**IEEE** *Access*

# Convolutional Neural Networks for Risso's dolphins identification

**R. MAGLIETTA[1], V. RENÒ[1], R. CACCIOPPOLI[2], E. SELLER[2], S. BELLOMO[3], F. C. SANTACESARIA[3], R. COLELLA[1], G. CIPRIANO[4,5], E. STELLA[1], K. HARTMAN[6], C. FANIZZA[3], G. DIMAURO[2] and R. CARLUCCI[4,5]**

[1]Institute of Intelligent Industrial Technologies and Systems for Advanced Manufacturing, National Research Council, via Amendola 122, Bari, Italy
[2]Department of Computer Science, University of Bari, Via Orabona 4 - 70125 Bari, Italy
[3]Jonian Dolphin Conservation, Viale Virgilio 102 - 74121 Taranto, Italy
[4]Department of Biology, University of Bari, Via Orabona 4 - 70125 Bari, Italy
[5]CoNISMa, Piazzale Flaminio 9 - 00196 Rome, Italy
[6]Nova Atlantis Foundation, Risso's Dolphin Research Centre, Rua Dr. A. F. Pimentel 11, 9930-309, Santa Cruz das Ribeiras, Pico, Azores, Portugal

Corresponding author: Rosalia Maglietta (e-mail: rosalia.maglietta@cnr.it).

**ABSTRACT** Photo-identification is one of the best practices to estimate the abundance of cetaceans and, as such, it can help to obtain the biological information necessary to decision-making and actions to preserve the marine environment and its biodiversity. The Risso's dolphin is one of the least-known cetacean species on a global scale, and the distinctive scars on its dorsal fin proved to be extremely useful to photo-identify single individuals. The main novelty of this paper is the development of a new method based on deep learning, called Neural Network Pool (NNPool), and specifically devoted to the photo-identification of Risso's dolphins. This new method also includes the unique function of recognizing *unknown vs known* dolphins in large datasets with no interaction by the user. Moreover, the new version of DolFin catalogue, collecting Risso's dolphins data and photos acquired between 2013-2018 in the Northern Ionian Sea (Central-eastern Mediterranean Sea), is presented and used here to carry out the experiments. Results have been validated using a further data set, containing new images of Risso's dolphins from the Northern Ionian Sea and the Azores, acquired in 2019. The performance of the NNPool appears satisfying and increases proportionally to the number of images available, thus highlighting the importance of building large-scale data set for the application at hand.

**INDEX TERMS** Cetaceans, classification, deep learning, photo-identification, Risso's dolphin

## I. INTRODUCTION

Top predators such as marine mammals help maintain functionality and resilience of the ecosystem, while actions that contribute to their conservation can generally be beneficial to marine biodiversity [1]. Conducting photo-identification (photo-ID) studies based on the recognition of single individuals through specific markers on their body can help to evaluate the abundance of cetaceans, providing relevant biological information usable for marine environment protection.
The Risso's dolphin *Grampus griseus* (Cuvier, 1812) is one of the least-known cetacean species on a global scale, with Mediterranean subpopulation ranked as Data Deficient by the IUCN Red List [2]. To bridge the gap of understanding this species, a key component is obtained through photo-ID studies. It is, in fact, their appearance, that makes the Risso's dolphins particularly suitable for this kind of research. Commonly, adult Risso's dolphins display extensive white scar-

ring on their bodies, solid grey at birth (see figure 1). These scars, most of which are presumably caused by infraspecific interactions [3], can appear as scratches, stains, or circular marks, and in some animals can cover most of their body surface. As a result, the unique markings on their dorsal fin can be successfully analyzed to identify single individuals. The state of the art for the automated photo-identification of Risso's dolphins is the algorithm SPIR (Smart Photo-identification of Risso's dolphin) [4], [5], where Scale Invariant Feature Transform (SIFT) [6] is applied to detect the keypoints over the scars on the dorsal fin of Risso's dolphins. The purpose of this photo-ID tool is to match the dolphin captured in the query image with the most similar dolphin, in terms of SIFT features, in a catalogue of *known* and labeled Risso's dolphins. Acting as a best-matching algorithm, the peculiarity of SPIR is that it still provides an answer in terms

of probability, even if the dolphin in the query image is *unknown*, that is it has never been sighted before, and for this reason its photos are not included in the reference catalogue. If we were to have photographs of new fins, for example acquired during a new survey, in order to bring about a more powerful photo-identification, the *unknown* class should be considered among the possible identities for the query dolphin. How could *unknown* dolphins be automatically detect? In general, machine learning algorithms, such as Support Vector Machine [7], [8], Random Forest [9], [10], Adaboost [11] and RUSBoost [12]–[14], provide us with strategies able to classify examples never seen before, taking into consideration all desirable classes, and in particular the *unknown* one. With this aim, in our previous paper [14] we provided an initial study on this matter, showing the RUSBoost performances when recognizing the *unknown* Risso's dolphins against one of the previously catalogued and *known* individuals. The selection of this algorithm depended on its ability to manage class imbalance, a straightforward skill considering that the number of images available for each Risso's dolphin, on which the model is trained, is small due to the low sighting frequency of this species [15]. The strategy proposed in [14] can be summarized in three different steps: fin segmentation (fin mask creation), feature extraction inside the mask and RUSBoost classification. The dorsal fin captured in the image was represented by a Super-SIFT, the vector containing the first three principal components of the SIFT computed inside the fin mask, in the Principal Component Analysis (PCA) analysis [16]. The choice fell on SIFT features, based on our previous paper [5], where we have demonstrated that the SIFT-based approach is more robust for the specific application of Risso's dolphin photo-ID. Results of RUSBoost photo-identification on a small data set were quite promising and highlighted the need to apply similar strategies to larger data sets.

Encouraged by the results, the study has been further pursued with several new contributions discussed in this paper. First of all, DolFin catalogue, previously published in [4], has been updated collecting Risso's dolphins data and photos acquired up to 2018 in the Northern Ionian Sea (Central-eastern Mediterranean Sea) by our research team. The updated version of this catalogue, presented here, was used to carry out experiments. Successively, the study proposed in [14] has been applied as follows to the DolFin catalogue. A number of $m$ RUSBoost classifiers were trained to identify *unknown* dolphins against each one of the $m$ *known* individuals selected in the DolFin catalogue. The selection of the $m$ *known* individuals, over the total number $n$ of *known* dolphins in DolFin catalogue was done automatically, based on the accuracy of the fin segmentation, as SIFT features are extracted inside the computed fin mask and are informative for that individual. Ideally, in the best case scenario, the fin mask is correctly computed when the mask contains all fin pixels and no sea pixels; in such a way, SIFT will only be computed over the dolphin?s fin, and no features will be extracted over the sea. Hence, the number $m$ of *known* selected individuals is

related to the number of dolphins whose fin?s image has been well segmented by the algorithm. It is clear that this selection of a subset of individuals aims to improve the performances of RUSBoos. In fact, the application of RUSBoost to all the $n$ *known* individuals worsens its performances. Now, a new contribution of this paper is RUSPool, a strategy focusing on the identification of *unknown vs known* Risso's dolphins, where the *known* dolphins are all the $m$ individuals previously chosen in the catalogue. In this case, the aim is not to predict the exact identity of the *known* individual. RUSPool consists of a smart merge, by means of a tailored filter, of the $m$ RUSBoost classifiers, trained as previously described.

Finally, the main novelty of this paper is the application of Convolutional Neural Network (CNN) [17] models to the Risso's dolphin photo-identification pipeline. In particular, the aim of this paper is to investigate the advantages and drawbacks of the modern CNN, when compared to RUSBoost, in the analysis of unbalanced data sets. CNN was selected among others because widely used in literature with excellent performances throughout many applications [18]–[25], and most importantly, because it does not require an explicit computation of external features, unlike RUSBoost which requires SIFT extraction, thus overcoming the previously discussed problems, related to fin segmentation and feature selection. Hence no selection of dolphins is done in the CNN analysis, and it will be applied over all the $n$ *known* dolphins listed in DolFin catalogue. A number of $n$ CNNs were trained to recognize the *unknown* dolphins against each one of the $n$ *known* individuals. A total number of $n$ CNNs was trained.

Lastly, a new methodology, called Neural Network Pool (NNPool), is presented here, faced with the specific task of recognizing *unknown vs known* Risso's dolphins, similarly to what has already been done with RUSPool. NNPool consists of a pool of the $n$ CNNs, trained as described. Its output simply consists of a major voting of the outputs of the $n$ CNNs. Both NNPool and RUSPool can be employed to automatically process extensive amounts of data, without user interaction, and their accuracy can be applied to identify *unknown* dolphins. NNPool is based on the photo-identification of $n$ individuals photo-identified, a number grater than the number $m$ of individuals used in RUSPool. However, the selection of the subset of $m$ dolphins for RUSPool is internal to the methodology and should be considered as a part of it. For this reason, the study of comparison between RUSPool and NNPool-based strategies is valid, both starting from the $n$ dolphins listed in DolFin. Finally, the performance of NNPool was compared to that of RUSPool, and experimental results were also validated using a further data set, collecting Risso's dolphins images acquired in 2019 from the Northern Ionian Sea and Azores.

In summary, the main contributions of the paper are: the development of NNPool, a new methodology devoted to the automated photo-identification of *unknown vs known* Risso's dolphins; a comparison study of the performances of CNN against RUSBoost algorithm for imbalanced data classifi-

**IEEE** *Access*



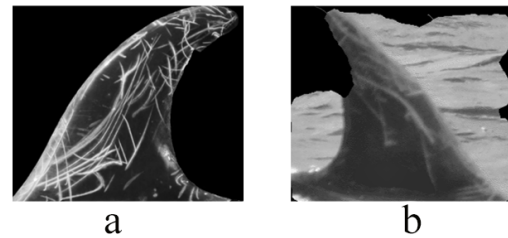FIGURE 1. Risso's dolphins *Grampus griseus*.



FIGURE 2. On the panel a (b), fin mask correctly (wrongly) computed is shown. The filter $f_R$ will discard the image b.

cation; the publication of the updated version of DolFin, a freely-accessible catalogue collecting photos and data of Risso's dolphins in the Gulf of Taranto.

## MATERIALS AND METHODS

### A. SURVEY AREA AND DATA COLLECTION

Sighting data of *G. griseus* were collected from July 2013 to August 2018 during vessel-based surveys conducted on board of a 12-m catamaran, investigating an area of about 960 km2, in the Gulf of Taranto (Northern Ionian Sea, Central-eastern Mediterranean Sea). Date, daytime, sea weather conditions, geographic coordinates, group size (number of specimens), and depth (m) were recorded. In addition, a collection of images for photo-ID were taken using a Nikon D3300 camera with Nikon AF-P Nikkor 70-300 mm, f4,5-6,3G ED lens. We found 93 different Risso's dolphins, photo-identified using the algorithm SPIR [4], [5] and their photographs are freely accessible using the new release of DolFin platform (http://dolfin.ba.issia.cnr.it/). Unfortunately, most of them were sighted only once, thus very few photos are available and the quality is not always good. Hence, in the present study, only fin images with dimension equal or superior to 200 x 200 pixels were used. Moreover, dolphins were selected according to the following: 1) each individual must be sighted in two or more different daily surveys; 2) at least 15 images must be available for a fin side of dolphin. The first condition serves to guarantee a minimum of variability in the illumination conditions when acquiring images; the second condition is imposed to ensure a minimum number of examples necessary to train the classifier. In this way, a number $n$=28 of individuals has been selected in the DolFin catalogue and used for this study. The left and right side of the dolphin's fin are considered and analysed independently, as if belonging to separate individuals.

To built the data set $D_R$ used for training RUSBoost classifiers, we introduced a tailored filter $f_R$: the fin area, computed using the mask, must cover less than 70% of the total image. As shown in the figure 2a, when the fin mask is correctly computed, i.e. the mask contains all the fin pixels and no sea pixels, then the mask area will be lower than a threshold, empirically fixed at the 70% of the total number of pixels. This is a crucial point. As the SIFT features, representative for each individual, will be extracted from within the fin

mask, the fin must be correctly segmented and exclude as many sea pixels as possible. In this manner, a subset of $m$=23 dolphins has been selected out of the $n$=28 chosen for this study. The data set $D_R$ contains 433 images belonging to the $m$=23 dolphins, whose names and percentages of images available are reported in table3.

To build the data set $D_{NN}$ used to train Convolutional Neural Networks, we used a resizing of all the images equal to 300 x 400 pixels because CNN requires images of equal size. Note that the filter $f_R$ used to build $D_R$ is not necessary now because, in this case, the input data for CNN is the full image and no mask is computed over the fin. Finally, $D_{NN}$ consists of 582 images of $n$=28 different dolphins, whose name and percentage of available images are reported in table 6.

Lastly, a validation data set $D_v$ was considered to validate experimental results, it containing 300 images of Risso's dolphin fins so detailed:

- 150 collected during daily surveys in 2019 in the Gulf of Taranto, using the previously described protocol and instruments, of which 40 belong to some of the 23 *known* dolphins, and 110 to the *unknown* dolphins;
- 150 collected in Azores. This dataset was obtained off Pico island covering approximately 540 km2 between May and September 2019. Risso's dolphins were first located from a land based look out (38.4078 N and 28.1880 W) using 25x80 binoculars (Steiner observer) [26], and encountered during ocean based surveys, using a 5.8 meters long zodiac, equipped with a 50 HP outboard engine. In the present study all this photos belong to the *unknown* class.

The *unknown* label is assigned to dolphins never seen during previous surveys and for this reasons these dolphins are not already catalogued in previous photo-id studies. On the contrary, the *known* class considers dolphins already listed in the reference catalogue, i.e. photo-identified in previous studies.

### B. RUSBOOST METHODOLOGY

In this section we illustrate the methodology, previously presented in [14], based on RUSBoost classifier and developed with the aim of identifying the name of the dolphin captured in a new photo is described. This methodology developed can be summarized as shown in figure 3. The PRE-
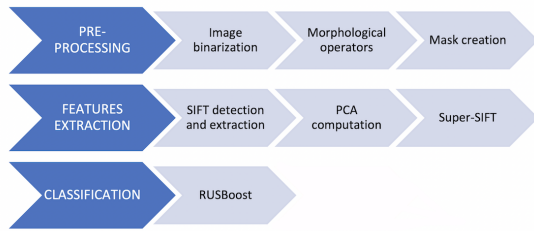
**FIGURE 3.** Experimental methodology, based on RUSBoost, implemented for the Risso's dolphins photo-identification.
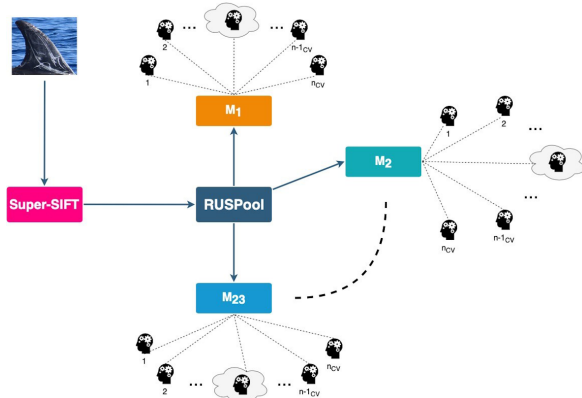


**FIGURE 4.** RUSPool method.

**TABLE 1.** Example of RUSPool output. The RUSPool output is a vector R containing $r_i$ values with i = 1, 2, ...,23, where $r_i$ is the number of positive predictions of the $M_i$ classifier, out of the $n_{CV}$ predictions.

| i | M_i | r_i (%) |
|---|---|---|
| 1 | TI_L | 99 |
| 2 | DELTA_R | 37 |
| 3 | HUGO_L | 30 |
| 4 | SVIRGOLO_R | 30 |
| 5 | SVIRGOLO_L | 25 |
| 6 | JAX_L | 23 |
| 7 | PERONI_R | 23 |
| 8 | ALT_R | 3 |
| 9 | ELE_R | 1 |
| 10 | ERARD_R | 1 |
| 11 | VITO_R | 1 |
| 12 | ZANTE_R | 1 |
| 13 | BLACK_L | 0 |
| 14 | CUPIDO_R | 0 |
| 15 | DALMATA_L | 0 |
| 16 | DUBBIO_L | 0 |
| 17 | EMME_R | 0 |
| 18 | JHONATAN_L | 0 |
| 19 | PINNA_L | 0 |
| 20 | PREZZEMOLO_L | 0 |
| 21 | PREZZEMOLO_R | 0 |
| 22 | SMILE_R | 0 |
| 23 | TRIS_L | 0 |

PROCESSING phase consists of the automated segmentation of the fin in the image, through 1) the image-binarization using Otsu's method [27], 2) application of morphological operators (opening and thinning, filling), 3) extraction of the contour from the B/W image and mask creation [4], [5]. In the FEATURE EXTRACTION step, the SIFT features are extracted in each fin image [5], and subsequently the PCA is performed and the first 3 principal components are selected to build the 3-dimensional super-SIFT descriptor for each image. The CLASSIFICATION step consists of two different processes and it is based on RUSBoost classifier, which is a boosting-based sampling algorithm designed to handle class imbalance [12], [13]. It combines Random Under-Sampling (RUS) and Adaboost [11]. RUS is a technique that randomly removes examples from the majority class until the desired balance is achieved. Firstly, one-vs-all RUSBoost algorithm is trained to classify the identity of one selected dolphin (*known* and already photo-identified, and labeled in the data set $D_R$) against "all" the other individuals, using a Cross Validation (CV) strategy. A one-vs-all RUSBoost classifier, $M_i$ with $i = 1, 2, ...23$, is built for each of the *m*=23 Risso's dolphins in $D_R$. The class "all" is made by the images of the remaining 22 dolphins in $D_R$ and simulates the *unknown* class in each model.

## C. RUSPOOL

The goal of this RUSPool (figure 4) is to recognize the *known* against *unknown* dolphins, without exactly identifying the

name of the individuals. It consists in the smart mixing of the $M_i$ RUSBoost classifiers, with i=1, 2, ..., 23, previously trained. Considering that each $M_i$ classifier is made of $n_{CV}$ trained models, the pool of RUSBoost trained classifiers is made by the 23 × $n_{CV}$ models. When a new image is presented in input to RUSPool, the super-SIFT are computed over the image, and this novel specimen is given as input to the pool of classifiers. Firstly, among the $n_{CV}$ models trained for each $M_i$ classifier, only those with $|score(1) - score(0)| > 0.11$ are taken into account, where score(1) and score (0) are respectively the posterior probabilities that the selected specimen belong to the *known* (1) or *unknown* (0) classes, and the cut-off of 0.11 has been empirically chosen. Then, the RUSPool output is a vector $R$ as the one shown in table 1, where each element $r_i \in R$ is the number of positive prediction for each $M_i$ classifier.

A filter (figure 5) has been developed over the vector $R$ to obtain the prediction of the identity of the dolphin in the new image, choosing between *known* or *unknown* label. The *strongest model* is the $M_i$ classifier with the highest $r_i$ value, named as $r_{strong}$. A contour of $r_{strong}$ is defined as $I = (\frac{n_{CV}}{2} + 1) - (\frac{2}{3} n_{CV})$, then if no classifier exists with $r_i$ contained in this contour (meaning that no $M_i$ classifiers has $r_i > r_{strong} - I$), the specimen was classified as known, otherwise the specimen was classified as *unknown*.

## D. CONVOLUTIONAL NEURAL NETWORK

Deep Learning is a type of machine learning in which a model learns how to perform classification tasks directly from images. This is advantageous because feature extraction is internal to the learning step, whilst RUSBoost required an explicit Super-SIFT features computation.

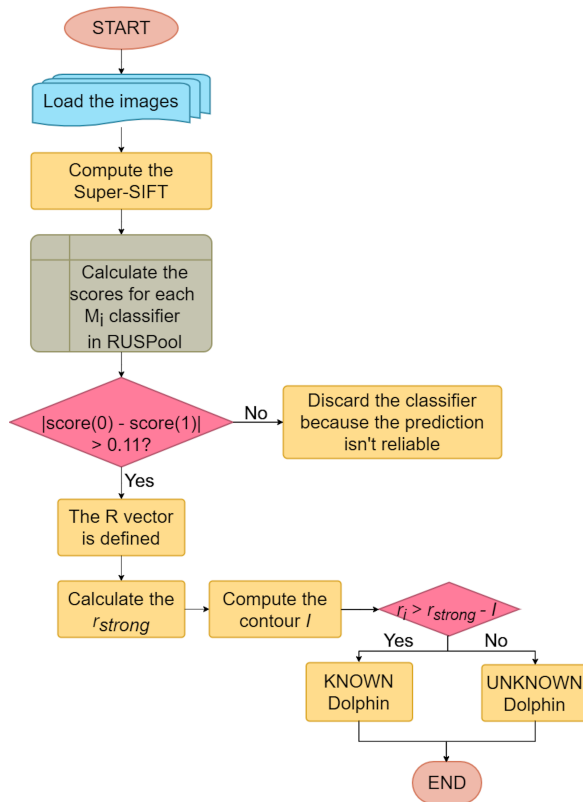In this paper we use one of the most popular algorithms for

**IEEE** *Access*



**FIGURE 5.** Flowchart of RUSPool filter.

Deep Learning, the Convolutional Neural Network (CNN) [17]. This algorithm can have dozens or hundreds of layers, each able to detect different features of an image. Like other neural networks, a CNN is composed of an input layer, an output layer, and many hidden layers in between. These layers perform operations that alter the data with the intent of learning features specific to the data. The most common layers are:

- Convolutional (Conv) which puts the input images through a set of convolutional filters, each activating a certain feature from the images.
- Rectified linear unit (ReLu) which allows faster and more effective training by mapping negative values to zero and maintaining positive values, thus adding non-linearity.
- Pooling (MaxPool) which simplifies the output by performing nonlinear down-sampling, reducing the number of parameters that the network needs to learn.
- Fully Connected (FC) which multiplies the input by a weight tensor and adds a bias vector. All neurones are connected together and typically, these kinds of layers are used to actually classify the extracted features from previous layers.
- SoftMax which allows to transform a set of values into probabilities associated to the classes.

By mixing these three types of layers, we obtain the architecture of the CNN we have built. In order to recognize the fin image, we apply three times the following combination of layers: Conv-ReLu-MaxPool, each time doubling the learned filters number on the convolutional layers from 8 to 16 to 32 (see figure 6).

After learning features the architecture of a CNN shifts to classification which is composed by two (FC + ReLu) layers, used to reduce the dimension to a k-dimension vector, where $k = 2$ is the number of classes, *known vs unknown*, to predict. This vector contains the probabilities that the input image belongs to each class. The final (FC3 + SoftMax + Classification) layers of the CNN use a SoftMax function to provide the classification output [28].

In order to better justify the CNN architectural choice, it is worth providing an overview of the number of learnable parameters involved in the training of the CNN proposed in this paper while comparing it to the homologous number for one of the CNN classifiers most known in literature. Our CNN requires $7.248.274$ parameters to be learned, compared to AlexNet (for example) that requires the update of $60.965.225$ parameters. The number of images required to train our classifier avoiding the overfitting is lower than the one required from other bigger networks. The straightforwardness of our architecture is also justified by the task that the network solves, i.e. the classification of dolphin individuals by analysing dorsal fins images. In fact, even if this kind of images can be different one from the other, the intrinsic variability of our dataset is lower when compared to other datasets used to train, for example, AlexNet, which aims to discriminate a generic input image among $1.000$ different classes. For these reasons, the number of images collected in our experiments is sufficient and has also been properly augmented and used to train a simple but effective CNN built from scratch.

A CNN model is then trained for each of the $n=28$ dolphins in the $D_{NN}$ data set with a one-vs-all technique and a CV strategy. We know that overfitting can be an issue that needs to be avoided at all costs. For this reason, we decided to execute an additional validation experiment to test the performance of NNPool on a dataset never used for training purposes. The results confirm the capability of the model to perform the classification task without extensively drop the performance.

Class imbalance is managed with a downsampling strategy. Given a dolphin $d_i$ among the $n=28$ dolphins in $D_{NN}$, let $n_i$ be the number of images available for $d_i$. Then, the *unknown* class is composed by $m_i$ images, where $m_i = 27 \times \kappa^*$ with $\kappa^* = min\{\kappa \in \mathbb{N} \,|\, 27 \times \kappa \geq n_i\}$, i.e. the first multiple of 27 greater than $n_i$. This way, $\kappa$ images for each of the remaining 27 individuals are taken into account. The training set is composed of $(n_i + m_i)$ photos. Subsequently, an oversampling technique based on image augmentation is used to increase the number of images in the training set. Two geometric transformations have been applied to every image: random rotation of $\pm 45$ degrees (over both y and x axes) and translation of $\pm 20$ pixels (over both axes, too).

**TABLE 2.** Example of NNPool output. The NNPool output is a vector P containing $p_i$ values with i = 1, 2, ..., 28, where $p_i$ is the probabilities of the fin in the photo belongs to the dolphin reported to the CNN$_i$ column. * refers to dolphins which were not analyzed by RUSBoost.

| i | CNN_i | p_i (%) |
|---|---|---|
| 1 | TI_L | 90 |
| 2 | BLACK_L | 50 |
| 3 | PINNA_L | 33 |
| 4 | * ZANTE_L | 30 |
| 5 | DALMATA_L | 27 |
| 6 | DUBBIO_L | 25 |
| 7 | * CUPIDO_L | 12 |
| 8 | PERONI_R | 12 |
| 9 | TRIS_L | 10 |
| 10 | * FRANGETTA_R | 8 |
| 11 | PREZZEMOLO_L | 7 |
| 12 | PREZZEMOLO_R | 6 |
| 13 | SVIRGOLO_L | 5 |
| 14 | * CARL_R | 3 |
| 15 | DELTA_R | 3 |
| 16 | SVIRGOLO_R | 1 |
| 17 | VITO_R | 1 |
| 18 | ERARD_R | 0,8 |
| 19 | EMME_R | 0,5 |
| 20 | ALT_R | 0,2 |
| 21 | CUPIDO_R | 0,2 |
| 22 | JHONATAN_L | 0,2 |
| 23 | * MENO_R | 0,1 |
| 24 | ELE_R | 0 |
| 25 | HUGO_L | 0 |
| 26 | SMILE_R | 0 |
| 27 | JAX_L | 0 |
| 28 | ZANTE_R | 0 |

### E. NNPOOL

Here we propose a new strategy, named NNPool, whose aim is to recognize *known vs unknown* dolphins without explicitly identifying the name of the known and previously labeled dolphin. Similarly to RUSPool, NNPool consists in the mixing of the CNN$_i$ networks, with i=1, 2, ..., 28, where CNN$_i$ is made of n$_{CV}$ trained models. NNPool is made of the (28 × n$_{CV}$) models.

Each time we want to predict the label of a new photo, it will be resized to the required dimension by the input layer of all the CNNs, which is 300x400 pixels. Subsequently, the photo can be used as input in NNPool, giving a $P$ vector as output (as shown in table 2). If there is only one $p_i \in P > 51\%$, the new photo will be labeled as *known*, otherwise it will be labeled as *unknown*.

### F. EVALUATION METRICS

The statistical measures used to evaluate the performances of the classifier are:

- accuracy = (TP+TN)/(TP+FN+FP+TN)
  percentage of predictions that are correct
- sensitivity = TP/(TP+FN)
  percentage of positive labeled instances that were predicted as positive
- specificity = TN/(FP+TN)
  percentage of negative labeled instances that were predicted as negative.

True Positive (TP) is the number of the actual positive data that are correctly classified; False Positive (FP) is the number of negative data classified as positive; True Negative (TN) is the number of the actual negative data that are correctly classified; False Negative (FN) is the number of positive data classified as negative. In our experiment the negative label N corresponds to the *unknown* class. The intuitive meaning of each measure is also listed.

## II. EXPERIMENTS AND RESULTS

All data are analyzed using Matlab (MathWorks, Natick, MA) and the experiments described in this paper have been conducted on a HP z840 Workstation equipped with Intel® Xeon® CPU E5-2699 v3 @ 2.30 GHz CPU, 256 GB RAM and Nvidia Quadro K5200 graphics card. In the first step of this study a one-vs-all RUSBoost model is built to recognize a selected dolphin against all the others in the data set D$_R$. A CV technique is used to evaluate the performances of the classifiers, In each round of the CV, the data set D$_R$ is divided into training and test set, as shown in Figure 7. The training set contains about 2/3 of the data, and the remaining examples are collected in the test set [29]–[31]. The RUSBoost classifier is trained on the first set and its performances are evaluated on the test set, results are averaged over the n$_{CV}$ = 100 rounds. The number of CV was empirically set. The imbalance of the data is varying from 9.5% to 3.5% (see table 3), reaching critical values for the classification task. Hence the selection of RUSBoost, among other machine learning algorithms, is focused. The number of cycles of RUSBoost was empirically set as NLearnCycles= 60 for all the classifiers. Moreover, the image quality surely impacts on the algorithm performances and therefore should be taken into account. Several methods to evaluate image quality have been discussed in literature [32], [33]. In this paper the Perception based Image Quality Evaluator (PIQE scores) [34] is used to evaluate the image quality and is computed for all the images used to built each M$_i$ classifiers. These are no-reference image quality scores, with values in the range [0, 100], inversely correlated to the perceptual quality of an image. The quality scale of the image based on its PIQE score is reported in table 4. A low PIQE score value indicates high perceptual quality and high score value indicates low perceptual quality. The accuracies, specificities and sensitivities of RUSBoost classification over the 23 dolphins are shown in table 3, where PIQE median value and Median absolute deviation (Mad) of images are reported in the last two columns.The classifier of ERARD$_R$ is built on the highest number of images, however having fair qualities, as shown by PIQE median and Mad, the specificity is good, as it corresponds to 82%. Similarly, CUPIDO$_R$, with an imbalancing of 5.3% of fair quality images, reached the highest accuracy and specificity, equal to 93%. A sensitivity of 90% was reached in the recognition of EMME$_R$, which is represented by lower number of images (corresponding to the 4.2% of the data set size), all having good qualities. When the imbalance decreases (lower than 5.5) and the image quality
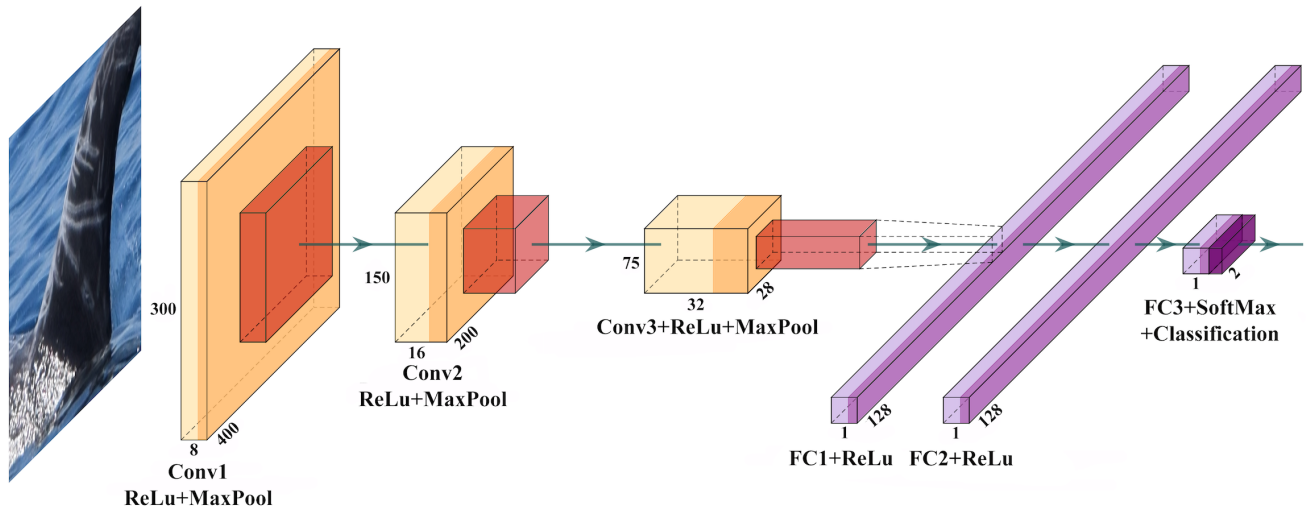
**FIGURE 6.** A graphical rapresentation of the NN build for each dolphin.

is fair or poor, the performance decreases as well, displaying low sensitivity value. A remark should be given on $ELE_R$ showing an excellent quality score, but $65\%$ of sensitivity, explainable by the $3.5\%$ class imbalance.

Hence the performances of the RUSBoost classifiers clearly depend on the number of images available for the training of the model, as well as on the quality of the images.

Successively to test the RUSPool performance, the images collected in the data set $D_v$ were used. The experimental results, shown in table 5, highlight that RUSPool is able to recognize the *unknown* dolphins with an accuracy $= 78\%$, sensitivity $= 58\%$ and specificity $= 81\%$.

CNN has been trained using the Stochastic Gradient Descent with Momentum method (momentum set to 0.9), with mini-batch dimension of (0.25 x Training set dimension), number of epochs 60 (with shuffle at every epoch) and initial learning rate of 0.00001. The loss function used was the cross entropy loss. The performances of CNN were evaluated using a CV strategy, with $n_{CV}=10$. In each round of the CV, the data set $D_{NN}$ is divided into training and test set, where the training set contains about 2/3 of the data, and the remaining examples are collected in the test set. The classifier is trained on the first set and its performances are evaluated on the test set, results are averaged over the $n_{CV}$ rounds and illustrated in table 6, showing that CNN overall behaviour outperforms that of RUSBoost, even with generally lower quality scores.

Figure 8 shows that when few images are available for the specimen (Percentage $\leq 3.5\%$), the CNN sensitivity decreases with the images quality (cyan line in the figure). Instead, when the number of images for the specimen increases (Percentage $> 3.5\%$), good sensitivity values are achieved even if fair or poor quality images are used (magenta line in
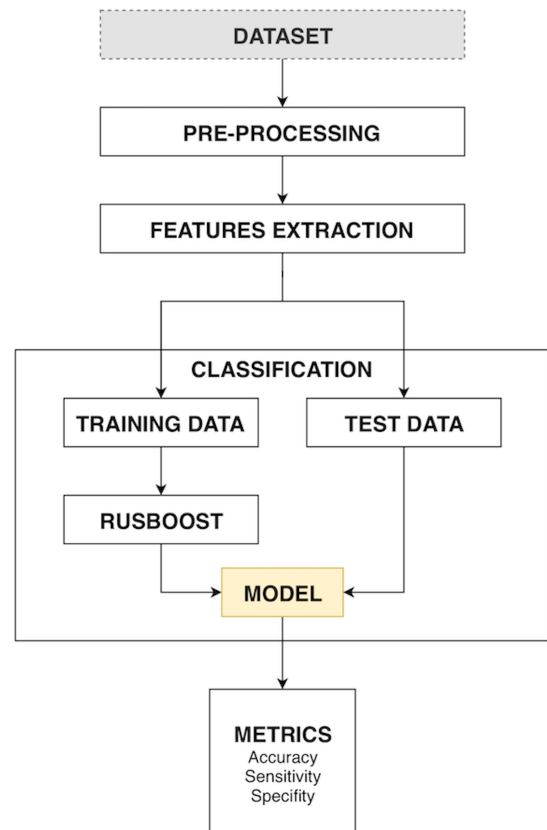


**FIGURE 7.** One round of cross validation technique employed to evaluate the RUSBoost performances.

the figure). In general, the performance of the CNNs is very

**TABLE 3.** The performances of RUSBoost (NLearnCycles= 60) in classifying the twenty-three dolphins (listed in Name) are shown in terms of accuracy, sensitivity and specificity computed using cross-validation strategy. The second column shows the percentage (%) of images of the data set available for each dolphin. Median, Mad (Median absolute deviation) and quality scale of the PIQE (Perception based Image Quality Evaluator) scores are shown in the last three columns.

| Name | Percentage | Accuracy | Sensitivity | Specificity | PIQE Median | PIQE Mad | quality scale |
|---|---|---|---|---|---|---|---|
| ERARD$_R$ | 9.5 | 0.82 | 0.81 | 0.82 | 44.14 | 3.17 | fair |
| PERONI$_R$ | 6.7 | 0.82 | 0.70 | 0.82 | 52.18 | 8.75 | poor |
| ALT$_R$ | 5,5 | 0.84 | 0.69 | 0.85 | 32.51 | 4.60 | good |
| CUPIDO$_R$ | 5.3 | 0.93 | 0.91 | 0.93 | 45.50 | 2.99 | fair |
| PREZZEMOLO$_R$ | 5.1 | 0.66 | 0.58 | 0.66 | 35.89 | 5.49 | fair |
| SVIRGOLO$_L$ | 4.6 | 0.79 | 0.71 | 0.80 | 43.87 | 7.08 | fair |
| JHONATAN$_L$ | 4.4 | 0.82 | 0.75 | 0.83 | 43.39 | 6.24 | fair |
| EMME$_R$ | 4.2 | 0.89 | 0.80 | 0.90 | 24.75 | 2.96 | good |
| SVIRGOLO$_R$ | 4.2 | 0.80 | 0.71 | 0.81 | 35.92 | 6.58 | fair |
| ZANTE$_R$ | 4.2 | 0.72 | 0.56 | 0.73 | 42.15 | 4.80 | fair |
| BLACK$_L$ | 3.9 | 0.79 | 0.63 | 0.79 | 30.05 | 9.49 | good |
| DALMATA$_L$ | 3.9 | 0.85 | 0.79 | 0.85 | 30.03 | 4.81 | good |
| DELTA$_R$ | 3.7 | 0.73 | 0.57 | 0.73 | 29.05 | 4.44 | good |
| JAX$_L$ | 3.7 | 0.74 | 0.50 | 0.75 | 40.41 | 10.75 | fair |
| DUBBIO$_L$ | 3.5 | 0.73 | 0.49 | 0.74 | 30.27 | 6.05 | good |
| ELE$_R$ | 3.5 | 0.81 | 0.65 | 0.81 | 15.28 | 1.73 | excellent |
| HUGO$_L$ | 3.5 | 0.79 | 0.64 | 0.79 | 31.29 | 5.19 | good |
| PINNA$_L$ | 3.5 | 0.71 | 0.58 | 0.71 | 35.57 | 4.56 | fair |
| PREZZEMOLO$_L$ | 3.5 | 0.76 | 0.73 | 0.76 | 39.98 | 2.98 | fair |
| SMILE$_R$ | 3.5 | 0.73 | 0.57 | 0.74 | 41.51 | 5.80 | fair |
| TI$_L$ | 3.5 | 0.88 | 0.80 | 0.88 | 28.55 | 5.21 | good |
| TRIS$_L$ | 3.5 | 0.93 | 0.88 | 0.93 | 32.57 | 2.59 | good |
| VITO$_R$ | 3.5 | 0.76 | 0.59 | 0.77 | 38.74 | 3.42 | fair |

**TABLE 4.** Quality scale of an image, based on its PIQE (Perception based Image Quality Evaluator) score.

| quality scale | score Range |
|---|---|
| Excellent | [0, 20] |
| Good | [21, 35] |
| Fair | [36, 50] |
| Poor | [51, 80] |
| Bad | [81, 100] |

**TABLE 5.** RUSPool and NNPool performance on the validation data set D$_v$.

| | Accuracy | Sensitivity | Specificity |
|---|---|---|---|
| RUSPool | 0.78 | 0.58 | 0.81 |
| NNPool | 0.87 | 0.70 | 0.90 |

good both in case of many low quality images as well as in case of a few high quality images.

The performances of NNPool, shown in table 5, were tested over the data set D$_v$, obtaining accuracy $= 87\%$, sensitivity $= 70\%$ and specificity $= 90\%$, higher than the values obtained with RUSPool.

A final remark should be given about the time required for the two analyzed classifiers. With this hardware configuration, the mean time for RUSPool training was 2288.53 seconds whilst NNPool required 1816.84 seconds. As far as classification time is concerned, i.e. the time required for a single unknown image to be classified by RUSPool or NNPool, the value is respectively 150 seconds for RUSPool and 25 seconds for NNPool.
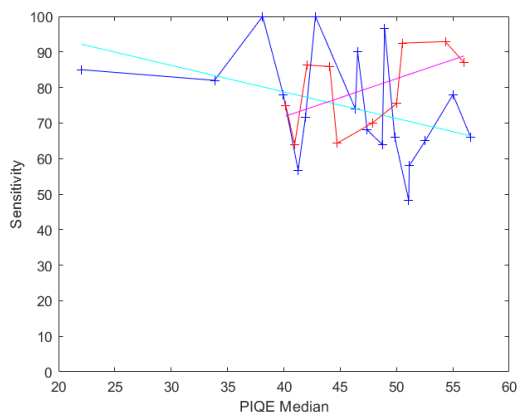
**FIGURE 8.** Plot of CNN sensitivities vs the PIQE median values (see table 6) computed on the data set $D_{NN}$. Blue (red) data refer to sensitivities of specimen with Percentage lower (greater) than $3.5\%$ (table 6). Cyan (magenta) line represents the linear trend of blue (red) data.

## III. CONCLUSION AND FUTURE WORKS

The performance of the proposed CNN-based algorithm appears, overall, satisfying, in addition to the fact that this system automatically processes large amounts of data with no interaction by the user. In particular, results highlighted by NNPool performances overcome those of RUSPool also on the validation set. Its ability to identify *unknown* Risso's dolphins, namely those dolphins never encountered during previous surveys, will open new frontiers to photo-identification studies, in particular, when paired with a photo-ID algorithm, such as SPIR [4]. An important advantage of NNPool is that it analyzes the full fin images, thus its performance is independent from the accuracy of the segmentation of the fin. And this is not a trivial task of the application at hand, keeping in mind that photos of wildlife taken on a background of sea, sun, or waves, are analyzed in the this study. Finally, it is also worth considering that the time required for classifying an unknown image with NNPool is significantly reduced if compared to the other method evaluated in this work, thus highlighting a practical advantage in terms of time complexity.

Moreover, our experimental results show that the evaluation metrics increase proportionally to the number of images available, highlighting the need to gather larger amounts of dolphins images. Thus, a strong effort is required to conduct large-scale studies on the Risso's dolphin photo-identification. Hence a further future goal will be to extend the study to larger data sets, which will be obtained by acquiring data during new surveys in our area of study, as well as by sharing data with other research groups that are offering their collaboration. DolFin portal, previously published by our research team in [4], and the updated version of which has been presented here, will facilitate the work in this regard.

## REFERENCES

[1] B. G., "Action Plan for Marine Mammals in Israel," Israel Marine Mammal Research & Assistance Center (IMMRAC), p. 101, 2017.

[2] S. Gaspari and A. Natoli, "Grampus griseus (Mediterranean subpopulation). The IUCN Red List of Threatened Species," IUCN, vol. e.T16378423A16378453, 2012.

[3] K. Hartman, "Risso's dolphin, Grampus griseus," Encyclopedia of Marine Mammals Third Edition, pp. 824–827, 2018.

[4] R. Maglietta, V. Reno, G. Cipriano, C. Fanizza, A. Milella, E. Stella, and C. R, "DolFin: an innovative digital platform for studying Risso's dolphins in the Northern Ionian Sea (North-eastern Central Mediterranean)," Scientific Reports, vol. 8, p. 17185, 2018, doi:10.1038/s41598-018-35492-3.

[5] V. Reno, G. Dimauro, G. Labate, E. Stella, C. Fanizza, G. Cipriano, R. Carlucci, and R. Maglietta, "A SIFT-based software system for the photo-identification of the Risso's dolphin," Ecological Informatics, vol. 50, pp. 95–101, 2019, doi:10.1016/j.ecoinf.2019.01.006.

[6] P. M. Panchal, S. R. Panchal, and S. K. Shah, "A comparison of SIFT and SURF," International Journal of Innovative Research in Computer and Communication Engineering, vol. 1, no. 2, pp. 323–327, 2013.

[7] M. Ambriola, R. Bellotti, M. Circella, R, and S. Stramaglia, "Supervised algorithms for particle classification by a transition radiation detector," Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, vol. 510, pp. 362–370, 2003.

[8] V. Vapnik, "An overview of statistical learning theory," IEEE Trans on Neural Networks, vol. 10, no. 5, pp. 988–999, 1999.

[9] R. Maglietta, N. Amoroso, S. Bruno, A. Chincarini, G. Frisoni, P. Inglese, S. Tangaro, A. Tateo, and R. Bellotti, "Random Forest classification for hippocampal segmentation in 3D MR images," in 12th International Conference on Machine Learning and Applications, 2013, pp. 264–267, doi:10.1109/ICMLA.2013.53.

[10] L. Breiman, "Random forests," Machine Learning, vol. 45, pp. 5–32, 2001.

[11] F. Yoav and R. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," Journal of Computer and System Sciences, vol. 55, pp. 119–139, 1997.

[12] C. Seiffert, T. Khoshgoftaar, J. V. Hulse, and A. Napolitano, "RUSBoost: a hybrid approach alleviating class imbalance," IEEE Trans. System, Man, Cybernetics Part A, vol. 40, no. 1, pp. 185–197, 2010.

[13] R. Maglietta, N. Amoroso, M. Boccardi, S. Bruno, A. Chincarini, G. Frisoni, P. Inglese, A. Redolfi, S. Tangaro, A. Tateo, R. Bellotti, and ADNI, "Automated hippocampal segmentation in 3D MRI using random undersampling with boosting algorithm," Pattern Analysis and Applications, vol. 19, p. 579, 2016, doi:10.1007/s10044-015-0492-0.

[14] R. Maglietta, A. Bruno, V. Reno, G. Dimauro, E. Stella, C. Fanizza, S. Bellomo, G. Cipriano, A. Tursi, and R. Carlucci, "The promise of machine learning in the Risso's dolphin Grampus griseus photo-identification," in IEEE International Workshop on Metrology for the Sea, 2018, doi:10.1109/MetroSea.2018.8657839.

[15] G. Bearzi, R. R. Reeves, E. Remonato, N. Pierantonio, and S. Airoldi, "Risso's dolphin Grampus griseus in the Mediterranean Sea," Mammalian Biology-Zeitschrift für Säugetierkunde, vol. 76, no. 4, pp. 385–400, 2011.

[16] I. Jolliffe, "Principal component analysis," International encyclopedia of statistical science, pp. 1094–1096, 2011.

[17] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016, http://www.deeplearningbook.org.

[18] D. Si, S. A. Moritz, J. Pfab, J. Hou, R. Cao, L. Wang, T. Wu, and J. Cheng, "Deep learning to predict protein backbone structure from high-resolution cryo-EM density maps," Scientific Reports, vol. 10, no. 1, mar 2020.

[19] A. Vahid, M. Mückschel, S. Stober, A.-K. Stock, and C. Beste, "Applying deep learning to single-trial EEG data provides evidence for complementary theories on action control," Communications Biology, vol. 3, no. 1, mar 2020.

[20] Z. Zhao, P. Zheng, S. Xu, and X. Wu, "Object detection with deep learning: A review," IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 11, pp. 3212–3232, Nov 2019.

[21] A. Hassan and A. Mahmood, "Convolutional recurrent deep learning model for sentence classification," IEEE Access, vol. 6, pp. 13 949–13 957, 2018.

[22] M. Kallenberg, K. Petersen, M. Nielsen, A. Y. Ng, P. Diao, C. Igel, C. M. Vachon, K. Holland, R. R. Winkel, N. Karssemeijer, and M. Lillholm, "Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring," IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1322–1331, May 2016.

[23] G. Dimauro, F. Girardi, M. Gelardi, V. Bevilacqua, and D. Caivano, "Rhino-Cyt: A System for Supporting the Rhinologist in the Analysis of Nasal Cytology," Intelligent Computing Theories and Application, pp. 619–630, 2018, doi:10.1007/978-3-319-95933-7_71.

[24] G. Dimauro, P. Colagrande, R. Carlucci, M. Ventura, V. Bevilacqua, and D. Caivano, "Crisprlearner: A deep learning-based system to predict crispr/cas9 sgrna on-target cleavage efficiency," Electronics, vol. 8(12), p. 1478, 2019, doi:10.3390/electronics8121478.

[25] G. Dimauro, G. Ciprandi, F. Deperte, F. Girardi, E. Ladisa, S. Latrofa, and M. Gelardi, "Nasal cytology with deep learning techniques," Int. J. Med. Inform., vol. 122, pp. 13–19, 2019, doi:10.1016/j.ijmedinf.2018.11.010.

[26] K. Hartman, F. Visser, and A. Hendriks, "Social structure of Risso's dolphins (Grampus griseus) at the Azores: a stratified community based on highly associated social units," Canadian Journal of Zoology, vol. 86, pp. 294–306, 2008.

[27] N. Otsu, "A threshold selection method from gray-level histograms," IEEE Transactions on Systems, Man, and Cybernetics, vol. 9(1), pp. 62–66, 1979.

[28] J. S. Bridle, "Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition," in Neurocomputing, F. F. Soulié and J. Hérault, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1990, pp. 227–236.

[29] N. Ancona, R. Maglietta, A. piepoli, A. D'Addabbo, R. Cotugno, M. Savino, S. Liuni, , M. Carella, G. Pesole, and F. Perri, "On the statistical assessment of classifiers using dna microarray data," BMC Bioinformatics, vol. 7, 2006.

[30] A. Bevilacqua, A. Brunetti, G. Trotta, G. Dimauro, K. Elez, V. Alberotanza, and A. Scardapane, "A novel approach for hepatocellular carcinoma detection and classification based on triphasic ct protocol," in 2017 IEEE Congress on Evolutionary Computation (CEC), 2017, pp. 1856–1863, doi:10.1109/CEC.2017.7969527.

[31] R. Maglietta, A. Milella, M. Caccia, and G. Bruzzone, "A vision-based system for robotic inspection of marine vessels," Signal, Image and Video Processing, vol. 3, pp. 471–478, 2018.

[32] G. Dimauro, "A new image quality metric based on human visual system," in Proceedings of IEEE International Conference on Virtual Environments Human-Computer Interfaces and Measurement Systems (VECIMS), 2012, pp. 69–73, doi:10.1109/VECIMS.2012.6273216.

[33] G. Dimauro, N. Altomare, and M. Scalera, "PQMET: A digital image quality metric based on human visual system," in Proceedings of 4th International Conference on Image Processing Theory, Tools and Applications (IPTA), 2014, pp. 1–6, doi:10.1109/IPTA.2014.7002002.

[34] N. Venkatanath, D. Praneeth, B. Chandrasekhar, S. Channappayya, and S. Medasani, "Blind image quality evaluation using perception based features," in Proceedings of the 21st National Conference on Communications (NCC) Piscataway, NJ: IEEE, 2015.

**ROSALIA MAGLIETTA** was born in Bari, Italy, in 1973. She received a PhD in Physics from the University of Bari 'Aldo Moro' (Italy), in 2009. Her main expertise is in artificial intelligence and her studies are currently focusing on the study and development of smart systems devoted to the conservation and protection of marine biodiversity from the threat of an ever-increasing anthropic pressure and climate change. She has been a researcher with the National Research Council of Italy (Bari), since 2007. She has co-authored more than 80 articles, published in international journals and conferences.

Dr. Maglietta is a member of the scientific committee of the Jonian Dolphin Conservation. She achieved the Best Conference Paper Award Prof. Cennamo at Workshop IEEE Metrology for the Sea 2018. Additional details: https://sites.google.com/site/liamaglietta2/home

**VITO RENÒ** was born in Bari, Italy, in 1988. He received the Master's Degree in Computer Science Engineering with honors from Polytechnic of Bari (Italy), in 2011. He also received his PhD in Electrical and Information Engineering from the same Polytechnic in 2017. His main expertise is in the development of key enabling technologies in the fields of computer vision and pattern recognition. He started working in 2011 and currently is a researcher at the National Research Council of Italy (Bari). He is co-author of 20+ scientific papers and one international patent.

Dr. Renò achieved the Best Conference Paper Award Prof. Cennamo at Workshop IEEE Metrology for the Sea 2018.

**ROCCO CACCIOPPOLI** was born in Bari, Italy, in 1995. Following the internship at the National Research Council of Italy (Bari) he will complete a Bachelor's Degree in Computer Science in 2020 at the University of Bari 'Aldo Moro' (Italy), defending a thesis on study and development of deep learning for the photo-identification of marine mammals.

**EMANUELE SELLER** was born in Bari, Italy, in 1995. He received a Bachelor's Degree in Computer Science and Technologies for Software Production from the University of Bari 'Aldo Moro' (Italy) in 2019. He is currently attending a first level Master's Degree at the University of Verona, Italy. During his internship (started in April 2019) at the National Research Council of Italy (Bari) he worked on the development of an intelligent system in the field of ecological informatics, for the automatic-photo identification of dolphins. His research interests include computer vision and automatic identification systems.

**STEFANO BELLOMO** was born in Bari, Italy, in 1988. He earned a Master of Science in Environmental Biology at University of Bari 'Aldo Moro' (Italy) in 2019. He has been collaborating with the Jonian Dolphin Conservation in Taranto since 2013 as marine mammals observer and, nowadays, he is currently a member of the scientific committee of the Jonian Dolphin Conservation. He acts as international education projects coordinator. His research interests are mainly related to ethology and photo-identification. He has co-authored 5 articles published in international conferences

Mr. Bellomo is member of the Italian Society of Marine Biology since February 2020.

**FRANCESCA CORNELIA SANTACESARIA** was born on December 7th 1992, in Bari. She earned a Master of Science in cellular and molecular biology in 2016 from University of Bari 'Aldo Moro' (Italy), where in 2019 she attended an advanced Master course in research communication and science education. She focused her studies on monitoring population dynamics, social structure and residency pattern of small cetaceans based on photo identification method. She began working as researcher at Jonian Dolphin Conservation since 2018. She has co-authored 5 articles published in international conferences.

**ROBERTO COLELLA** was born in Bari, Italy, in 1973. He graduated in Physics at the University of Bari 'Aldo Moro' (Italy) in 2003. His main interests include the development of experimental setups and the automation of measurement processes. He has been a researcher with the National Research Council of Italy (Bari), since 2019 and has co-authored 10 articles published in international journals and conferences.

**GIULIA CIPRIANO** was born in Putignano, Italy, in 1987. She received her PhD degree in Environmental Science in 2019 and recently, she became researcher in Ecology at the Department of Biology, University of Bari 'Aldo Moro' (Italy). Her main research interests focus on the application of biological, statistical and mathematical models to marine ecology and population dynamics focused on cetacean. She shows expertise in the implementation and analysis of biological/environmental data in GIS environment. She has co-authored more than 30 scientific publications in congress proceedings, national and international ISI journals.

Dr. Cipriano achieved the Best Conference Paper Award Prof. Cennamo at Workshop IEEE Metrology for the Sea 2018.

**ETTORE STELLA** was born in Bari, Italy, in 1960. He received a Laurea degree in computer science from the University of Bari 'Aldo Moro' (Italy) in 1984. His professional experiences are on computer vision, robotics, industrial automation and high-perfomance computing. He has been working for the National Research Council of Italy (Bari) since 1987, where he has been research director since 2010. He was contract professor at University of Basilicata (Italy) for the telecommunications engineering faculty in the period 2004-2009. He is co-author of more than 150 papers on international journals, on book chapter, and on proceedings of international conferences. He was co-author of international patents.

**KARIN HARTMAN** was born in Amsterdam, The Netherlands in 1974. She received a PhD in Biology from the University of the Azores, Ponta Delgada, Portugal, in 2015. She co-founded the Nova Atlantis Foundation in 2002, aiming to study and protect cetacean species in the Azores. She is studying Risso's dolphins for over 20 years, focusing on the social ecology and behaviour. At present, she is field director of the Nova Atlantis Foundation and has built up the largest database and photo-identification catalogue of Risso's dolphins (n=1250 individuals) in the world. In 2018, Karin was invited to write the chapter on Risso's dolphins for the Encyclopaedia of Marine Mammals. She is co-author on +50 international communications at conferences and papers.

Dr. Karin Hartman is the chair and co-founder of the International Grampus Workgroup, a collective of 40 scientists aiming to collaborate and share data of Risso's dolphins. In 2018, she produced the film "Scars, Politics in the Big Blue", a film that won several awards for best drone and wildlife footage in international film festivals.

**CARMELO FANIZZA** was born in Taranto, Italy, in 1976. He received a bachelor's degree in Mariculture Sciences from the University of Bari 'Aldo Moro' (Italy) in 2001. His main expertise is in scientific research on cetaceans, along with cetacean-sighting expeditions aimed at directly involving students, tourists and citizens. He founded the Jonian Dolphin Conservation in 2009 (www.joniandolphin.it), where he currently acts as President, and participated to the realization of the Euro-Mediterranean Center of the Sea and Cetaceans, Ketos, inaugurated in Taranto, Italy, in 2019. Ketos hosts a museum, an area dedicated to tourist services, a sector specifically set aside for start-ups and social entrepreneurship, and a VR library overlooking the sea. He has co-authored more than 20 articles published in international conferences and journals.

Dr. Fanizza and the Jonian Dolphin Conservation were recognized as one of the 21 Italian Excellence organizations at Expo Milan in 2015.

**GIOVANNI DIMAURO** was born in Taranto, Italy, in 1964. He received a Laurea degree in computer science from the University of Bari 'Aldo Moro' (Italy) in 1987, where he is currently an Associate Professor with the Computer Science Department. He teaches computer programming and multimedia systems. His research interests include e-health, multimedia systems, and pattern recognition with applications in medicine, such as new diagnosis technology for anemia and Parkinson's disease. He has authored more than 150 articles, which have been published in scientific journals, proceedings, and books. He is the author of 2 patents, of which the latter pertains to the field of noninvasive anemia estimation. Additional details: http://www.di.uniba.it/ dimauro/.

**ROBERTO CARLUCCI** was born in Taranto, Italy, in 1970. He holds a PhD in Environmental Science and he is Assistant Professor in Ecology at the Department of Biology, University of Bari 'Aldo Moro' (Italy). His main research interests focus on the application of biological, statistical and mathematical models to marine ecology, population dynamics and fishery stock assessment. He has co-authored 105 scientific publications in congress proceedings, national and international ISI journals and he is a reviewer for Italian and international ISI journals.

Prof. Carlucci attended the Working Group on Fishery Stock Assessment of Demersal Species at FAO General Fisheries Commission for the Mediterranean as responsible of the GSA19 (Western Ionian Sea). He is the scientific responsible of the Jonian Dolphin Conservation onlus. He is member of the Commission Internationale pour l'Exploration Scientifique de la Méditerranée Task Force on Sharks and Rays.

● ● ●

**TABLE 6.** The performances of CNNs in classifying the twenty-six dolphins (listed in Name) are shown in terms of accuracy, sensitivity and specificity computed using cross-validation strategy. The second column shows the percentage (%) of images of the data set available for each dolphin. Median, Mad (Median absolute deviation) and quality scale of the PIQE (Perception based Image Quality Evaluator) scores are shown in the last three columns. * refers to dolphins which were not analyzed by RUSBoost.

| Name | Percentage | Accuracy | Sensitivity | Specificity | PIQE Median | PIQE Mad | quality scale |
|---|---|---|---|---|---|---|---|
| $ERARD_R$ | 8.8 | 0.89 | 0.86 | 0.89 | 44.07 | 9.86 | fair |
| $ALT_R$ | 7,5 | 0.84 | 0.64 | 0.86 | 40.96 | 8.60 | fair |
| $PERONI_R$ | 5.7 | 0.82 | 0.70 | 0.83 | 47.85 | 5.05 | fair |
| $JHONATAN_L$ | 5.3 | 0.94 | 0.75 | 0.95 | 50.01 | 8.00 | poor |
| $PREZZEMOLO_R$ | 5.2 | 0.82 | 0.75 | 0.83 | 40.16 | 9.32 | fair |
| $CUPIDO_R$ | 4.9 | 0.89 | 0.87 | 0.90 | 55.99 | 6.71 | poor |
| $SVIRGOLO_L$ | 4.3 | 0.88 | 0.92 | 0.88 | 50.53 | 7.67 | poor |
| * $CARL_R$ | 3.9 | 0.89 | 0.86 | 0.90 | 42.05 | 7.26 | fair |
| $ZANTE_R$ | 3.8 | 0.86 | 0.64 | 0.87 | 44.73 | 9.67 | fair |
| $SVIRGOLO_R$ | 3.6 | 0.83 | 0.92 | 0.83 | 54.33 | 5.37 | poor |
| $BLACK_L$ | 3.3 | 0.89 | 0.65 | 0.91 | 52.55 | 6.64 | poor |
| $SMILE_R$ | 3.3 | 0.86 | 0.72 | 0.86 | 41.92 | 4.93 | fair |
| $DELTA_R$ | 3.1 | 0.77 | 0.57 | 0.78 | 41.27 | 6.26 | fair |
| $EMME_R$ | 3.1 | 0.86 | 0.85 | 0.86 | 22.04 | 7.59 | good |
| $VITO_R$ | 3.1 | 0.75 | 0.90 | 0.75 | 48.94 | 8.90 | fair |
| * $FRANGETTA_R$ | 3.1 | 0.90 | 0.97 | 0.91 | 38.11 | 8.67 | fair |
| * $MENO_R$ | 3.1 | 0.91 | 1 | 0.91 | 46.55 | 6.68 | fair |
| $DALMATA_L$ | 3.0 | 0.86 | 0.49 | 0.88 | 51.08 | 5.79 | poor |
| $PINNA_L$ | 2.8 | 0.86 | 0.58 | 0.87 | 51.13 | 6.84 | poor |
| $TI_L$ | 2.8 | 0.88 | 0.78 | 0.88 | 55.01 | 3.41 | poor |
| $TRIS_L$ | 2.8 | 0.88 | 1 | 0.88 | 42.80 | 4.97 | fair |
| $HUGO_L$ | 2.6 | 0.87 | 0.64 | 0.87 | 48.73 | 4.82 | fair |
| $JAX_L$ | 2.6 | 0.90 | 0.66 | 0.91 | 56.58 | 6.34 | poor |
| * $CUPIDO_L$ | 2.6 | 0.88 | 0.66 | 0.89 | 49.84 | 5.87 | fair |
| $DUBBIO_L$ | 2.6 | 0.89 | 0.68 | 0.89 | 47.40 | 6.71 | fair |
| $PREZZEMOLO_L$ | 2.6 | 0.91 | 0.74 | 0.92 | 39.95 | 4.41 | fair |
| $ELE_R$ | 2.6 | 0.92 | 0.78 | 0.93 | 46.34 | 7.49 | fair |
| * $ZANTE_L$ | 2.6 | 0.93 | 0.82 | 0.94 | 33.85 | 7.89 | good |