be done without flattening the natural variability that characterizes the unique acoustic behavior of each individual dolphin, both alone and in pods.

V. CONCLUSION

This research highlighted the potential of integrating deep learning with acoustic monitoring to address complex environmental challenges. The current results suggest that artificial intelligence techniques can significantly contribute to dolphin monitoring, even under challenging conditions like those in the open sea. Advancements in AI and machine learning, indeed, have the potential to revolutionize how marine ecosystems are studied and monitored. The automatic identification of whistles through artificial intelligence also opens novel possibilities for dolphin conservation in the marine environment; the identification of a trigger could be associated with the development of new and more efficient pingers, the creation of alarm systems (for example, in port areas), and so on. Further studies are nonetheless needed, focusing on refining the CNN model, improving the data preparation, expanding the dataset, and exploring the application of similar techniques to other marine species and environments.

ACKNOWLEDGMENT

A special thanks to Stefano Furlati, Barbara Marchiori, the dolphin trainers and all the Oltremare family experience park staff for their support in the experiment.

REFERENCES

[1] M. O. Lammers and J. N. Oswald, "Analyzing the acoustic communication of dolphins," in *Dolphin Communication and Cognition: Past, Present, and Future*. Cambridge, MA, USA: MIT Press, 2015, pp. 107–138.

[2] W. W. L. Au, *The Sonar of Dolphins*. New York, NY, USA: Springer, 1993.

[3] T. Akamatsu, D. Wang, K. Nakamura, and K. Wang, "Echolocation range of captive and free-ranging baiji (*Lipotes vexillifer*), finless porpoise (*Neophocaena phocaenoides*), and bottlenose dolphin (*Tursiops truncatus*)," *J. Acoust. Soc. Amer.*, vol. 104, pp. 2511–2516, Oct. 1998.

[4] M. O. Lammers, W. W. L. Au, and D. L. Herzing, "The broadband social acoustic signaling behavior of spinner and spotted dolphins," *J. Acoust. Soc. Amer.*, vol. 114, no. 3, pp. 1629–1639, Sep. 2003.

[5] K. J. Palmer, K. L. Brookes, I. M. Davies, E. Edwards, and L. Rendell, "Habitat use of a coastal delphinid population investigated using passive acoustic monitoring," *Aquatic Conservation, Mar. Freshwater Ecosyst.*, vol. 29, no. S1, pp. 254–270, Sep. 2019.

[6] P. M. Thompson, K. L. Brookes, and L. S. Cordes, "Integrating passive acoustic and visual data to model spatial patterns of occurrence in coastal dolphins," *ICES J. Mar. Sci.*, vol. 72, no. 2, pp. 651–660, Jan. 2015.

[7] M. Azzolin, A. Gannier, M. O. Lammers, J. N. Oswald, E. Papale, G. Buscaino, G. Buffa, S. Mazzola, and C. Giacoma, "Combining whistle acoustic parameters to discriminate Mediterranean odontocetes during passive acoustic monitoring," *J. Acoust. Soc. Amer.*, vol. 135, no. 1, pp. 502–512, Jan. 2014.

[8] M. Gregorietti, E. Papale, M. Ceraulo, C. de Vita, D. S. Pace, G. Tranchida, S. Mazzola, and G. Buscaino, "Acoustic presence of dolphins through whistles detection in Mediterranean shallow waters," *J. Mar. Sci. Eng.*, vol. 9, no. 1, p. 78, Jan. 2021.

[9] O. M. Serra, F. P. R. Martins, and L. R. Padovese, "Active contour-based detection of estuarine dolphin whistles in spectrogram images," *Ecol. Informat.*, vol. 55, Jan. 2020, Art. no. 101036.

[10] Y. Shiu, K. J. Palmer, M. A. Roch, E. Fleishman, X. Liu, E.-M. Nosal, T. Helble, D. Cholewiak, D. Gillespie, and H. Klinck, "Deep neural networks for automated detection of marine mammal species," *Sci. Rep.*, vol. 10, no. 1, pp. 1–12, Jan. 2020.

[11] J.-J. Jiang, L.-R. Bu, F.-J. Duan, X.-Q. Wang, W. Liu, Z.-B. Sun, and C.-Y. Li, "Whistle detection and classification for whales based on convolutional neural networks," *Appl. Acoust.*, vol. 150, pp. 169–178, Jul. 2019.

[12] B. N. Korkmaz, R. Diamant, G. Danino, and A. Testolin, "Automated detection of dolphin whistles with convolutional networks and transfer learning," *Frontiers Artif. Intell.*, vol. 6, Jan. 2023, Art. no. 1099022.

[13] C. Jin, M. Kim, S. Jang, and D.-G. Paeng, "Semantic segmentation-based whistle extraction of indo-pacific bottlenose dolphin residing at the coast of Jeju island," *Ecol. Indicators*, vol. 137, Apr. 2022, Art. no. 108792.

[14] L. Li, G. Qiao, S. Liu, X. Qing, H. Zhang, S. Mazhar, and F. Niu, "Automated classification of tursiops aduncus whistles based on a depth-wise separable convolutional neural network and data augmentation," *J. Acoust. Soc. Amer.*, vol. 150, no. 5, pp. 3861–3873, Nov. 2021.

[15] D. Duan, L.-G. Lü, Y. Jiang, Z. Liu, C. Yang, J. Guo, and X. Wang, "Real-time identification of marine mammal calls based on convolutional neural networks," *Appl. Acoust.*, vol. 192, Apr. 2022, Art. no. 108755.

[16] J. Schneider, A. Klüner, and O. Zielinski, "Towards digital twins of the oceans: The potential of machine learning for monitoring the impacts of offshore wind farms on marine environments," *Sensors*, vol. 23, no. 10, p. 4581, May 2023.

[17] Nauta Scientific. (2024). *SQ26-05 Hydrophone, Sensor Technology*. Accessed: May 22, 2024. [Online]. Available: https://www.nauta-rcs.it/EN/Hydrophones/SensorTech/

[18] F. Di Nardo, R. De Marco, A. Lucchetti, and D. Scaradozzi, "A WAV file dataset of bottlenose dolphin whistles, clicks, and pulse sounds during trawling interactions," *Sci. Data*, vol. 10, no. 1, p. 650, Sep. 2023.

[19] O. Abdel-Hamid, L. Deng, and D. Yu, "Exploring convolutional neural network structures and optimization techniques for speech recognition," in *Proc. Interspeech*, Lyon, France, Aug. 2013, pp. 1173–1175.

[20] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," in *The Handbook of Brain Theory and Neural Networks*. Cambridge, MA, USA: MIT Press, 1998.

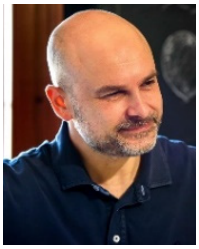
[21] P. Tyack, "Whistle repertoires of two bottlenosed dolphins, *Tursiops truncatus*: Mimicry of signature whistles?" *Behav. Ecol. Sociobiol.*, vol. 18, pp. 251–257, Feb. 1986.

[22] G. La Manna, N. Rako-Gospić, D. S. Pace, S. Bonizzoni, L. Di Iorio, L. Polimeno, F. Perretti, F. Ronchetti, G. Giacomini, G. Pavan, G. Pedrazzi, H. Labach, and G. Ceccherelli, "Determinants of variability in signature whistles of the Mediterranean common bottlenose dolphin," *Sci. Rep.*, vol. 12, no. 1, p. 6980, May 2022.

[23] V. M. Janik, D. Todt, and G. Dehnhardt, "Signature whistle variations in a bottlenosed dolphin, tursiops truncatus," *Behav. Ecol. Sociobiol.*, vol. 35, no. 4, pp. 243–248, Oct. 1994.

[24] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[25] D. Gillespie, M. Caillat, J. Gordon, and P. White, "Automatic detection and classification of odontocete whistles," *J. Acoust. Soc. Amer.*, vol. 134, no. 3, pp. 2427–2437, Sep. 2013, doi: 10.1121/1.4816555.



DAVID SCARADOZZI (Senior Member, IEEE) received the B.Eng., M.Eng., and Ph.D. degrees, the degree in electronic engineering from Università Politecnica delle Marche, in 2001, and the Ph.D. degree in intelligent artificial systems. Since 2007, he has been an Assistant Professor with the Department of Information Engineering, Università Politecnica delle Marche (UNIVPM). He is an accredited Engineer and an Assistant Professor with Dipartimento di Ingegneria dell'Informazione and International Mobility/Networking Delegate of the Engineering Faculty, UNIVPM. At UNIVPM, he is a member of the Professors Board of the Ph.D. School in Information Engineering and the Chair of the courses: "Model Identification and Data Analysis" (bachelor's degree in automation engineering); "Design and Optimization of the Control Systems" (master's degree in automation engineering); and "Advanced Virtual Instruments for Simulation and Control of Complex Systems" (Ph.D. school in information engineering). He is the author of about 90 publications in refereed international journals, books, and conferences. His research interests include control and optimization of dynamical systems, robotics and automation (motion and interaction control problems in distributed agents; rapid prototyping and mechatronics), underwater robotics and marine technologies, focusing on tools for 3D scientific documentation of sea operative surveys for marine protected areas and archaeological sites study using divers, AUVs, ROVs and other technological devices, educational robotics, with special interests devoted to all the aspects regarding the identification and assessment of education processes, study and development of new robotic tools and lesson plans for teaching e-STrEM (environmental science technology robotics engineering maths) subjects, in formal, and nonformal education.

ROCCO DE MARCO received the bachelor's degree in computer science from Università di Camerino, Italy, in 2006. From 2002 to 2005, he was an IT Manager with SGD Adriatico Centrale, Italian Institute for the Physics of Matter (INFM). Since 2005, he has been a Technical Research Collaborator with Italian National Research Council (CNR) developing acquisition devices and performing data curation and analysis, in particular of marine data. In 2019, he was the responsible of the computation service of the Institute for Biological Resources and Marine Biotechnology (IRBIM). He has the author of 17 peer reviewed articles and more than 40 technical/scientific reports.

DANIEL LI VELI received the Mar.Biol. and M.S. degrees. Since 2019, he has been a Research Fellow with CNR-IRBIM, focusing on applied research in marine biology. His expertise includes marine resource management, fishing technology, and assessing the impact of fishing. His primary research activities involve testing devices to reduce bycatch of protected marine species, studying gear selectivity, addressing fishing discards, and promoting sustainable marine resource management.

ALESSANDRO LUCCHETTI is currently a Senior Researcher with more than 20 years of experience in fishing technology, fisheries management and biology, and fisheries impact assessment. He is a coordinator and scientific responsible of numerous international and national projects. He is the author and co-author of more than 70 publications in international journals and more than 200 other publications, including university textbooks, and manuals. He has collaborated as an Expert with Ministries, European Commission, FAO, NGOs, and private companies. He has lectured in several Italian universities and in other European and non-European countries. He has lectured for 15 years for the Coast Guard personnel on fisheries control systems and related regulations.

LAURA SCREPANTI (Member, IEEE) received the B.S. degree in biomedical engineering, the M.Sc. degree in electronic engineering, and the Ph.D. degree in information engineering from Università Politecnica delle Marche, Ancona, Italy, in 2011, 2014, and 2020, respectively. After a postdoctoral position with Università Politecnica delle Marche. From 2015 to 2016, she was a Research Assistant with the Department of Biomedical Sciences and public health for the development of a mechatronic infrastructure gathering biometric signals from SCUBA divers during their underwater activity. During her Ph.D., she worked on modeling and identification of learning systems during educational robotics activities. She is currently a Research Fellow with the Department of Education, Cultural Heritage and Tourism, University of Macerata. She is an Adjunct Professor of systems modeling and identification with Università Politecnica delle Marche and a Teacher of real-time systems at the "Fondazione ITS Nuove Tecnologie per il Made in Italy." She has been appointed as an Expert on Educational Robotics and innovative STEAM methodologies to fight against gender stereotypes by the national institute of documentation, innovation and research in education (INDIRE). Her research interests include Educational Robotics, systems modeling and identification, and smart sensors development. She is a member of the IEEE Education Society and the IEEE Robotics and Automation Society.

FRANCESCO DI NARDO received the master's degree in electronic engineering and the Doctor of Philosophy (Ph.D.) degree in artificial intelligence systems from Università Politecnica delle Marche, Ancona, Italy, in 2000 and 2005, respectively. He is currently a Senior Staff Scientist with the Laboratory of Modeling, Analysis and Control of Dynamical System (LabMACS), Dipartimento di Ingegneria dell'Informazione, Università Politecnica delle Marche, and the Head of the Section "Acquisition Systems and Data Processing." He is the Board of Italian Society of Movement Analysis in Clinics (SIAMOC) and a Senior Fellow of the Interuniversity Centre, Bioengineering of the Human Neuromusculoskeletal System (BOHNES). Università Politecnica delle Marche, where he is currently associated. In this and other fields, he is the author and co-author of about 150 publications, including full articles in refereed international journals, chapters in international books, and international conference papers. His main research interests include biomedical and biological signal processing (filtering, feature extraction, pattern recognition, time–frequency analysis, application of neural networks to biosignals, statistical gait analysis) and interpretation.

Open Access funding provided by 'Università Politecnica delle Marche' within the CRUI CARE Agreement