Detecting the presence-absence of bluefin tuna by automated analysis of medium-range sonars on fishing vessels

Jon Uranga[1]*, **Haritz Arrizabalaga**[1], **Guillermo Boyra**[1], **Maria Carmen Hernandez**[2,3], **Nicolas Goñi**[1], **Igor Arregui**[1], **Jose A. Fernandes**[4], **Yosu Yurramendi**[2], **Josu Santiago**[1]

1 AZTI-Tecnalia, Marine Research Division, Pasaia, Spain, **2** Computer Science and Artificial Intelligence Department, University of the Basque Country, Donostia, Spain, **3** Centre for the Research and Technology of Agro-Environmental and Biological Sciences, University of Trás-os-Montes and Alto Douro, Vila Real, Portugal, **4** Plymouth Marine Laboratory, Plymouth, United Kingdom

* juranga.research@gmail.com

Abstract

This study presents a methodology for the automated analysis of commercial medium-range sonar signals for detecting presence/absence of bluefin tuna (*Tunnus thynnus*) in the Bay of Biscay. The approach uses image processing techniques to analyze sonar screenshots. For each sonar image we extracted measurable regions and analyzed their characteristics. Scientific data was used to classify each region into a class ("tuna" or "no-tuna") and build a dataset to train and evaluate classification models by using supervised learning. The methodology performed well when validated with commercial sonar screenshots, and has the potential to automatically analyze high volumes of data at a low cost. This represents a first milestone towards the development of acoustic, fishery-independent indices of abundance for bluefin tuna in the Bay of Biscay. Future research lines and additional alternatives to inform stock assessments are also discussed.

Introduction

The Atlantic bluefin tuna (*Thunnus thynnus*) is an emblematic species exploited for several centuries that has supported economically important industrial fisheries [1]. The International Commission for the Conservation of Atlantic Tunas (ICCAT) manages two Atlantic bluefin tuna stocks, the western stock that spawns in the Gulf of Mexico, and the eastern stock that spawns in the Mediterranean. Both stocks have been overfished in recent decades [2] and currently they are under recovery plans. Furthermore, the scientific community has warned about the large uncertainty surrounding the eastern stock status [3], which is being addressed with a set of research programs under the Atlantic-wide Research Programme for bluefin Tuna (GBYP) promoted by ICCAT. In order to be able to quantify the effects of the implemented recovery plan, it is of outmost importance to be able to monitor changes in abundance and stock status through accurate indicators.

Fisheries independent scientific surveys are used to monitor the stock abundance of many groundfish and small pelagics [4]. Absolute and relative stock abundance estimates are useful

to inform management of exploited fish stocks. Many of the uncertainties associated with our ability to estimate fish stock abundances can be linked directly to limitations in the spatial coverage of our sampling systems [5]. For example, in the case of scientific acoustic surveys, highly precise narrow vertical beam acoustic equipment might fail to detect aggregations if these are sparsely distributed or if fish are aggregated in the unsampled surface. In such situations, the use of commercial fishing vessels and their acoustic equipment allows for substantial increase in the spatial coverage. In fact, major progress has been made in the use of this information as the basis for stock assessment [6, 7, 8, 9], as well as to analyze fish behavior [10], vessel avoidance [11] and fish distribution [12].

In tuna stock assessments, time series of standardized catch per unit effort (CPUE) indices are used as proxies for relative abundance. However, these series, based on fishery data, have known analytical challenges, such as lack of scientific design, correlated observations, non-random sampling or variable catchability [13], and do not necessarily reflect trends in population abundance. In the case of bluefin tuna, the drastic reduction in fishing opportunities as part of the recovery plan has affected the CPUE indices, and the Standing Committee on Research and Statistics (SCRS) of ICCAT has recommended urgently developing fisheries independent indices of abundance [14].

There are very few fishery-independent surveys for tuna, and other highly mobile species with wide distributional ranges, because the cost associated with research vessels covering the whole distribution area is prohibitive. Moreover, it is not possible to account for the uncertainty associated with this type of surveying (e.g. double counting). Therefore, some fishery independent surveys for tuna have focused on early life stages (larvae) or spawners whose distributional range is much more concise and spatially limited to spawning areas [15]. When the focus has been on juveniles and adults (with high migration capabilities) airplanes have been used instead of research vessels to provide broad distribution coverage in reasonable timeframes and with reasonable costs [16, 17], estimating the approximate horizontal shape of the visible portion of schools [18]. Some sonar and echosounder-based acoustic surveys have also been implemented to monitor southern bluefin tuna recruitment [19], together with trolling transects surveys [20].

The standardized CPUE of the Bay of Biscay baitboat fleet is used as the only abundance index for the juvenile fraction of the entire eastern stock [21, 22]. Catchability by baitboats can be affected by several factors including food availability, feeding behavior and stomach repletion [23, 24]. These variables are difficult to incorporate during the CPUE standardization process. Consequently, inter-annual variability could induce bias in the abundance indices (e.g. a large tuna biomass could yield a low baitboat CPUE if plenty of food is available in the environment and tunas are not attracted by the bait). However, Bay of Biscay baitboats use Omni mode Medium Range Sonars (MRS) to search for tuna, and omni directional sonars have proven to be useful tools for characterizing large pelagic schools [25, 26]. Thus, the information obtained by these sonars could provide data about the number and size of tuna schools in the search area, independent of food availability and feeding behavior. These sonars are analog and non-scientific, used only for display, and all the information collected is lost as soon as it disappears from the screen. Thus, our approach is to record sonar screen shots in a large number of fishing vessels during the tuna fishing campaigns and design an automated methodology for analyzing these images as a way to utilize the data currently wasted. The automated processing of images has been proven to be useful in biological studies and it is a fast-evolving area of research [27, 28, 29].

In summary, the estimation of bluefin tuna abundance in the Bay of Biscay using fishery independent methods remains challenging, but new technologies, datasets and approaches provide new opportunities to address the challenge. The main objective of this study is to

develop an automated image analysis procedure for detecting presence-absence of bluefin tuna in commercial sonar images, plus a validation of the procedure based on data mining. The utility of the procedure to track abundance of juvenile bluefin tuna in the Bay of Biscay is also discussed. This constitutes a first milestone towards the longer-term objective of developing new fishery independent indices of abundance for Atlantic bluefin tuna based on acoustics.

Materials and methods

The research presented in this manuscript involved no endangered or protected species. No experimentation with animals was performed and no specific field permits were required as the scientific observations were conducted on commercial fishing activities regulated by the International Commission for the Conservation of Atlantic Tunas (ICCAT). No other ethical issues applied to the present research project.

The study area is delimited by the activity of the baitboat fleet in the southeast corner of the Bay of Biscay, between 43–47˚N and 2–6˚W, from June to October (Fig 1). The Bay of Biscay represents a relatively small fraction of the total bluefin tuna habitat in the Atlantic [30]. However, it is the most important known feeding area for juveniles during their feeding migration to the Northeast Atlantic around summer [31].

Pole and line fishing with live bait is the traditional fishing technique used by the Basque fleet fishing for bluefin tuna in the Bay of Biscay since the early 1950s. Live bait (mainly small horse mackerel, sardine, mackerel and anchovy) is caught with a small purse seine and kept in water tanks onboard. Tuna schools can be spotted visually at large distances and then detected acoustically by sonar, once the school is within the detection range of the sonars. When the boat is close to the tuna school, live bait is thrown into the water to keep the tuna next to the boat, while the boat sprays water so that it is not seen by the tuna. At this point, baited hooks are used to catch the tuna.

In this study we created a reference dataset of sonar images with known categories ("tuna" or "no tuna", based on tuna presence and absence data observed by scientists) to validate an image analysis and classification procedure. This dataset was used to test the methodology developed in this study which consists of several steps: 1) Image acquisition and categorization based on scientific data, 2) Features extraction, 3) Training dataset elaboration and 4) Model training and evaluation.

Image acquisition and categorization based on scientific data

The images processed in this study were obtained from the commercial sonar MAQ 90 kHz. This omni-directional MRS is used by the majority of the Bay of Biscay baitboat fleet. The searching range of the sonar varies with sea conditions and skipper preferences but, in general, range settings of 100–300 m are used when searching for tuna, with a slight tilt of minus 5–7˚ off the horizontal and narrow vertical and horizontal beam widths (5˚).

The screen dumps were acquired using an image acquisition device composed of 400MHz video splitter, an external VGA Capture Device and a laptop with a script for continuous data acquisition. The images selected for this study correspond to six different trips from two scientific tuna surveys conducted in summer 2009 and 2011. The scientific surveys were conducted using a baitboat that behaved similar to the rest of the commercial baitboat fleet. Thus, the area searched during the scientific surveys significantly overlapped the fishing area used by the commercial fleet (Fig 1). The main activities conducted by the scientists during the surveys were characterization of the vessel activities, recording of MAQ sonar screenshots and SIMRAD EK 60 signal, tuna tagging and biological sampling (length measurements as well as collection of genetic tissue). The presence of bluefin tuna in the sonar was validated when bluefin
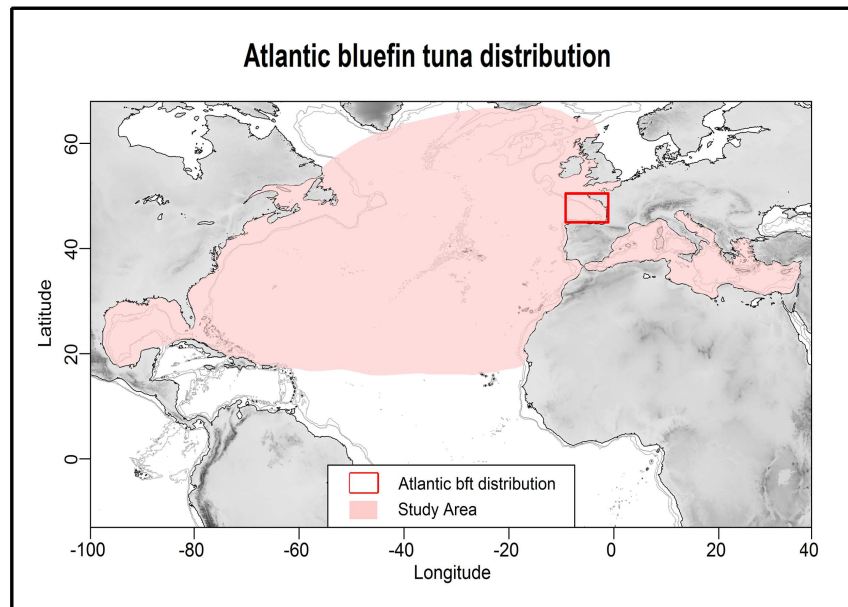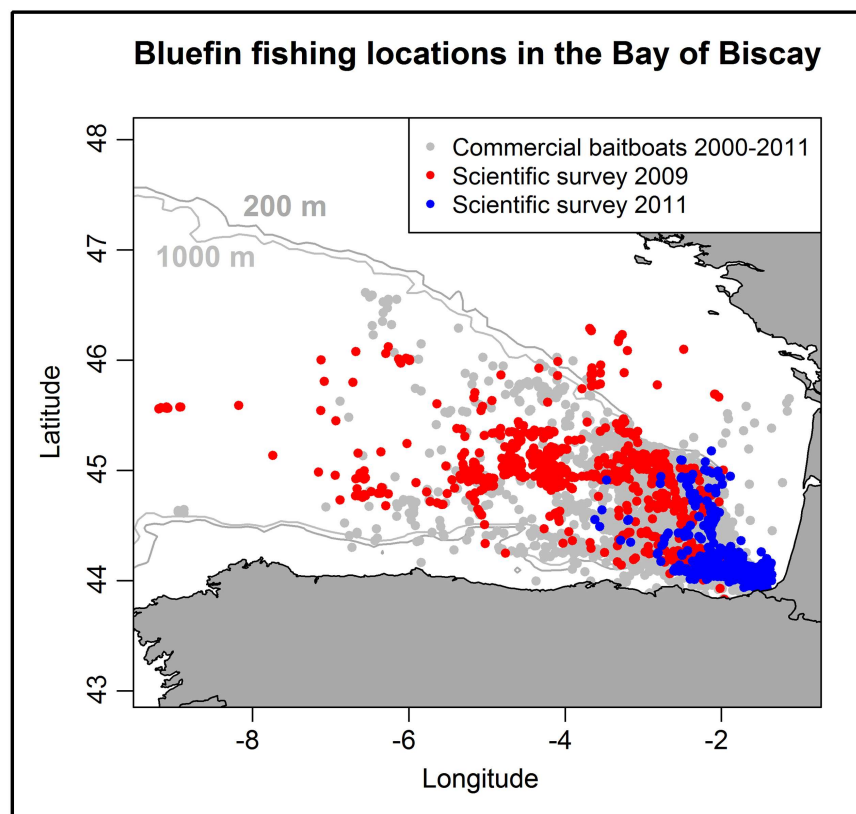
Fig 1. The study area. A) Atlantic bluefin tuna distribution based on ICCAT catch data for the period 2000–2013 [14]. B) The study area, bluefin tuna fishing locations based on logbook data [22] and scientific surveys conducted in 2009 and 2011.

doi:10.1371/journal.pone.0171382.g001

tuna was the only specie caught during fishing operations. Presence of bluefin tuna was annotated in the scientific logbooks, and this information was used to classify the images under "tuna" and "no tuna" categories. For this study, the reference dataset was built by selecting a balanced set of images, with 1397 images of bluefin tuna presence and 1398 images of bluefin tuna absence. Bluefin tuna absence was defined as lack of tuna echo in the image and lack of tuna catch. With the aim to include the main types of images recorded, the reference dataset included images with different background colors as well as images with and without surface noise (Fig 2).

Features extraction

The image processing application was developed with a Java software and it consisted of three steps: pre-processing, segmentation and extraction of characteristics.

Pre-processing. The pre-processing phase removed the non-relevant parts of the sonar screen image. The screen of the MAQ sonar has two main regions, the echogram display circle and the menu panel (Fig 2). The menu panel provides user information on the operation and system control settings whereas the echogram represents the acoustic data. During the pre-processing we divided the sonar screen into these two basic regions and then focused on the echogram. In the echogram, we worked with the upper half of the circle, as the tuna schools are not clearly detected in the lower half due to the vessel's wake. Furthermore, the schools were observed to appear first in the upper part of the echogram because the vessels move faster than the fish. The sonar display was set up in such a way that the forward observations were located at the top of the screen. Additionally, the echogram was cleaned of noise and sonar display lines and marks, such as cursor crosses, vessel tracks or range circumferences were removed from the echogram (Fig 3).

Segmentation. In the segmentation phase, the selected part of the echogram was partitioned into sub-images or blobs. First, the zero-valued (i.e., black) pixels were considered background and removed; whereas the non-zero (i.e., colored) pixels were grouped, using the 8 adjacency rule, into blobs. Then, in order to reduce the size of the training dataset, the blobs containing less than 100 pixels were removed. We believe that this decision is conservative since the smallest tuna school observed by expert judgement in the reference dataset contained 415 pixels, and so it does not restrict the utility of the classification algorithm developed.

Extraction of characteristics. The remaining blobs were considered tuna candidates and were subject to a characteristics extraction process. For each one, 20 morphologic characteristics were measured related to area, perimeter, position, smallest rectangle containing the blob, best ellipse fitting the blob, aspect ratio, circularity, solidity, greatest distance between any pair of pixels of the blob (known as Feret or Feret's diameter), the projections of Feret's diameter on the axes, the angle of Feret's diameter with respect to the horizontal axis and the minimum value of the Feret's diameter. Finally, the blobs were labeled with two possible categories: "tuna" and "no-tuna", according to scientific observations.

Training dataset elaboration

Based on the reference images, a training dataset of blobs was created to train automatic classification programs and to test their efficiency before they were used to classify new unsupervised images (e.g. those collected onboard commercial fishing vessels without an observer
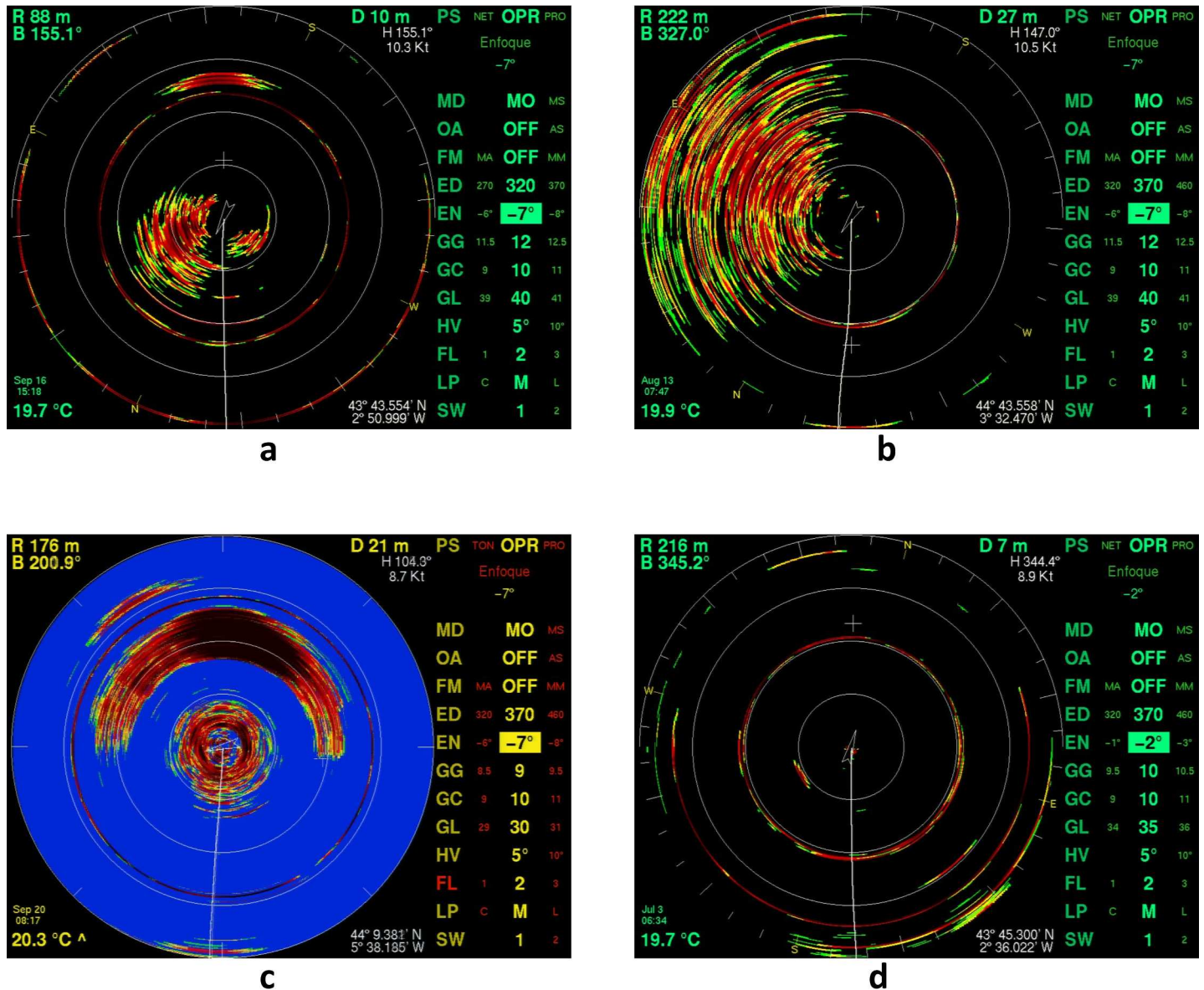
Fig 2. Main types of images recorded. Typical cases of echograms: tuna, black background (a); no tuna, noise (b); tuna, blue background (c) and no tuna (d).

doi:10.1371/journal.pone.0171382.g002

onboard). The training dataset included the categories "tuna" (presence) and "no-tuna" (absence), and is available as S1 Dataset.

From the 1397 presence and 1398 absence images in the reference images, after the features extraction, we obtained 22501 blobs for constructing the training dataset: 1497 were positive examples (presence) and 21004 were negative examples (absence), as shown in Table 1. The resulting ratio between positive/negative instances was 1/14.03, which shows that we had an unbalanced training dataset, due to the fact that images with tuna blobs also contained many other blobs that were not tuna. Subsampling and oversampling methods are available to manage unbalanced datasets [32]. For this purpose we applied a Synthetic Minority Oversampling Technique [33] to oversample the minority cases and a Spread Sample filter [34] to subsample the majority instances with the Weka software [35].
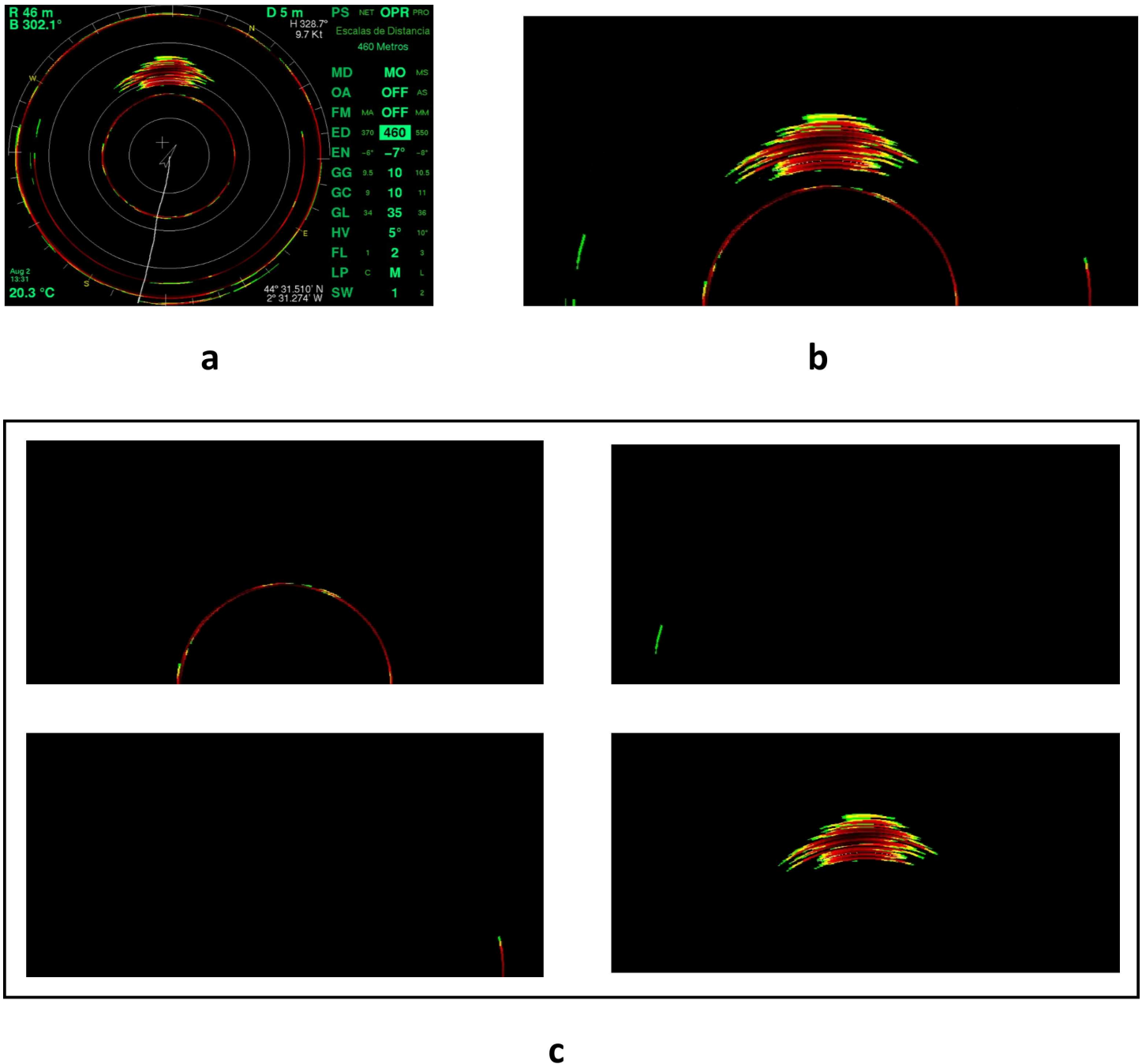
Fig 3. Image pre-processing phase. Sequential steps of the features extraction procedure: (a) original image, (b) image pre-processing, and (c) segmentation of the operative part of the echogram into "blobs".

doi:10.1371/journal.pone.0171382.g003

As a result, we constructed three training datasets: a complete dataset (TOTAL) with 22501 instances; (ii) an oversampled dataset (SMOTE) with 23998 instances; and (iii) a subsampled dataset (SPREAD) with 11999 instances, with 20 morphological characteristics in each one (Table 1).

Table 1. Ratios between presence and absence cases for the three databases: the original database (TOTAL), a subsampled database (SMOTE) and an oversampled database (SPREAD).

	Tuna	No Tuna	Ratio
TOTAL	1497	21004	14.03
SMOTE	2994	21004	7.02
SPREAD	1497	10502	7.01

doi:10.1371/journal.pone.0171382.t001

Model training and evaluation

A first experiment was performed to evaluate the merits of using only some of the 20 characteristics available in the dataset. We compared the classification performance of the reduced datasets containing a subset of characteristics with the performance of the dataset containing the whole set of characteristics. The subset of characteristics in the reduced datasets were selected using four attribute selection filters: ChiSquared, InfoGain, Support Vector Machine (SVM) and Stepwise [34]. The Stepwise method provided an optimum number of characteristics (13 in our case), while the rest of the attribute selection filters were applied at fixed numbers of characteristics (ranging between 3 and 19 in steps of 2). In each case, the attribute selection filter selected the most powerful combination of characteristics. On the reduced datasets, we applied the "five replications of two fold cross-validation" methodology (5x2cv). With this methodology, in each of the five replications, the available data were randomly partitioned into two equal sized datasets, a training dataset and a testing dataset, so that each data point had a chance of being validated. Using the Random Forest (RF) classification algorithm [36], a classification model was generated with each training dataset and validated on the testing dataset [37]. To compare their relative performance, the Kappa [38] and Accuracy values of the reduced datasets were compared to those of the complete dataset. A corrected resampled t-test was also performed to test the null hypothesis of whether the classification with the reduced dataset yielded the same accuracy as when using the complete dataset. This experiment was run under R [39], making calls to Weka software. BioSeqClass [34] and MASS [40] packages were used for this purpose.

Once the optimum number of characteristics was determined, five classification methods were applied to each of the three different datasets (TOTAL, SMOTE and SPREAD): RF [36], SVM [41, 42], Multilayer Perceptron (MLP) [43, 44], Iterative Dichotomiser 3 (J48 in WEKA) [45] and Instance-Based learner with fixed neighborhood (IBK) [46]. RF, MLP, IBK and J48 classifications were applied using Weka software and the SVM was applied using R software.

To evaluate the effectiveness and efficiency of classification methods we estimated the average validation indices for sensitivity, specificity, Kappa and Area Under the Curve (AUC). These validation indices are calculated using a confusion matrix which evaluates the predictive accuracy of presence-absence models on a set of test data for which the true values are known. The confusion matrix is defined by the true positive rate (TP, presence was correctly predicted by the model), the true negative rate (TN, absence was correctly predicted by the model), the false negative rate (FN, the model incorrectly predicted absence) and the false positive rate (FP, the model incorrectly predicted presence).

Sensitivity and specificity were calculated by the caret R package [47], as follows:

$$Sensitivity = \frac{TP}{TP + FN} \qquad \text{(Eq 1)}$$

$$Specificity = \frac{TN}{FP + TN} \qquad \text{(Eq 2)}$$

Sensitivity measures the efficiency of the algorithm in correctly classifying positive cases, and specificity measures the efficiency of the algorithm in correctly classifying negative cases.

Kappa and AUC, both are calculated by the PresenceAbsence R package [48]. Kappa is a measure of agreement between the classifications and the true classes. It's calculated as the difference between the relative observed agreements (p_o) and the relative agreements expected by chance (p_e) divided by the maximum possible agreement: It is known as the "chance-corrected proportion of agreement" [38] and it is calculated as follows:

$$Kappa = \frac{p_o - p_e}{1 - p_e}$$

(Eq 3)

AUC, is a common evaluation metric for binary classification problems and represents the area under the receiver operating characteristic (ROC) curve [49]. ROC graphs are two-dimensional graphs in which the TP rate is plotted on the Y axis and the FP rate is plotted on the X axis. It ranges between 0 and 1. When the classifier is very good, the TP rate will increase quickly and the area under the curve will be close to 1. If the classifier has a random behavior, the TP rate will increase linearly with the FP rate and the area under the curve will be close to 0.5.The scale most commonly used for model evaluation implies that a model with an AUC value of 0.95 or higher is excellent; between 0.85 and 0.95 is good; between 0.75 and 0.85 is acceptable; and below 0.75 is poor [49].

The validation indices (Kappa, Sensitivity, Specificity and AUC) were computed after executing 30 runs of the classification algorithm with 10-fold cross-validation in order to avoid overfitting and to achieve stable results [50].

Results

In the first experiment, the characteristics that were most consistently selected by the different attribute selection filters were *Area*, *Major* and *Minor* (Fig 4), which are correlated, suggesting that the size of the blob is most informative about the tuna or no-tuna category. However, both Kappa and accuracy values increased as the number of characteristics increased. The trends for both Kappa and accuracy were similar: highest gains occurred for reduced number of characteristics (up until 9), but classification performance continued to gradually improve afterwards, albeit at lower rates. Overall, none of the reduced datasets (including only a subset of the characteristics) improved the performance of the complete dataset. According to statistical t-tests, a similar performance was achieved only when a high number of characteristics were included in the reduced dataset (17 or 19 characteristics, depending on the attribute selection method, Fig 5). Thus, since our main goal was to achieve the best classification performance, and computing time was not a limiting factor, we decided to use the complete dataset (with 20 characteristics) instead of a reduced dataset.

Regarding the bluefin tuna classification study, with the original dataset (TOTAL), acceptable results were obtained for all algorithms (Fig 6). AUC values were between 0.87 and 0.97 with a difference in performance between algorithms of around 10%, such as between SVM and MLP. Sensitivity estimates varied between 0.73 and 0.79, indicating that all algorithms classify positive ("tuna") instances with similar efficiency. For specificity, all algorithms obtained very high results (> 0.95) with minor differences between them. Consequently, most negative ("no tuna") instances were correctly recognized. Kappa values also ranged from 0.74 to 0.79, thus evidencing good ratios between true positives and true negatives.

For both SPREAD and SMOTE, due to the use of more balanced datasets, the results generally improved for all the indices. This was not the case for the specificity, which showed lowest variation between datasets, and were high (> 0.95) in all instances.
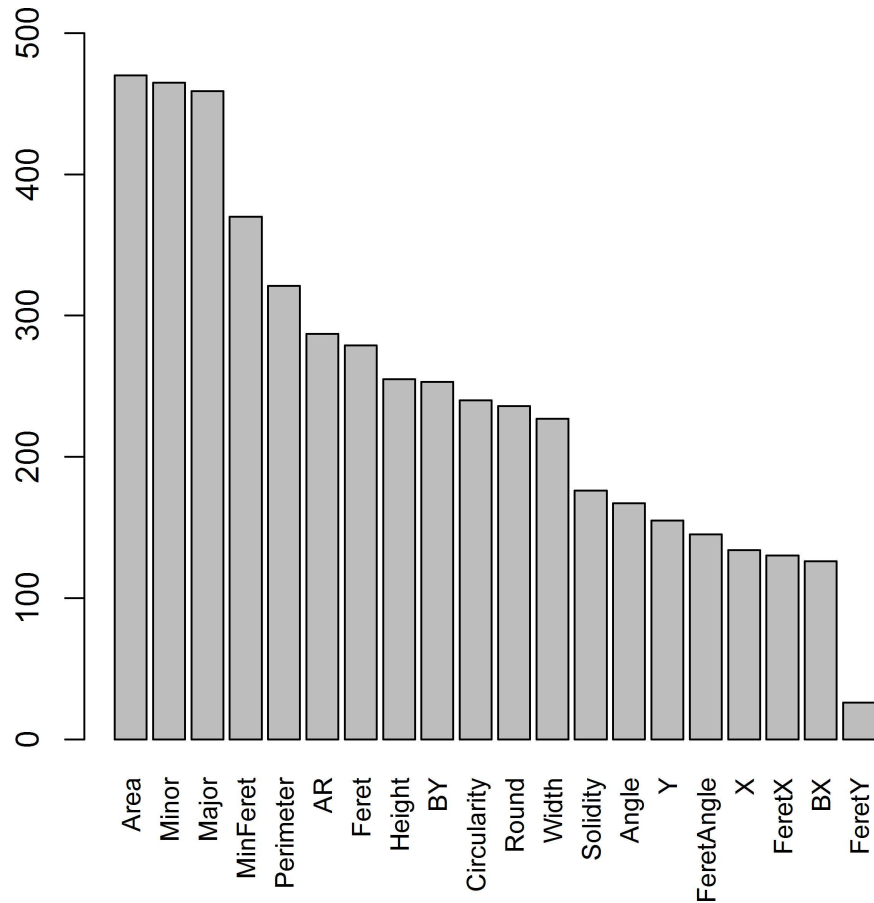
Fig 4. Frequency with which the blob characteristic were selected by the different attribute selection filters during the experiment to evaluate the merits of using reduced datasets.

With the SPREAD dataset the performance of the different algorithms (as measured by AUC and sensitivity) improved with respect to the TOTAL dataset. AUC estimates ranged between 0.90 and 0.98 and were higher than with the TOTAL dataset in all instances. Sensitivity values varied from 0.82 to 0.86 and the lowest value was higher than any of the ones obtained with the TOTAL dataset. Although highest specificity and Kappa were scored by SVM, highest sensitivity and AUC values were scored by the RF.

Finally, the SMOTE dataset obtained the best general accuracy, especially in terms of sensitivity, Kappa and AUC, since scores for these three indices where higher than those obtained with the TOTAL and SPREAD datasets, in all cases. AUC, Kappa and sensitivity values varied from 0.91 to 0.99, 0.83 to 0.87, and 0.84 to 0.90 respectively. SVM showed the best specificity, but RF was the algorithm showing best AUC, sensitivity and Kappa scores.

Discussion

A semi-automated image processing and a supervised classification validation method have been developed and applied to detect the presence/absence of bluefin tuna in sonars that are routinely used by the fishing vessels targeting this species. The results of the classification validation tests show that all algorithms have good classification efficiency. Among the three datasets used in the experiment, the TOTAL dataset obtained a good overall performance, but
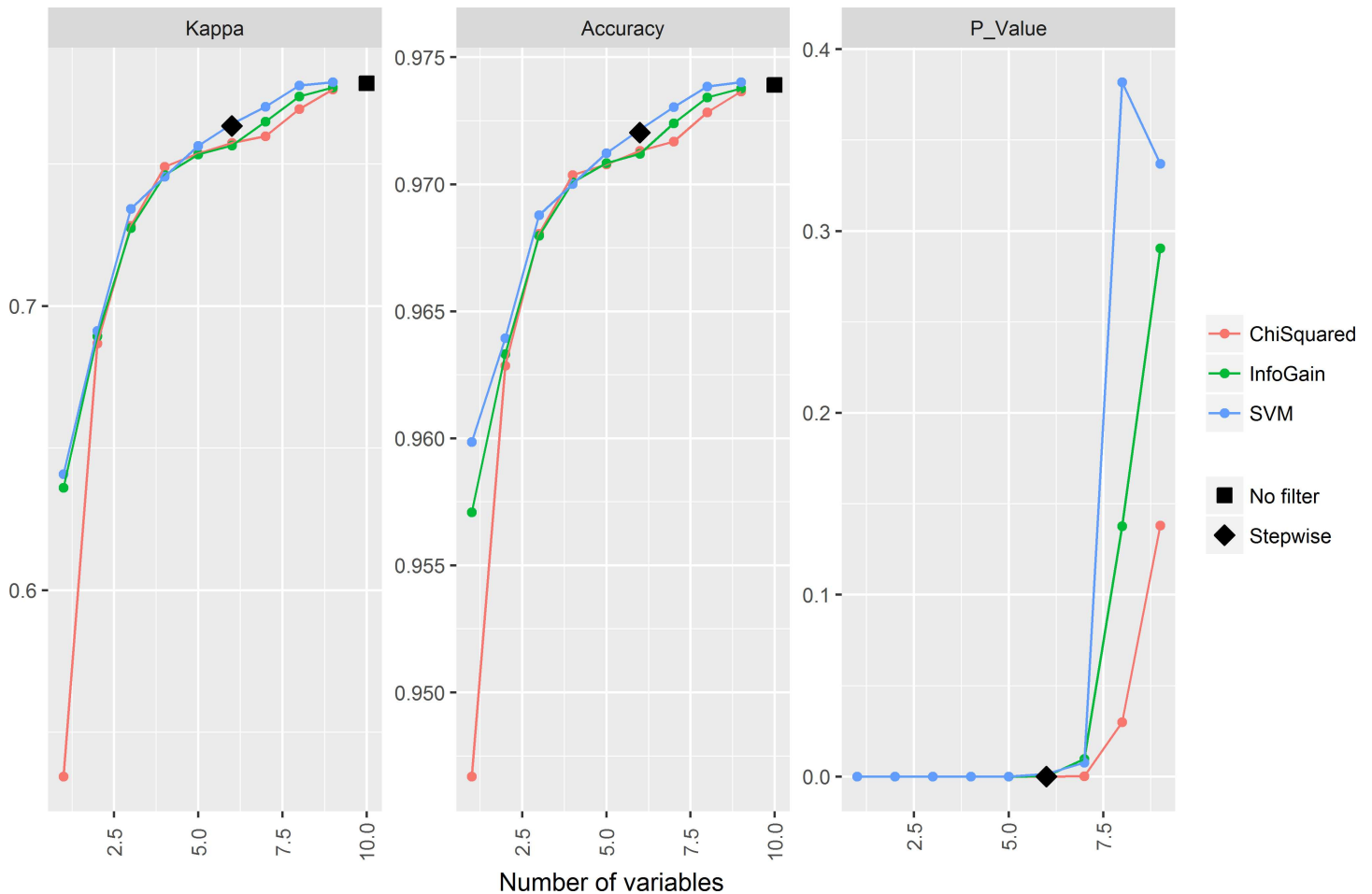
Fig 5. Comparison between the complete dataset and the reduced datasets. Values for *Kappa*, *Accuracy* and *P_Value* (obtained from a corrected resampled t-test) are shown.

doi:10.1371/journal.pone.0171382.g005

balanced datasets SPREAD and SMOTE subsequently improved the general performance. The RF algorithm applied to the SMOTE dataset provided the highest accuracy among the tested algorithms. Nevertheless, although different machine learning algorithms were compared, the main objective of the experiment was not to select the best algorithm. The overall good performance in classifying "tuna" and "no tuna" cases allowed us to validate the proposed methodology. The particular algorithms can be selected on a case by case basis, considering additional constraints (e.g. computing time) in particular future applications. In fact, MLP and SVM require substantially larger calculation time, which can be an additional consideration to guide selection in specific applications such as the processing of massive amounts of data (e.g. obtained from monitoring programs in the whole fleet throughout the whole fishing season), or when the speed of the analysis is critical (e.g. for near real time monitoring of resource abundance and distribution).

On one hand, the results of this work indicate that the designed methodology has an appreciable morphologic discriminatory capacity with the processed images. On the other hand, it should be taken into account that the ratio of positives and negatives in the set of images used in this experiment may not be representative of the ratio in the commercial fishing trips conducted by the baitboat fleet in the Bay of Biscay (where a higher percentage of negative cases is expected). This will have to be taken into account when the model is applied to datasets
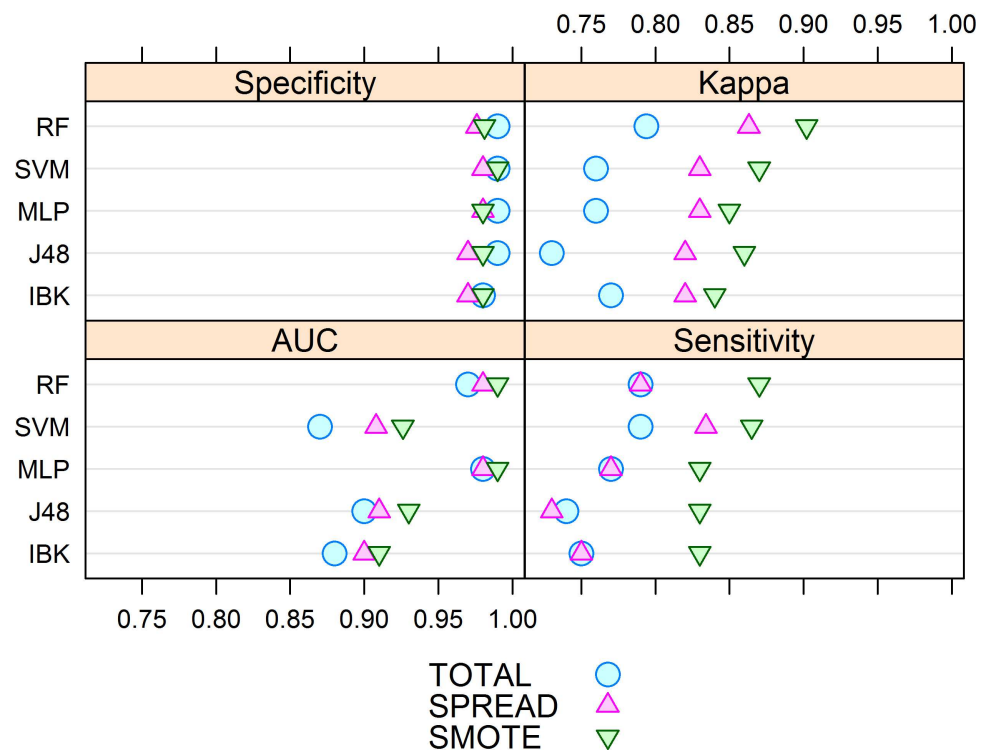
Experimental Results

Fig 6. Experiment results. Specificity, sensitivity, AUC and Kappa values for the three datasets: TOTAL, a complete dataset with 22501 instances; SPREAD, an oversampled dataset with 23998 instances; and SMOTE, a subsampled dataset with 11999 instances. The Y axis represents the classification method used: Random Forest (RF), Support Vector Machine (SVM), Multilayer Perceptron (MLP), J 48 and IBK.

doi:10.1371/journal.pone.0171382.g006

obtained e.g. during an entire fishing campaign by estimating the real presence/absence ratios and using algorithms that properly deal with uncompensated datasets. In addition, in such an extensive application of the model, the classification will have to be semi-supervised. Although a decrease in efficiency of the classification might be expected, the generalization of the model will likely increase since a larger variety of situations will be encountered [51].

In order to enhance the strengths of this methodology, several future research lines are being developed. First, following [8], a flexible Optical Character Recognition (OCR) method to extract metadata from sonar screens (sonar signal range, tilt, gains, speed, heading, as well as additional information) is being developed so that extra information can be introduced to guide classification on an image by image basis. This will also allow providing standardized tuna school sizes, since they can be specially affected by the gain settings. And second, tools for temporal tracking of schools should be used to identify the same school in a set of sequenced images. This is a necessary step in order to be able to quantify the total number of schools observed as well as to characterize their size. Additionally, these sonar observations could be paired with additional bluefin tuna presence/absence data from logbooks and/or scientific observers, as they become available, to allow a continuous improvement of the reference data-set used to train the algorithms.

A third research line will consist in combining the MRS data with scientific echosounder data [25]. The main purpose of extracting the number and size of schools from MRS screen-shots is to provide an index of abundance of bluefin tuna; for instance, something of the type

of a sonar mapping [52]. It is clear, though, that the data obtained from MRS images might not be as precise as those from standard acoustic-trawl surveys, based on echointegration [53] of data recorded by calibrated scientific echosounders [54]. However, currently, there are no ongoing acoustic surveys estimating the abundance of Atlantic bluefin tuna in the Bay of Biscay nor anywhere else, due to the large spatial distribution and high mobility of this species. In addition, typical single-vessel acoustic-trawl surveys have spatial-temporal limitations that could be overcome by an extensive implementation of this methodology [5]. Taking this into account, we plan to combine the extensive sonar mapping based on this methodology with the density distribution of the schools measured by a scientific echosounder. This would allow us to overcome the inherent uncertainty of the analog sonar images and also the sampling limitation of the standard, single vessel scientific echosounder acoustics. In practical terms, the low cost of the data acquisition device and the automation of the process would allow it to be applied extensively (in many vessels and through large periods of time) while carrying scientific echosounders in one or a few of the vessels. Additionally, the side-scan sonars increase the volume sampled near the surface and thus may constitute an adequate sampler of near-surface distribution species as the bluefin tuna while feeding in the Bay of Biscay. Side scan sonars have been successfully used for fishery work in other areas in the past [55, 56, 57], and allow in-season decisions on the spatial and temporal sub-allocation of the total allowable catch [7].

The morphological differences between the tuna and no-tuna blobs allow for their discrimination. In fact, tuna blobs were generally larger, more elongated and showed a more horizontal alignment. Some of these characteristics are ecologically meaningful, and the measurements obtained in the different blobs can inform e.g. about the size and shape of the bluefin tuna schools aggregated in the Bay of Biscay during the summer feeding season (Fig 7). Bauer et al [58] classified the size of tuna schools based on the surface disturbance observed by airplanes. Similarly, the measurements of the area of the blobs classified as "tuna" could be used to provide estimates of the size of the different schools in the future.

The Standing Committee on Research and Statistics of ICCAT has recurrently highlighted the need for developing fishery independent indices of abundance, given the problems associated with existing CPUEs and their inability to accurately track biomass changes, especially in recent years after the implementation of the recovery plan [14]. Our study can be considered a first milestone towards getting more accurate indices of abundance for juvenile bluefin tuna in the Bay of Biscay, and this can be pursued in two ways: (i) On one hand, the automated procedure presented in our study could be applied to MRS images recorded onboard commercial fishing vessels during their commercial operations. The bluefin tuna detections per unit of effort (DPUE, in number of schools per time unit) could be standardized, just in a similar way to the CPUE observations of the commercial fleet [22], currently used in the bluefin tuna assessment model. The signal of inter annual variability in bluefin tuna abundance can be isolated by removing the variability in DPUE due to other variables like month, area, or skipper skill, and this time series of standardized DPUE could be used as an index of relative abundance to tune the stock assessment models. Compared to the standardized CPUE that is currently used, the standardized DPUE index would have the advantage that the detections by the sonar, unlike the catch, would be independent from factors affecting the bluefin catchability by baitboats, such as the availability of tuna forage in the environment. It would, therefore, in principle better reflect the real abundance of bluefin tuna in the Bay of Biscay, compared to the standardized CPUE that is based on what the baitboat fleet was able to finally catch. However, the variability of factors affecting detection of bluefin tuna by the sonars would need to be considered in the DPUE standardization process. Since vessels could use different sonar settings at different times, it is important to standardize DPUE observations to standard sonar range, tilt, and gain values.
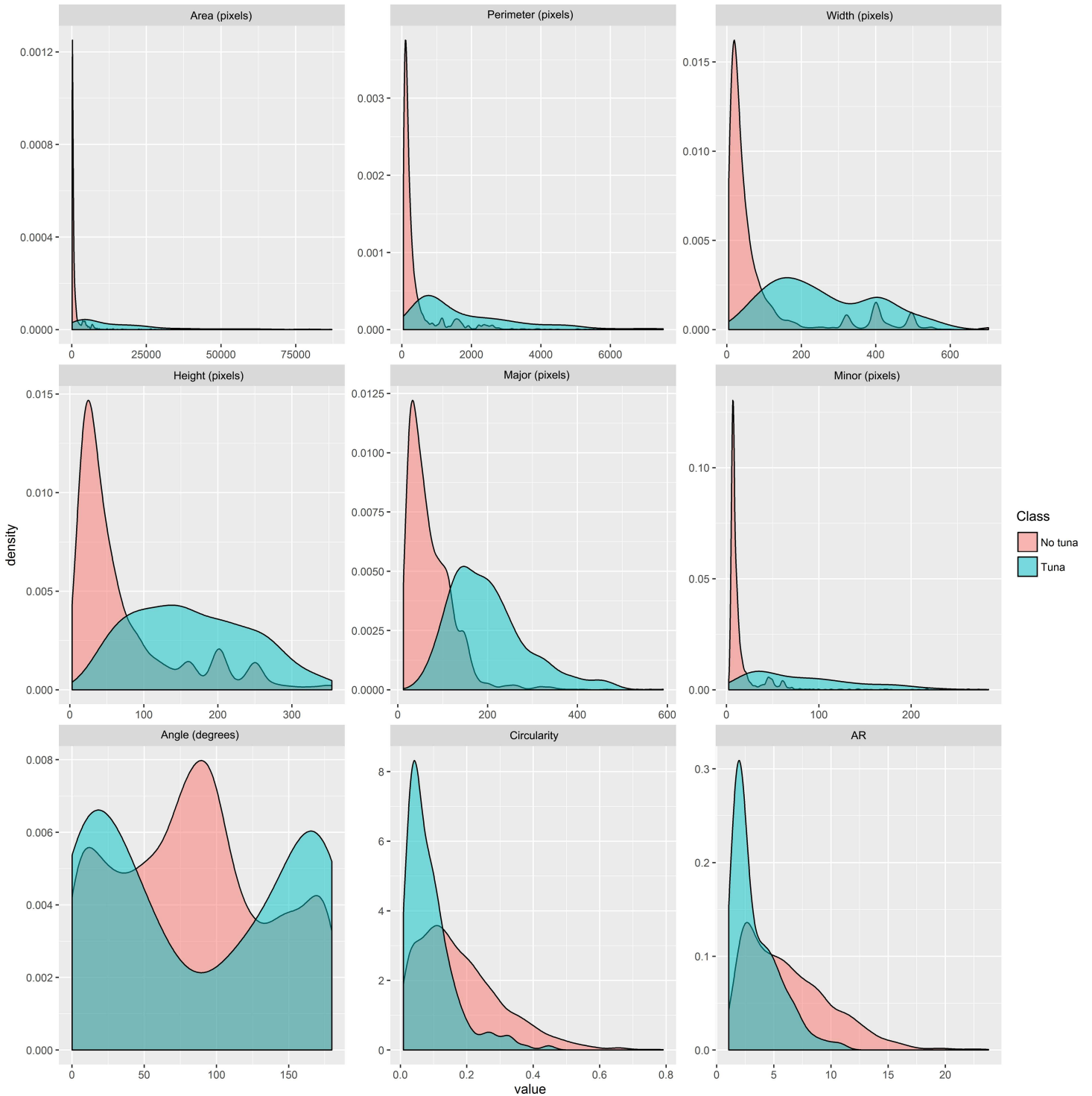
Fig 7. Density plots of the measured characteristics for tuna and no-tuna blobs. Only ecologically meaningful characteristics, related to size and shape of the schools, are plotted. "Angle" refers to the angle (0–180˚) between the X axis of the image and the primary axis of the best fitted ellipse to the blob contour; "circularity" is proportional to the ratio between the area and the squared perimeter, with a value of 1 representing a perfect circle and a value of 0 representing an increasingly elongated shape; Aspect Ratio (AR) is the ratio between the primary and secondary axes of the fitted ellipse., The area that a single pixel represents ranges between 0.69 m2 and 0.92 m2 (depending on the gain setting of the sonar).

doi:10.1371/journal.pone.0171382.g007

(ii) On the other hand, transect based systematic surveys covering the Bay of Biscay onboard commercial baitboats equipped with MRS could be designed and conducted yearly to quantitatively estimate the bluefin tuna school density (in number of schools per area unit). Such a time series could be used as a relative index of abundance to tune stock assessment models. Given the relatively high mobility of tunas (compared to small pelagics or demersal resources), ideally the systematic surveys would involve several commercial boats equipped with MRS so that the whole area of distribution can be searched in few days. This is advantageous compared to when a single boat prospects all the area (which is often the case when scientific acoustic equipment is used to estimate total biomass), because the probabilities of immigration, emigration and double counting schools are diminished. Bluefin tuna concentrate in a relatively small area while feeding during summer in the Bay of Biscay (Fig 1), which provides a unique opportunity to conduct systematic abundance surveys on this widely distributed species [30].

The relatively non-expensive methodology presented in this study can also be adjusted to other tuna and non-tuna pelagic fisheries by adapting the analyses to the specific type of sonar, output signal and display (see also [8]). This provides an interesting alternative to standard acoustic-trawl surveys, especially when targeting species of high mobility and/or near surface distribution. It thus provides an opportunity to use commercial fishing vessels as observatories of the pelagic ecosystem, and commercial sonars as tools to track changes in abundance of commercial species [6, 8].

Supporting information

S1 Dataset. Complete training dataset of tuna and no-tuna blobs. "Blob_ID" is a unique blob identifier (as a concatenation of survey, year, time and blob number); "Area" and "Perimeter" of the blob are in number of pixels; "BX" and "BY" refer to the upper left corner coordinates of the smallest rectangle housing the blob; width and height refer to the dimensions (in pixels) of such rectangle; "X", "Y", "Major", "Minor" and "Angle" refer to the coordinates of the centroid, the size of the principal and secondary axes, as well as the angle (with respect to the horizontal axis) of the best fitting ellipse; "Circularity" is proportional to the ratio between the area and the squared perimeter, with a value of 1 representing a perfect circle and a value of 0 representing an increasingly elongated shape; Feret's diameter, or "Feret" is the longest distance between any two points along the selection boundary, also known as maximum caliper; "FeretAngle" is the angle (0–180 degrees) of the Feret's diameter and "MinFeret" is the minimum caliper diameter; "FeretX" and "FeretY" refer to the starting coordinates of the Feret diameter; "Aspect Ratio" (AR) is the ratio between the primary and secondary axes of the fitted ellipse; "Roundness" is the inverse of AR; "Solidity" is the ratio between the area and the convex area of the blob; and "Class" refers to the "tuna" or "no-tuna" category.
(ZIP)

Author contributions

Conceptualization: HA GB MCH JU.

Data curation: JU NG IA.

Formal analysis: JU HA GB MCH.

Funding acquisition: HA NG JS JU.

Investigation: JU HA GB MCH NG IA.

Methodology: JU HA GB MCH JAF YY.

Project administration: HA GB NG JS JU.

Resources: JU HA GB MCH NG JS.

Software: JU HA GB MCH.

Supervision: HA GB MCH YY.

Validation: JU GB MCH NG JAF.

Visualization: JU HA MCH IA JS.

Writing – original draft: JU HA MCH GB.

Writing – review & editing: JU HA GB MCH NG IA JAF YY JS.

References

1. Fromentin JM, Powers JE. Atlantic bluefin tuna: population dynamics, ecology, fisheries and management. Fish and Fisheries. 2005; 6(4):281–306.

2. ICCAT. Report of the 2012 Atlantic bluefin tuna Stock Assessment Session (Madrid, Spain—September 4–11, 2012). Collective Volume of Scientific Papers ICCAT. 2013; 69(1): 1–198.

3. Fromentin JM, Bonhommeau S, Arrizabalaga H, Kell LT. The spectre of uncertainty in management of exploited fish stocks: The illustrative case of Atlantic bluefin tuna. Marine Policy. 2014; 47:8–14.

4. Petitgas P, Cotter J, Trenkel V, Mesnil B. Fish stock assessments using surveys and indicators. Aquatic Living Resources. 2009; 22(2):119–1.

5. Mayer L, Li Y, Melvin G. 3D visualization for pelagic fisheries research and assessment. ICES Journal of Marine Science: Journal du Conseil. 2002; 59(1):216–225.

6. Dalen J, Karp WA. Collection of acoustic data from fishing vessels. 287. International Council for the Exploration of the Sea; 2007.

7. Melvin G, Stephenson R, Power M, Fife F, Clark K. Industry acoustic surveys as the basis for in-season decisions in a co-management regime. Herring: Expectations for a new millennium University of Alaska Sea Grant, AK-SG-01-04, Fairbanks (This volume). 2001.

8. Brehmer P, Lafont T, Georgakarakos S, Josse E, Gerlotto F, Collet C. Omnidirectional multibeam sonar monitoring: applications in fisheries science. FISH and Fisheries. 2006; 7(3):165–179.

9. Misund OA. Underwater acoustics in marine fisheries and fisheries research. Reviews in Fish Biology and Fisheries. 1997; 7(1):1–34.

10. Brehmer P, Georgakarakos S, Josse E, Trygonis V, Dalen J. Adaptation of fisheries sonar for monitoring schools of large pelagic fish: dependence of schooling behaviour on fish finding efficiency. Aquatic Living Resources. 2007; 20(4):377–384.

11. Gerlotto F, Freon P. Some elements on vertical avoidance of fish schools to a vessel during acoustic surveys. Fisheries Research. 1992; 14(4):251–259.

12. Melvin G, Li Y, Mayer L, Clay A. Commercial fishing vessels, automatic acoustic logging systems and 3D data visualization. ICES Journal of Marine Science: Journal du Conseil. 2002; 59(1):179–189.

13. Maunder MN, Sibert JR, Fonteneau A, Hampton J, Kleiber P, Harley SJ. Interpreting catch per unit effort data to assess the status of individual stocks and communities. ICES Journal of Marine Science: Journal du Conseil. 2006; 63(8):1373–1385.

14. ICCAT. Report of the Standing Committee on Research and Statistics (SCRS) (Madrid, Spain, 3 to 7 October 2016). 2016. 429 pp. Available from: https://www.iccat.int/Documents/Meetings/Docs/2016_SCRS_ENG.pdf.

15. García A, Alemany F, De la Serna J, Oray I, Karakulak S, Rollandi L, et al. Preliminary results of the 2004 bluefin tuna larval surveys off different Mediterranean sites (Balearic Archipelago, Levantine Sea and the Sicilian Channel). Collective Volume of Scientific Papers ICCAT. 2005; 58:1261–1270.

16. Antonio Di Natale M, Justel-Rubio A. ICCAT Atlantic-wide Research Programme for Bluefin Tuna (GBYP) activity report for 2013 (extension of Phase 3 and first part of Phase 4). Collect Vol Sci Pap ICCAT. 2014; 70(2):459–498.

17. Lutcavage M, Kraus S, Hoggard W. Aerial survey of giant bluefin tuna, Thunnus thynnus, in the great Bahama Bank, Straits of Florida, 1995. Fishery Bulletin. 1997; 95(2):300–310.

18. Weber TC, Lutcavage ME, Schroth-Miller ML. Near resonance acoustic scattering from organized schools of juvenile Atlantic bluefin tuna (Thunnus thynnus). The Journal of the Acoustical Society of America. 2013; 133(6):3802–3812. doi: 10.1121/1.4802646 PMID: 23742334

19. Itoh T, Tsuji S. Review of acoustic monitoring survey–analyses of data for eight years. CCSBT; 2004. 03.

20. Itoh T, Sakai O, Tokuda D. Report of the piston-line trolling monitoring survey for the age-1 southern bluefin tuna recruitment index in 2011/2012. CCSBT; 2012.

21. Rodriguez-Marín E, Arrizabalaga H, Ortiz M, Rodrıguez-Cabello C, Moreno G, Kell L. Standardization of bluefin tuna, Thunnus thynnus, catch per unit effort in the baitboat fishery of the Bay of Biscay (East-ern Atlantic). ICES Journal of Marine Science: Journal du Conseil. 2003; 60(6):1216–1231.

22. Santiago J, Arrizabalaga H, Ortiz M, Goñi N. Updated standardized bluefin tuna CPUE index of the Bay of Biscay baitboat fishery (1952–2014). Collect Vol Sci Pap ICCAT. 2016; 72(7):1694–1714.

23. Arreguin-Sanchez F. Catchability: a key parameter for fish stock assessment. Oceanographic Literature Review. 1997; 1(44):58.

24. Stoner A. Effects of environmental variables on fish feeding ecology: implications for the performance of baited fishing gear and stock assessment. Journal of Fish Biology. 2004; 65(6):1445–1471.

25. Miquel J, Delgado de Molina A, Ariz J, Delgado de Molina R, Déniz S, Díaz N, et al. Acoustic Selectivity in Tropical Tuna (Experimental Purse-Seine Campaign in the Indian Ocean).Western and Central Pacific Fisheries Commission. 2006; 'WCPFC-SC2', FT WP-8.

26. Trygonis V, Georgakarakos S, Dagorn L, Brehmer P. Spatiotemporal distribution of fish schools around drifting fish aggregating devices. Fisheries Research. 2016; 177, 39–49.

27. Irigoien X, Fernandes JA, Grosjean P, Denis K, Albaina A, Santos M. Spring zooplankton distribution in the Bay of Biscay from 1998 to 2006 in relation with anchovy recruitment. Journal of Plankton Research. 2009; 31(1):1–17.

28. Fernandes JA, Irigoien X, Boyra G, Lozano JA, Inza I. Optimizing the number of classes in automated zooplankton classification. Journal of Plankton Research. 2009; 31(1):19–29.

29. Bachiller E, Fernandes JA, Irigoien X. Improving semiautomated zooplankton classification using an internal control and different imaging devices. Limnol Oceanogr Methods. 2012; 10:1–9.

30. Arrizabalaga H, Dufour F, Kell L, Merino G, Ibaibarriaga L, Chust G, et al. Global habitat preferences of commercially valuable tuna. Deep Sea Research Part II: Topical Studies in Oceanography. 2015; 113:102–112.

31. Goñi N, Arrizabalaga H. Seasonal and interannual variability of fat content of juvenile albacore (Thun-nus alalunga) and bluefin (Thunnus thynnus) tunas during their feeding migration to the Bay of Biscay. Progress In Oceanography. 2010; 86:115–123

32. Zarauz L, Irigoien X, Fernandes JA. Modelling the influence of abiotic and biotic factors on plankton dis-tribution in the Bay of Biscay, during three consecutive years (2004–06). Journal of Plankton Research. 2008; 30(8):857–872.

33. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling tech-nique. Journal of artificial intelligence research. 2002; 16:321–357.

34. Witten IH, Frank E. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann; 2005.

35. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. ACM SIGKDD explorations newsletter. 2009; 11(1):10–18.

36. Breiman L. Random forests. Machine learning. 2001; 45(1):5–32.

37. Dietterich TG. Approximate statistical tests for comparing supervised classification learning algorithms. Neural computation. 1998; 10(7):1895–1923. PMID: 9744903

38. Wood, J. M. Understanding and computing Cohen's kappa: A tutorial. WebPsychEmpiricist. Web Jour-nal. 2007 [cited 2016 Dec 15]. Available from: http://wpe.info/vault/wood07/Wood07.pdf

39. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Com-puting, Vienna, Austria. 2013. ISBN 3-900051-07-0; 2014.

40. Venables WN, Ripley BD. Modern applied statistics with S-PLUS. Springer Science & Business Media; 2013.

41. Burges CJ. A tutorial on support vector machines for pattern recognition. Data mining and knowledge discovery. 1998; 2(2):121–167.

42. Cortes C, Vapnik V. Support-vector networks. Machine learning. 1995; 20(3):273–297.

43. Bishop CM. Neural networks for pattern recognition. Oxford university press; 1995.

44. Haykin S, Network N. A comprehensive foundation. Neural Networks. 2004; 2(2004).

45. Quinlan JR. Improved use of continuous attributes in C4. 5. Journal of artificial intelligence research. 1996; 4:77–90.

46. Aha DW, Kibler D, Albert MK. Instance-based learning algorithms. Machine learning. 1991; 6(1):37–66.

47. Kuhn M. Caret package. Journal of Statistical Software. 2008; 28(5).

48. Freeman EA, Moisen G. PresenceAbsence: An R package for presence absence analysis. Journal of Statistical Software. 2008. 23(11): 31.

49. Fawcett T. An introduction to ROC analysis. Pattern recognition letters. 2006; 27(8):861–874.

50. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Ijcai. vol. 14; 1995. p. 1137–1145.

51. Chapelle O.; Scholkopf B. Zien A. Semi-Supervised Learning (Chapelle, O. et al., Eds.; 2006) IEEE Transactions on Neural Networks, IEEE, 2009, 20, 542–542.

52. Smith P. E. The horizontal dimensions and abundance of fish schools in the upper mixed layer as measured by sonar. Proceedings of an international symposium on biological sound scattering in the ocean, 1970, 31.

53. Dragesund O, Olsen S. On the possibility of estimating year-class strength by measuring echo-abundance of 0-group fish. Pattern recognition letters. 1965.

54. Simmonds J, MacLennan DN. Fisheries acoustics: theory and practice. John Wiley & Sons; 2008.

55. Hewitt R. Developments and use of sonar mapping for pelagic stock assessment in the California current. Fisheries Bulletin. 1976; 74:281–300.

56. O Driscoll RL, McClatchie S. Spatial distribution of planktivorous fish schools in relation to krill abundance and local hydrography off Otago, New Zealand. Deep Sea Research Part II: Topical Studies in Oceanography. 1998; 45(7):1295–1325.

57. Melvin GD. Observations of in situ Atlantic bluefin tuna (Thunnus thynnus) with 500-kHz multibeam sonar. ICES Journal of Marine Science: Journal du Conseil. 2016; May 20: fsw077.

58. Bauer RK, Bonhommeau S, Brisset B, Fromentin JM. Aerial surveys to monitor bluefin tuna abundance and track efficiency of management measures. Marine Ecology Progress Series. 2015; 534: 221–34.