Explainable Classification Methods for Fish Species Detection using Hydroacoustic Data

Lucas Tubino Bonifácio Costa
Computational Modeling PhD Program
Federal University of Rio Grande, Campus Carreiros
Rio Grande RS, Brazil
email: tubinobcosta@gmail.com

Graçaliz Pereira Dimuro
Centro de Ciências Computacionais
Federal University of Rio Grande, Campus Carreiros
Rio Grande RS, Brazil
Departamento de Estadística, Informática y Matemáticas
Universidad Pública de Navarra, Campus Arrosadia
Pamplona, Spain
email: gracaliz@gmail.com

Eduardo Nunes Borges
Centro de Ciências Computacionais
Federal University of Rio Grande, Campus Carreiros
Rio Grande RS, Brazil
email: enborges@gmail.com

Giancarlo Lucca
Computational Modeling PhD Program
Federal University of Rio Grande, Campus Carreiros
Rio Grande RS, Brazil
email: giancarlo.lucca88@gmail.com

Leonardo R. Emmendorfer
Centro de Ciências Computacionais
Federal University of Rio Grande, Campus Carreiros
Rio Grande RS, Brazil
email: leonardo.emmendorfer@gmail.com

Stefan Cruz Weigert
Oceanography Department
Federal University of Rio Grande, Campus Carreiros
Rio Grande RS, Brazil
email: stefanweigert@gmail.com

***Abstract*—This work aims to evaluate explainable classification methods for the detection of fish species from hydroacoustic data acquired by echo sounders at a region near the coastline of south and southeastern Brazil. Decision trees and fuzzy rule-based methods were adopted. The fitted models were evaluated by quality measures based on the performance of the classifiers and also by an expert which analyzes the usefulness of the rules on describing the schools. The models learned by the algorithms performed well for the available data and were able to represent the documented behavior of the species considered in the studied region, according to the literature.**

***Index Terms*—hydroacoustics, fish classification, explanability, fuzzy rule-based classification systems**

I. INTRODUCTION

The automatic identification of the prevailing species in a fish school allows a non-invasive approach for the analysis of marine life. The approach is also useful for increasing the capture of fish from economically relevant species as well as reducing the fishery of organisms that would be discarded, therefore contributing to the preservation of several species [1]. Catches of the most important pelagic fish along the coast of Brazil represent hundreds of thousands of tons per year.

Modern fishing boats equipped with echo sounders can move freely around fishing grounds and digital echograms can be checked online from the boats or from the fleet headquarters [1], [2]. An echo sounder is an instrument that transmits and receives sound vertically through the water column [2]. This equipment typically works attached to the underside of a vessel, sending acoustic signals to detect fish, mollusks, zooplankton, or other objects in the water column, generating acoustic records that appear on echograms. Modern hydroacoustic systems allow the detection, quantification, and identification of fish species [3].

In the context of fisheries research, the classification and identification of acoustic targets traditionally combine knowledge about the distribution and behavior patterns of constituent species, which includes the analysis of acoustic and catch data [2]. Currently, the method most widely adopted to classify acoustic records is the characterization of echotypes, where certain patterns in acoustic records are visually identified by a specialist who, based on information called descriptors, determines a class for the echo-record. Thus, a key question related to this topic is that the estimate's precision is highly related to the quality of the interpretation of the echo-records by fishermen and researchers.

The application of those techniques has been limited by an inability to objectively discriminate among taxonomic groups of sound scatterers [4]. To date, although echo-sound software programs are used onboard or remotely to process echograms, most workers in commercial fishery adopt subjective identi-

This study was supported by PNPD/CAPES (464880/2019-00) and CAPES Financial Code 001, CNPq (301618/2019-4) and FAPERGS (17/2551-0000872- 3, 19/2551-0001279-9, 19/ 2551-0001660).

fication methods based on visual interpretation of echograms with taxonomic discrimination based on “rules of thumb”.

Computerized systems for fish school detection and sizing came into major use with the onset of the computer technology era in the mid-1970s [5], [6]. Classification algorithms were already applied for species identification from hydroacoustic data [7]. Recently the random forest classification algorithm [8] was applied for the identification of *Thunnus thynnus* (Atlantic bluefin tuna) from sonar images in the region of the Bay of Biscay [9].

The aforementioned classification problem is characterized by subjective features such as the shape or behavior of the schools, which motivates the adoption of fuzzy classifiers. The explainable artificial intelligence, from a broader perspective, aims to build intelligent systems that are intelligible by users [10]. This work proposes and evaluates explainable classifications models for the automatic detection of fish species from hydroacoustic data acquired by echo sounders. The adoption of explainable models allows experts to evaluate the resulting rules and assess their applicability for practical purposes.

II. Species identification by fishery acoustics

For part or all of their life, fish may spend their time as associated with other fish in shoals. The terms shoal and school are sometimes used interchangeably, but it is useful to retain school for the situation in which the fish in a shoal show a high degree of coordination in their spatial positions within the shoal [11], which might indicate that the fishes in that shoal are mostly from the same species.

The identification of the prevailing species in a school allows improving the effectiveness of fishery by enhancing the success rate of capturing fish from species of commercial interest while preserving fish from other species [1]. In multispecies environments, especially where schools are small, interspersed, and have a low or varying catchability, net sampling is often unworkable even as a rough means of taxonomic identification. Net samples in practice cannot achieve spatial or temporal sampling which is comparable with that of acoustic sampling [4].

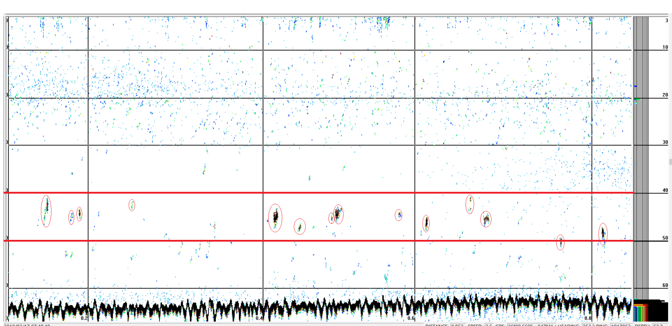

Fig. 1. Echogram obtained from a split-beam echo sounder. Red ellipses indicate the presence of fish schools. Reproduced under permission of the Fisheries Technology and Hydroacoustics Laboratory (FURG, Brazil).

Figure 1 represents a single echogram, where the areas circulated in red represent schools. Any structure graphically registered in an echogram is called an echo-record, therefore many echo-records might be registered in the same echogram. The classification of echo-records is performed through the characterization of echotypes. An echotype is a morphologically consistent pattern of the echo-record, characterized by a set of descriptors [12].

The determination of the correct class of an echo-record is achieved by the analysis of several types of descriptors by a specialist. Energetic descriptors are related to backscatter data from the operating frequency of the echo sounder and/or the color visible in the echogram, which is usually related to the biomass of individuals. Morphological descriptors are related to the shape and extent of the echo-record, and also its clustering degree, which varies as a function of the density of individuals, among other factors. Spacial descriptors include the location and depth of the echo-record. Figure 2 illustrates the characterization of spatial descriptors. Temporal descriptors represent the date and time of detection of the echo-record.

The classification of echo-records requires fishing trawls for the correct identification of the predominant fish species. Hauls trawl depth must be monitored to be related to echo-records found in the same depth. Net sampling must occur at the same time the echogram is obtained, which allows the species identification and weighing of fishes onboard.

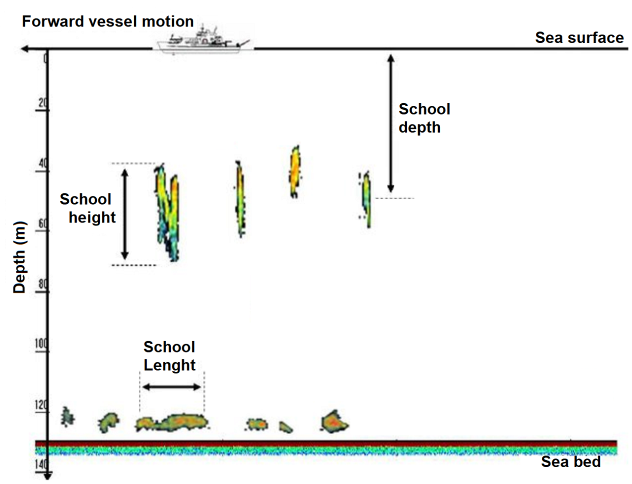

Fig. 2. Examples of echo-records associated to fish school and their descriptors in an echogram. Adapted from [7].

Some descriptors might present high inconsistency or a wide variability, notably the morphological and biological ones. This results from variations on the conditions for sound wave propagation due to changes in the water temperature and/or vertical migration of fish, its prey, or other organisms present in the water column [12]. Those features motivated the adoption of fuzzy logic as a suitable alternative for the classification of echo-records since this approach is recognized as being more robust to imprecise, ambiguous data.

III. Explainable classification methods

Explainable Artificial Intelligence (XAI) is a research field that aims to make AI systems results more understandable to humans While the term is relatively new, the problem of explainability has existed since the mid-1970s when researchers studied explanation for expert systems [10]. An explainable intelligent system would be able to provide auditable and provable ways to defend algorithmic decisions, which leads to building trust. Additionally, a model that can be explained and understood is one that can be more easily improved [10].

A. C4.5

Among the XAI methods available from the literature, the inference of decision trees represents a simple rather effective approach. C4.5 [13] is a recursive decision tree learning algorithm that adopts the information gain ration to choose the attribute which better explains the class at each stage. The algorithm builds a decision tree following a top-down approach.

A base case for the recursion occurs when all the objects at a given node belong to the same class. Also, a parameter regulates the minimum number of instances per leaf. The method is widely adopted for object classification on several domains [14]. C4.5 adopts a crispy approach for the split over each attribute selected therefore the rules generates from the algorithm represent abrupt transitions between classes. Those sharp decision boundaries are questionable and often unnatural [15].

B. FARC-HD

According to [16], classification methods based on fuzzy rules better encompass the whole information available, even though it comes from expert knowledge, empirical measures, or mathematical models. A fuzzy association rule can be considered to be a classification rule if the antecedent contains fuzzy item sets and the consequent part contains only one class label [17].

FARC-HD (*Fuzzy Association Rule-based Classification method for High-Dimensional problems*) [17] is a computationally efficient fuzzy-based algorithm based on association discovery. Figure 3 illustrates the general scheme of the FARC-HD method.

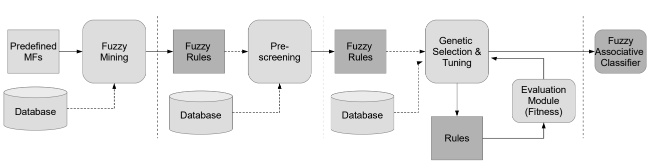

Fig. 3. Scheme of the FARC-HD method. Adapted from [17].

At the first stage, the fuzzy association rule extraction generates an initial set of rules comprising all classes in the dataset from an input set of predefined triangular membership functions which could be uniformly distributed, for instance. Candidate rule prescreening stage performs a selection from the best rules generated previously. Finally, a genetic approach is adopted for the tuning of the rules selected. The authors adopted a fixed number of five lingustic variables for the fuzzy representation of each attribute in the evaluation of the method [17]. Figure 4 illustrates five triangular lingustic variables and the respective membership functions for an arbitrary attribute.

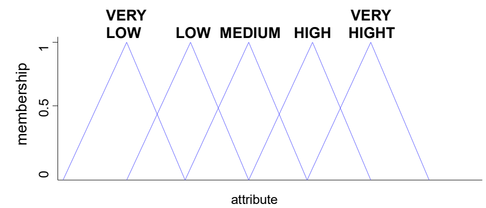

Fig. 4. Lingustic variables and their respective triangular membership functions for an arbitrary attribute.

C. FURIA

FURIA [15] extends the rule learning algorithm RIPPER [18] on allowing the adoption of fuzzy and non-ordered rules. Essentially, a fuzzy rule is obtained through replacing intervals by fuzzy intervals, namely fuzzy sets with trapezoidal membership function. The fuzzy antecedents successively learned by FURIA are open fuzzy half intervals.

At the building stage, an initial ruleset is obtained. Rules are grown by the inclusion of antecedents. Each rule can raise one or more “optimizations”, which are variations from it. During the optimization phase, in each iteration, rules are fuzzified greedily. The fuzzification is evaluated for every antecedent A_i in terms of its purity ρ_i, which is defined as:

$$\rho_i = \frac{p_i}{p_i + n_i} \tag{1}$$

where $p_i = \sum_{x \in D^i_+} \mu_{A_i}(x)$ and $n_i = \sum_{x \in D^i_-} \mu_{A_i}(x)$. Each $\mu_{A_i}(x)$ is the degree to which the antecendent A_i covers each x, where x is an instance. D^i_+ and D^i_+ are the subsets of positive and negative instances respectively, with respect to antecendent A_i. The fuzzification is then performed for the antecedent with the greatest purity. The process is repeated until all antecedents have been fuzzified [15].

The model is built following a k-fold approach, where one fold is used for pruning. The user must inform the number of optimizations for each rule and the number folds, among other parameters. The algorithm has been successfully applied to classification problems from several fields including medicine [19], urban traffic prediction [20], and climate [15].

IV. Methodology

Fish hydroachoustic data for this study[1] were collected using a Simrad EK500 scientific split-beam echo sounder[2] operating at 38kHz. Data were obtained from two cruises

[1]Data provided by the Fisheries Technology and Hydroacoustics Laboratory at the Federal University of Rio Grande, Brazil (http://www.io.furg.br/)

[2]Simrad Fisheries; Lynnwood, USA

which occurred between 2009 and 2010 [1] at a region near coastline of south and southeastern Brazil, between the Santa Marta cape (28° 36' S) and São Tomé cape (22° 02' S) (Figure 5).

The cruises were carried out along pre-established grids during day and night, at a speed of 10 knots. The coverage of the cruises, calculated as the ratio of the number of miles prospected to the total area, was more than 20%. Mid-water trawl net sampling was performed whenever schools were detected. The net, designed to catch small pelagic fish, had wings and square with a mesh of 400 mm between knots, gradually decreasing to 50 mm in the tunnel and 20 mm in the bag, plus an internal 12 mm mesh bag. The net was kept open by the use of two doors with 4 m^2 each. The hauls were performed at speeds between 3 and 4 knots for a period that depended on the size of the schools [21].

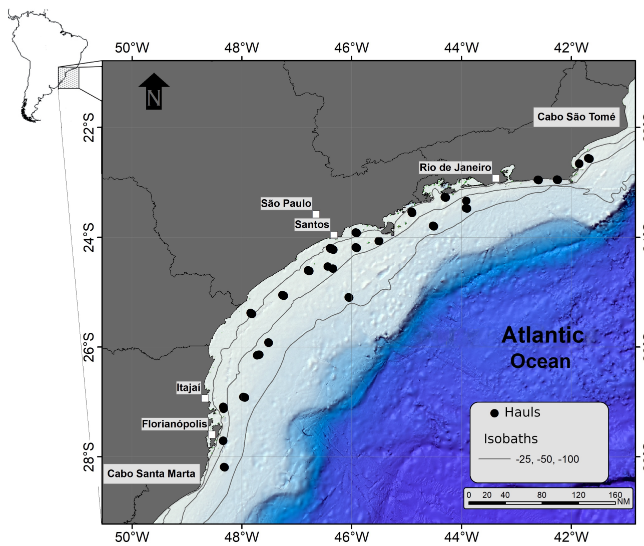

Fig. 5. Catch positions of the cruises in the study area. Isobaths refer to local depth. Reproduced under permission of the Fisheries Technology and Hydroacoustics Laboratory (FURG, Brazil).

Attribute	Average	s.d.	Median	Max.	Min.
SL (°)	24.8708	1.2979	24.6123	28.1939	22.5632
WL (°)	46.3506	1.6991	46.7773	48.3383	41.6664
LD (m)	42.8569	17.6523	34.0000	86.2000	19.4000
SD (m)	25.0027	13.5780	21.8000	58.0000	3.8000
ID (m)	28.1625	13.2570	24.3500	62.4000	6.3000
SH (m)	3.1713	2.5047	2.3000	15.9000	0.5000
SW (m)	10.9206	11.9705	7.6000	129.2000	0.9000

TABLE I
AVERAGE, STANDARD DEVIATION (S.D.), MEDIAN, MAXIMUM AND MINIMUM VALUES FOR EACH ATTRIBUTE OF THE DATASET.

Hydroacoustic data were digitally stored as raw acoustic data, which can be viewed in the form of echogram images using the MOVIES+ version 3.4b (IFREMER) software. The crew prepared data sheets of quantity (number and biomass) and percentage of the total biomass of the species present in each fishing haul with an echogram record and, from these data, the predominant species of fish in the haul was identified.

1. **if** LD $\leq$ 64.2 **and** SL $\leq$ 22.96 **then**
   class = *Engraulis anchoita*
2. **if** LD $\leq$ 64.2 **and** SL $>$ 22.96 **then**
   class = *Dactylopterus volitans*
3. **if** LD $\leq$ 64.2 **and** SL $>$ 24.61 **then**
   class = *Engraulis anchoita*
4. **if** 34.8 $<$ LD $\leq$ 64.2 **and** 23.34 $<$ SL $\leq$ 24.62 **and** SD $\leq$ 16.2 **then**
   class = *Dactylopterus volitans*
5. **if** LD $\leq$ 34.8 **and** 23.34 $<$ SL $\leq$ 24.62 **then**
   class = manjuba
6. **if** 34.8 $<$ LD $\leq$ 64.2 **and** 23.34 $<$ SL $\leq$ 24.62 **and** SD $>$ 16.2 **then**
   class = others
7. **if** LD $>$ 64.2 **and** LD $\leq$ 70.5 **then**
   class = *Trichiurus lepturus*
8. **if** LD $>$ 70.5 **and** SW $\leq$ 6.5 **then**
   class = *Trachurus lathami*
9. **if** LD $>$ 70.5 **and** SW $>$ 6.5 **then**
   class = *Dactylopterus volitans*

Fig. 6. The set of rules generated from C4.5.

1. **if** SL **is** RoughlyToNorth **and** LD **is** AvgToShallow **and** SD **is** RoughlyShallow **then**
   class = *Dactylopterus volitans* (CF = 0.9363)
2. **if** SL **is** MuchToNorth **and** SD **is** RoughlyShallow **then**
   class = *Dactylopterus volitans* (CF = 0.4459)
3. **if** WL **is** RoughlyToEast **and** SW **is** AvgToNarrow **then**
   class = *Dactylopterus volitans* (CF = 1.0)
4. **if** WL **is** RoughlyToEast **and** SH **is** AvgToHigh **then**
   class = *Dactylopterus volitans* (CF = 0.8822)
5. **if** LD **is** AvgToShallow **and** SD **is** VeryShallow **then**
   class = *Dactylopterus volitans* (CF = 0.8561)
6. **if** WL **is** MuchToEast **then**
   class = *Engraulis anchoita* (CF = 0.8348)
7. **if** SL **is** RoughlyToSouth **then**
   class = *Engraulis anchoita* (CF = 0.9011)
8. **if** WL **is** MuchToWest **and** SD **is** RoughlyShallow **then**
   class = *Engraulis anchoita* (CF = 0.9995)
9. **if** WL **is** RoughlyToEast **and** LD **is** RoughlyDeep **then**
   class = *Trachurus lathami* (CF = 0.8616)
10. **if** SL **is** MuchToNorth **and** WL **is** MuchToWest **then**
    class = manjuba (CF= 0.8722)
11. **if** WL **is** RoughlyToWest **and** LD **is** RoughlyShallow **and** SD **is** RoughlyShallow **then**
    class = manjuba (CF: 0.7430)
12. **if** SL **is** RoughlyToNorth **and** WL **is** RoughlyToWest **and** LD **is** VeryDeep **then**
    class = others (CF =0.9466)
13. **if** SL **is** MuchToNorth **and** LD **is** AvgToShallow **and** SD **is** AvgToShallow **then**
    class = others (CF = 0.9637)
14. **if** WL **is** RoughlyToWest **and** SD **is** VeryDeep **then**
    class = others (CF = 0.9852)
15. **if** SD **is** VeryDeep **then**
    class = *Trichiurus lepturus* (CF = 0.9449)
16. **if** SL **is** MuchToSouth **and** LD **is** VeryDeep **then**
    class = *Trichiurus lepturus* (CF = 1.0)

Fig. 7. The set of rules generated from FARC-HD.

Through the analysis of morphological, spatial, and temporal descriptors, the echo-records detected were characterized into echotypes, according to common characteristics and consistent standards, according to the works carried out by [21] and also by [1].

Each entry in the fishery dataset represents the description of a fishing haul and its respective fish schools, containing the following data: South Latitude (SL) and West Longitude (WL) in decimal degrees of the coordinates of the school; Location Depth (LD) in meters which refers to the depth of the local where each school was detected; Superior and Inferior depth in meters (SD and ID respectively), which refer to the upper and lower depth of the school; School Height and Width in meters (SH and SW respectively). Those fields are metadata obtained from the echogram. Additionally, a class represents the prevailing species as recognized by the crew after the school is hauled.

Class	Attribute						
	SL	WL	LD	SD	ID	SH	SW
Dactylopterus volitans	23.53^d	44.46^d	47.09^b	10.43^c	17.34^c	6.91^a	29.02^a
Engraulis anchoita	25.27^b	46.71^b	39.17^c	24.59^b	27.53^b	$2.96^{b,c}$	9.03^b
Trachurus lathami	23.80^d	44.51^d	77.87^a	14.77^c	17.82^c	$3.05^{a,b,c}$	4.35^c
Trichiurus lepturus	27.03^a	47.99^a	65.90^a	51.44^a	55.08^a	$3.67^{a,b}$	15.55^a
manjuba	24.33^c	46.46^c	28.33^d	20.80^b	23.15^b	2.36^c	8.61^b
others	$24.07^{c,d}$	45.56^d	$44.12^{b,c}$	27.76^b	30.32^b	2.55^c	8.93^b
p-value (K-W)	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01

TABLE II

AVERAGE VALUES OF EACH ATTRIBUTE FOR EACH CLASS AND P-VALUES FROM KRUSKAL-WALLIS [22] (K-W) TESTS. DIFFERENT LOWERCASE LETTERS (WITHIN A COLUMN) REPRESENT SIGNIFICANT DIFFERENCES BETWEEN THE AVERAGES, ACCORDING TO THE METHOD OF MULTIPLE COMPARISONS [23] AT SIGNIFICANCE LEVEL α=0.05.

1. **if** SL **is** ABOVE_22.6 **and** LD **is** BETWEEN_34.4_52.3 **and** SD **is** BELOW_15.3 **then**
   class = *Dactylopterus volitans* (CF = 0.93)
2. **if** SL **is** BETWEEN_22.9_23.5 **then**
   class = *Dactylopterus volitans* (CF = 0.9)
3. **if** SL **is** BETWEEN_22.6_22.9 **then**
   class = *Dactylopterus volitans* (CF = 0.69)
4. **if** WL **is** ABOVE_46.8 **and** LD **is** BELOW_64.8 **then**
   class = *Engraulis anchoita* (CF = 0.99)
5. **if** SL **is** BETWEEN_22.7_23.3 **then**
   class = *Engraulis anchoita* (CF = 0.95)
6. **if** SL **is** BELOW_22.7 **then**
   class = *Engraulis anchoita* (CF = 0.87)
7. **if** LD **is** BETWEEN_70.5_84 **and** SL **is** ABOVE_23.5 **then**
   class = *Trachurus lathami* (CF = 0.93)
8. **if** SL **is** BELOW_25.1 **and** LD **is** BELOW_47.2 **and** WL **is** ABOVE_46.4 **then**
   class = *manjuba* (CF = 0.98)
9. **if** LD **is** BELOW_22 **then**
   class = *manjuba* (CF = 0.93)
10. **if** SD **is** BELOW_24.7 **and** WL **is** BELOW_46.4 **then**
    class = *others* (CF = 0.94)
11. **if** WL **is** BELOW_46.4 **and** SL **is** BETWEEN_24.2_24.6 **then**
    class = *others* (CF = 0.8)
12. **if** SL **is** BETWEEN_24.2_24.6 **then**
    class = *others* (CF = 0.86)
13. **if** ID **is** ABOVE_27.3 **and** SL **is** ABOVE_26.9 **then**
    class = *Trichiurus lepturus* (CF = 0.96)
14. **if** SL **is** ABOVE_27.7 **then**
    class = *Trichiurus lepturus* (CF = 0.7)

Fig. 8. The set of rules generated from FURIA.

Class	Algorithm		
	C4.5	FARC-HD	FURIA
Dactylopterus volitans	0.710	0.875	0.921
Engraulis anchoita	0.982	0.979	0.988
Trachurus lathami	0.820	0.981	1.000
Trichiurus lepturus	0.977	0.988	0.977
manjuba	0.962	0.971	0.980
others	0.708	0.907	0.916

TABLE III

F-MEASURES BY CLASS FOR ALL ALGORITHMS CONSIDERED.

Results from fishing hauls revealed schools composed majorly by pelagic fish from four species: *Engraulis anchoita* (Argentine anchovy), *Dactylopterus volitans* (flying gurnard), *Trachurus lathami* (rough scad) and *Trichiurus lepturus* (largehead hairtail). The class manjuba refers to fish from the family *Engraulidae*, other than *E. anchoita* which were aggregated under this name. The class "others" refers to schools without a predominant species, where the species with the higher frequency represents under 50% of the total of individuals. Also, schools of other species were categorized under this class.

Table I describes the attributes that represent the descriptors in the dataset. Average values of each attribute for each class are shown in Table II. P-values resulting from Kruskal-Wallis [22] tests for the difference on the means for each attribute are also shown. Significative differences between classes were detected for all attributes considered. Table II illustrates the pertinency of each class to each group at a significance level α=0.05, after the method of multiple comparisons [23] which reveals that classes are well-separated. The class *Trichiurus lepturus* presented higher LD and also higher SL and WL (i.e., sampling points which are away from the coast). Both *Trichiurus lepturus* and *Dactylopterus volitans* classes presented higher SW and SH. A total of 408 schools were identified in classes, with the following distribution: 165 instances of *Engraulis anchoita*, 33 of *Dactylopterus volitans*, 26 of *Trachurus lathami*, 43 of *Trichiurus lepturus*, 102 of the class "manjuba" and 39 instances of the class "others".

The evaluation of the results from the three algorithms C4.5, FARC-HD and FURIA is performed after 5-fold cross-validation technique is applied [24]. The average accuracies and F-measures [25] are obtained for each algorithm. Parameter setting for the algorithms adopted was performed as follows. The parameter of C4.5 which regulates the minimum number of instances per leaf was set to 5. We set the number of lingustic variables for the fuzzy representation of each variable to 6 in FARC-HD. The number of optimizations was set to 1 and the number of folds to 4 in FURIA.

V. RESULTS

The average accuracies from cross-validation computed for C4.5, FURIA FARC-HD are 92.2%, 96.1%, and 97.8% respectively. Table III shows the average F-measures by class for the three algorithms considered. The higher values were obtained from the FURIA algorithm for all classes, except for *Trichiurus lepturus*.

Figure 6 shows the 9 rules generated by C4.5. The attributes LD (which is the root of the tree) and SL were the most frequent, appearing in 9 and 6 of the rules, respectively.

From the 16 rules generated by FARC-HD (Figure 7), the most relevant attributes are SL, WL, and LD, which are related to the description of the location where the school was detected. Triangular membership functions are generated

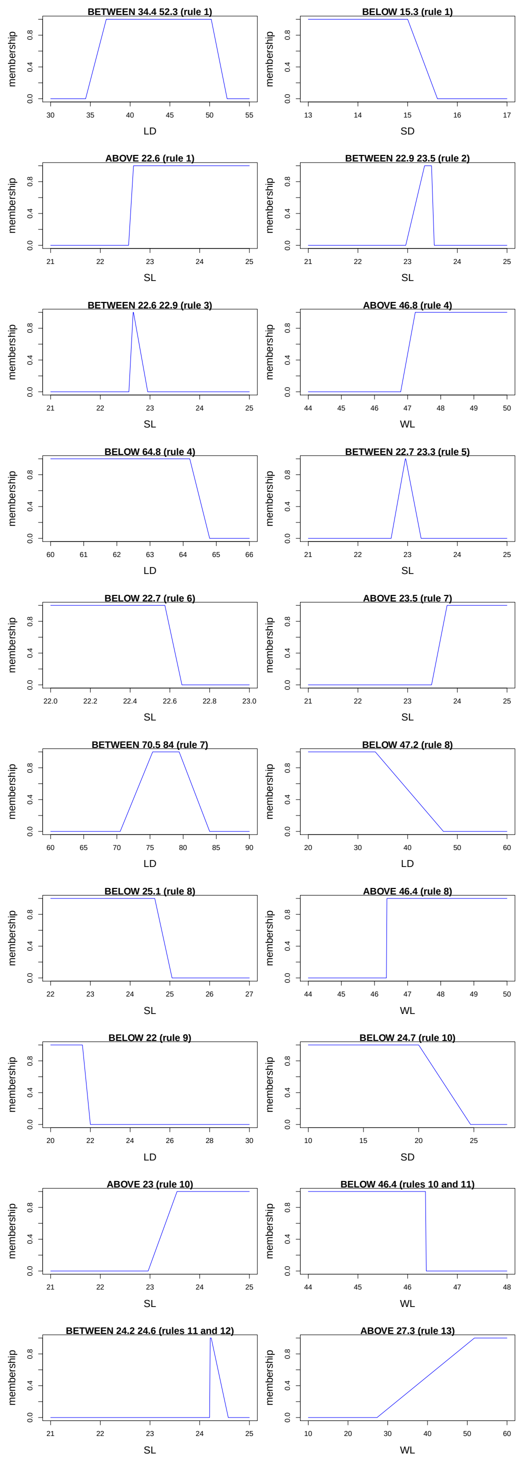

Fig. 9. Membership functions generated from FURIA.

by FARC-HD for each variable[3]. FURIA generated 14 rules (Figure 8), where 12 of them adopted the attribute SL. Other frequent attributes are SD and WL, similarly to FARC-HD. The membership functions corresponding to the rules generated from FURIA are shown in Figure 9.

The rules were analyzed by specialists, which pointed out that results from FURIA are especially coherent with the actual behavior, noticeably with respect to the spatial distribution shown by the corresponding species. The classes "manjuba" and "others" were harder to be correctly classified by the FURIA algorithm. The statistical analysis of those two classes indicated that they differed only in terms of LD (lower in "manjuba") and WL (lower in "others"). This might result from the composition of both classes, which are formed by mixed schools (with more than one species), whose biological needs and ecological interactions can generate greater variability in their spatiotemporal distributions and in the morphology of schools in the water column.

The attributes SL and LD were revealed as very important for the composition of the rules. Those descriptors are directly related to other abiotic (eg, temperature and water salinity) and biotic (food availability) factors that influence the places where these marine beings inhabit [26], [27]. According to [27], depth is the main structuring factor of marine megafauna communities (macroscopic animals).

Morphological descriptors SH and SW were not relevant for the rule's composition. This results from a relatively small variability of the way fish schools are formed in this region, which do not vary much between species. In general the schools had an average height of 2.4 to 3.1 meters and average width between 4.4 to 8 meters, except the schools of *Dactylopterus volitans* and *Trichiurus lepturus* which presented greater sizes.

VI. Conclusion

The adoption of fuzzy rules was revealed as an efficient approach for the classification of hydroacoustic records of fishing hauls. All F-measure values obtained from the FURIA algorithm in the classification of the records are above 0.91. The superiority of FURIA over other fuzzy classifiers has been reported in the literarure [15], [28], notably under the case of unbalanced classes. This can be noticed from our case since FURIA was able to deliver the highest overall F-measure from the class which has the lowest number of instances (*Trachurus lathami*). According to [28] this behavior results from the fact that the accuracy of FURIA is not restricted by the choice of a linguistic partition, and the antecedents of the rules change dynamically when an instance appears that is not covered by the rule. The replacement of rules is performed by its minimal generalizations [29].

Future work should consider higher number of species, other locations, and the differentiation on the periods of the year. The methodology could be applied to other types of hydroacoustic data. Also, fuzzy rule-based classification systems based on the generalization of the Choquet integral can also be applied, as in [30]–[33]

[3] A nominal representation is proposed for each variable as follows. The lingustic variables for the WL attribute are: MuchToWest, RoughlyToWest, CenterToWest, CenterToEast, RoughlyToEast, MuchToEast. The linguistic variables for the SD attribute are: VeryShallow, RoughlyShallow, AvgToShallow, AvgToDeep, RoughlyDeep, VeryDeep. The other variables are represented similarly.

REFERENCES

[1] L. S. P. Madureira, M. P. Pinho, S. C. Weigert, J. L. Coletto, P. L. Costa, and R. C. Valderrama, "Fishing up and down the marine food web with hydroacoustics. in: Acoustics in underwater geosciences symposium," *IEEE/OES RIO Acoustics*, vol. 8, no. 3, pp. 1 – 5, 2015.

[2] J. K. Horne, "Acoustic approaches to remote species identification: a review," *Fisheries oceanography*, vol. 9, no. 4, pp. 356–371, 2000.

[3] J. Simmonds and D. N. MacLennan, *Fisheries acoustics: theory and practice*. John Wiley & Sons, 2008.

[4] G. Rose and W. Leggett, "Hydroacoustic signal classification of fish schools by species," *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 45, no. 4, pp. 597–604, 1988.

[5] V. Trygonis, S. Georgakarakos, and E. J. Simmonds, "An operational system for automatic school identification on multibeam sonar echoes," *ICES journal of Marine Science*, vol. 66, no. 5, pp. 935–949, 2009.

[6] R. Hewitt, "Developments and use of sonar mapping for pelagic stock assessment in the california current," *Fisheries Bulletin*, vol. 74, pp. 281–300, 1976.

[7] R. Lefort, R. Fablet, and J.-M. Boucher, "Weakly supervised learning: application to fish school recognition," in *New Advances in Intelligent Signal Processing*. Springer, 2011, pp. 203–223.

[8] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[9] J. Uranga, H. Arrizabalaga, M. C. Hernandez, I. Onandia, G. Boyra, and J. Santiago, "Detecting the presence-absence of north atlantic albacore tuna by automated analysis of medium-range sonars on fishing vessels," *Collect. Vol. Sci. Pap. ICCAT*, vol. 76, no. 8, pp. 75–84, 2020.

[10] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52 138–52 160, 2018.

[11] R. J. Wootton, *Fish ecology*. Springer Science & Business Media, 1991.

[12] S. C. Weigert and L. S. P. Madureira, "Registros acústicos biológicos detectados na zona econômica exclusiva da região nordeste do brasil–uma classificação em ecotipos funcionais," *Atlântica, Rio Grande*, vol. 33, no. 1, pp. 15–32, 2011.

[13] J. R. Quinlan, *C4. 5: programs for machine learning*. Elsevier, 2014.

[14] S. Ruggieri, "Efficient c4. 5 [classification algorithm]," *IEEE transactions on knowledge and data engineering*, vol. 14, no. 2, pp. 438–444, 2002.

[15] J. Hühn and E. Hüllermeier, "FURIA: an algorithm for unordered fuzzy rule induction," *Data Mining and Knowledge Discovery*, vol. 19, no. 3, pp. 293–319, 2009.

[16] H. Ishibuchi, *Classification and Modeling with Linguistic Information Granules: Advanced Approaches to Linguistic Data Mining*. Berlin, Heidelberg: Springer-Verlag, 2009.

[17] J. Alcala-Fdez, R. Alcala, and F. Herrera, "A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning," *IEEE Transactions on Fuzzy Systems*, vol. 19, no. 5, pp. 857–872, 2011.

[18] W. W. Cohen, "Fast effective rule induction," in *Machine Learning proceedings 1995*. Elsevier, 1995, pp. 115–123.

[19] L. Verma, S. Srivastava, and P. Negi, "A hybrid data mining model to predict coronary artery disease cases using non-invasive clinical data," *Journal of medical systems*, vol. 40, no. 7, p. 178, 2016.

[20] G. Asencio-Cortés, E. Florido, A. Troncoso, and F. Martínez-Álvarez, "A novel methodology to predict urban traffic congestion with ensemble learning," *Soft Computing*, vol. 20, no. 11, pp. 4205–4216, 2016.

[21] C. L. D. B. Rossi-Wongtschowski, A. M. Vaz-dos Santos, and C. C. Siliprandi, "Checklist of the marine fishes collected during hydroacoustic surveys in the southeastern brazilian bight from 1995 to 2010." *Arquivos de Zoologia*, pp. 45: 73–88, 2014.

[22] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-criterion variance analysis," *Journal of the American Statistical Association*, vol. 47, no. 260, pp. 583–621, 1952.

[23] S. Siegel and J. J. Castellan, *Nonparametric Statistics for the Behavioral Sciences (2nd ed.)*. Mcgraw-Hill Book Company, 1988.

[24] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 36, no. 2, pp. 111–133, 1974.

[25] D. M. Powers, "Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, 2011.

[26] C. M. Lalli and T. R. Parsons, *Biological oceanography: an introduction*. Amsterdam: Elsevier, 1997.

[27] P. A. S. d. Costa, M. M. Mincarone, A. d. C. Braga, A. S. Martins, H. P. Lavrado, M. Haimovici, and A. P. d. C. Falcão, "Megafaunal communities along a depth gradient on the tropical Brazilian continental margin," *Marine Biology Research*, vol. 11, no. 10, pp. 1053–1064, 2015.

[28] A. Palacios, K. Trawiński, O. Cordón, and L. Sánchez, "Cost-sensitive learning of fuzzy rules for imbalanced classification problems using furia," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 22, no. 05, pp. 643–675, 2014.

[29] M. Eineborg and H. Boström, "Classifying uncovered examples by rule stretching," in *International Conference on Inductive Logic Programming*. Springer, 2001, pp. 41–50.

[30] G. P. Dimuro, G. Lucca, B. Bedregal, R. Mesiar, J. A. Sanz, C.-T. Lin, and H. Bustince, "Generalized cf1f2-integrals: From choquet-like aggregation to ordered directionally monotone functions," *Fuzzy Sets and Systems*, vol. 378, pp. 44 – 67, 2020, theme : Aggregation Operations.

[31] G. P. Dimuro, J. Fernández, B. Bedregal, R. Mesiar, J. A. Sanz, G. Lucca, and H. Bustince, "The state-of-art of the generalizations of the choquet integral: From aggregation and pre-aggregation to ordered directionally monotone functions," *Information Fusion*, vol. 57, pp. 27 – 43, 2020.

[32] H. Bustince, R. Mesiar, J. Fernandez, M. Galar, D. Paternain, A. Altalhi, G. Dimuro, B. Bedregal, and Z. Takáč, "d-choquet integrals: Choquet integrals based on dissimilarities," *Fuzzy Sets and Systems*, 2020.

[33] G. Lucca, G. P. Dimuro, J. Fernández, H. Bustince, B. Bedregal, and J. A. Sanz, "Improving the performance of fuzzy rule-based classification systems based on a nonaveraging generalization of CC-integrals named $C_{F_1 F_2}$-integrals," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 1, pp. 124–134, Jan 2019.