Fisheries Research 194 (2017) 42–49

Contents lists available at ScienceDirect

Fisheries Research

journal homepage: www.elsevier.com/locate/fishres

Full length article

A 'machine learning' technique for discriminating captive-reared from wild Atlantic bluefin tuna, *Thunnus thynnus* (Osteichthyes: Scombridae), based on differential fin spine bone resorption

Francesco Campobasso[a], Annarita Fanizzi[b], Giambattista Bello[c], Nicoletta Santamaria[c], Aldo Corriero[c,*]

[a] Department of Economics and Mathematics, University of Bari Aldo Moro, Largo Abbazia S. Scolastica, 70124 Bari, Italy
[b] I.R.C.C.S. Istituto Tumori 'Giovanni Paolo II', Viale Orazio Flacco 65, 70124 Bari, Italy
[c] Department of Emergency and Organ Transplantation, Section of Veterinary Medicine and Animal Production, University of Bari Aldo Moro, S.P. per Casamassima km. 3, 70010 Valenzano, BA, Italy

ARTICLE INFO

Keywords:
Atlantic bluefin tuna
Tuna farming
Spine bone resorption
Supervised classification
Random forest

ABSTRACT

The Atlantic bluefin tuna (ABFT) fishery is regulated by the International Commission for the Conservation of Atlantic Tunas (ICCAT), which establishes the allowable annual yield and the minimum capture size, and allocates capture quotas to the Contracting Parties. Despite fishery monitoring, a considerable amount of captures escapes ICCAT control. In the Mediterranean Sea, the purse seine fishery supports ABFT farming, a capture-based aquaculture activity that involves catching fish from the wild and rearing them in sea cages for a few months. The first spine of the cranial dorsal fin undergoes a continuous bone remodeling process consisting in old bone (primary bone) resorption and new bone (secondary bone) apposition. A marked increase of spine bone resorption was shown in captive-reared ABFT with respect to wild specimens. In this paper, the Random Forest (RF), a Computer Aided Detection system, was applied to distinguish captive-reared from wild ABFT based on fish age, fish fork length, total surface of spine cross section, and surface of remodeled bone tissue in the spine cross section (sum of reabsorbed bone tissue and secondary cancellous bone). The RF system was also compared to the Logistic Regression method (LR). The percentages of properly classified animals, either wild or captive-reared, with respect to the overall number of animals, i.e. *accuracy*, was 95.3 ± 2.6% and 79.0 ± 5.1% for RF and LR, respectively. The percentages of the properly classified captive-reared specimens, i.e. *sensitivity*, were 93.5 ± 3.1% and 75.8 ± 5.3% for RF and LR, respectively. The percentages of the properly classified wild specimens was 96.7 ± 2.2% and 81.4 ± 4.9%, for RF and LR, respectively. The proposed technique appears to be a reliable investigation tool anytime the suspicion arises that illegally caught ABFT are sold as aquaculture products.

1. Introduction

The Atlantic bluefin tuna (ABFT), *Thunnus thynnus* (Linnaeus, 1758) (Osteichthyes: Scombridae), is a very highly priced tuna species, which is subjected to an intense fishing pressure. The fishery of ABFT is regulated by the International Commission for the Conservation of Atlantic Tunas (ICCAT). This recognizes two different ABFT stocks, the western Atlantic stock and the eastern Atlantic stock; the geographical range of the latter includes the Mediterranean Sea. The two stocks are individually managed by ICCAT, according to specific regulations (ICCAT recommendations and resolutions are available at http://iccat.int/en/RecsRegs.asp). The ICCAT establishes the maximum annual yield in terms of fish biomass (total allowable catches; TACs) and retains the option to suspend all the ABFT fisheries in the event that any periodical stock assessment detects a serious threat of fishery collapse. The TACs are distributed among the ICCAT Contracting Parties (Countries that adhere to ICCAT), so that each of them is assigned a fishing quota. Specific ICCAT Observer Programs aimed at monitoring the amount of catches by fishing vessels and traps are currently implemented (https://www.iccat.int/en/ROPbft.htm). Since 1990s, despite more and more management measures have been introduced, illegal, unreported and unregulated tuna fishing practices have increased (Miyake et al., 2004).

In March 2010, the Convention on International Trade in

* Corresponding author.
E-mail address: aldo.corriero@uniba.it (A. Corriero).

http://dx.doi.org/10.1016/j.fishres.2017.05.008
Received 7 March 2016; Received in revised form 12 May 2017; Accepted 15 May 2017

Endangered Species of Wild Fauna and Flora (CITES) discussed the proposal by the Principality of Monaco to include ABFT in the Appendix I (http://www.cites.org/eng/cop/15/prop/E-15-Prop-19.pdf), which lists threatened species whose international trade is prohibited. The proposal was not adopted by the conference because of the opposition from some nations. The inclusion of ABFT in CITES Appendix A would have seriously affected the economic sustainability of the fisheries, mainly the purse seine fishery, whose activity fuels the international ABFT market. In fact, the majority of purse seine catches is destined for the Japanese market (Mylonas et al., 2010).

In the Mediterranean Sea, the purse seine fishery supports ABFT fattening and farming aquaculture activities, which are exclusively based on fish captured from the wild. Presently no commercial complete life cycle ABFT aquaculture exists. In Mediterranean tuna farms, wild-caught fish are reared for periods ranging from 3 months to 2 years (for a review on tuna farming and fattening see Mylonas et al., 2010). Rearing of adult ABFT individuals for a few months (usually 3–7 months) is classified as 'fattening' and aims at increasing the fish fat mass through a diet based on small, high fat content pelagic fish. In addition, tuna fattening allows a controlled input in the market, avoiding flooding it with ABFT during its short fishing term, which is instrumental in keeping ABFT quotations high. The tuna farming activity is authorized by ICCAT in the Adriatic Sea (Croatia) only and involves the capture of immature individuals and their rearing in captivity for up to 2 years (Ticina et al., 2007).

Since early 2000, an intense research effort aimed at transforming tuna fattening and farming activities in self-sustained aquaculture has been undertaken (Berkovich et al., 2013; Corriero et al., 2009, 2007; De Metrio et al., 2010; Micera et al., 2010; Mylonas et al., 2007; Pousis et al., 2012, 2011; Rosenfeld et al., 2012). Under the stimuli of this scientific activity, carried out in the framework of different EU, national and regional projects, a true ABFT aquaculture industry is being developed so that, in December 2014 the first ABFT individuals born and grown out in captivity in a Spanish fish farm were placed on the market (cf. newspaper La Verdad, Murcia, Spain, 06 December 2014). Moreover, in very recent times (July 2016), the closure of the ABFT life cycle with the production of F1 generation in captivity has been achieved (announcement of the Spanish Institute of Oceanography; http://www.mispeces.com/nav/actualidad/noticias/noticia-detalle/El-IEO-cierra-el-ciclo-biolgico-del-atn-rojo-Atlntico-en-cautividad/#.V53AOriLTy2). The ABFT fully produced in captivity, which are morphologically identical to wild individuals, are not managed by ICCAT and, therefore, their production is outside the quota allowance system. Hence, once the ABFT self-sustained commercial aquaculture will be eventually established, there will be the real possibility that illegally caught ABFT might be smuggled into the market as individuals born in captivity, since no tools to distinguish between wild and reared ABFT are presently available.

The ABFT is provided with median (dorsal and anal) and paired (pectoral and pelvic) fins. The dorsal cranial fin is supported by 12–15 spiny rays (spines), the caudal dorsal one is supported by a spine followed by 11–13 soft rays (rays) (Tortonese, 1975). The first spine of the cranial dorsal fin is used for age determination since its transverse sections displays well-defined growth marks, called annuli, which are interpreted as periodic events (Corriero et al., 2005; Luque et al., 2014; Santamaria et al., 2015, 2009). Growth marks are caused by the progressive apposition of bone tissue on the external surface of the spine, which becomes apparent as an ordered series of alternate opaque and translucent rings, corresponding to a faster spring-summer and a slower autumn-winter apposition, respectively, which parallels body growth (Cort, 1991; Megalofonou and De Metrio, 2000; Santamaria et al., 2009). The ABFT first dorsal spine bone tissues undergo dynamic processes: while new compact bone is added on the spine outer surface, in its inner part (the so-called core or nucleus) a physiological progressive bone resorption occurs (Cort, 1991; Megalofonou and De Metrio, 2000; Santamaria et al., 2009, 2015). In a recent paper, Santamaria et al. (2015) assessed the spine bone apposition and resorption in both wild (aged 1–13 years) and captive-reared (aged 2–11 years) Mediterranean ABFT. They reported that: a) the spine section surface grows isometrically with respect to body size; b) the fraction of spine compact bone progressively decreases with both fish size and age; c) the phenomenon of spine bone resorption is dramatically enhanced in captive-reared ABFT individuals with respect to wild animals.

The aim of the present paper was to set up a method to discriminate between wild and captive-reared ABFT by means of a Computer Aided Detection (CAD) system trained to recognize the fish origin on the basis of a number of parameters related to spine bone resorption. The present 'machine learning technique', the so-called Random Forest (Breiman, 2001), as applied to the spine-related data, was compared to a classical technique of supervised classification, i.e. the Logistic Regression, in order to both evaluate its strength and assess the feasibility of its adoption for practical purposes. In addition, the most discriminating variables among the tested parameters were identified, so to warrant a more effective and straightforward use of the herein proposed method.

2. Material and methods

2.1. Sampling and spine measurements

The ABFT specimens which provided the data for the present study are the same used in Santamaria et al. (2015). The sampling procedure is concisely reported below; further details are reported in Santamaria et al. (2015). In all 428 ABFT specimens (186 wild and 242 captive-reared) were sampled over the eight-year period 2003–2010 in several Mediterranean sites (Fig. 1). Wild fish were caught by commercial vessels whereas captive-reared specimens were sampled in the framework of the following research projects: EU project REPRODOTT, EU project SELFDOTT and Italian project ALLOTUNA funded by the regional government of the Apulia region. The captive-reared ABFT were in fact wild-born, that is collected from the sea and kept in rearing cages for a period ranging from a few months to three years. The fish fork length, *FL*, was measured to the nearest cm. The first spine of the cranial dorsal fin was removed (Fig. 2a) from the fish and processed in the laboratory. A low speed diamond saw (Buehler, Isomet) was used to cut it transversally at a distance of half the maximum spine diameter from the condyle, according to the spine sectioning standard procedure (Luque et al., 2014), and obtain a 0.7 mm thick cross-section (Fig. 2b). An age (*AGE*) was assigned to each fish according to Santamaria et al. (2009). The following measurements were taken on each spine section (Fig. 2c):

SD, spine diameter (= maximum transverse diameter of the spine section);

TS, total surface;

RS, reabsorbed part surface (= surface of reabsorbed bone + surface of remodeling bone);

CT, compact bone thickness (= maximum thickness of the spine compact bone layer).

TS and *RS* are the same raw data used in Santamaria et al. (2015); *SD* and *CT* are new measurements, i.e. previously unreported.

The measurements were taken on spine section images, using an interactive function (i.e. measurements of operator-selected surfaces by a specific image analysis software function), by means of the image analysis software Quantimet 500 W (Leica, Wetzlar, Germany).

A table reporting all measurements is available in Supplementary material.

2.2. Supervised classification methods

2.2.1. The Random Forest classifier

The raw data deriving from the measurements were used to develop a predictive system to detect the fish origin, either wild or captive-

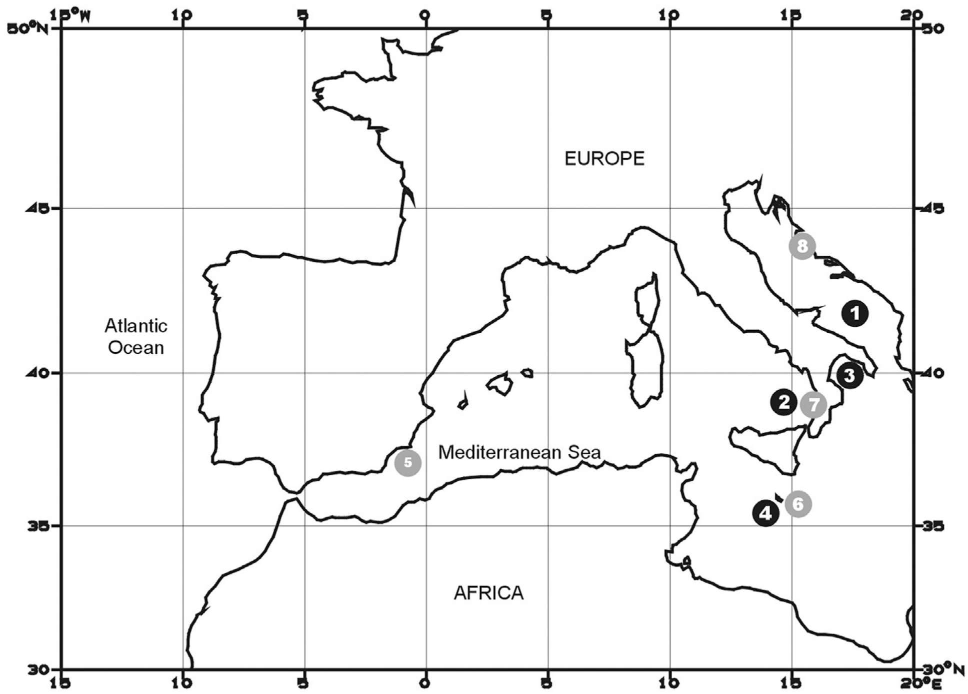

Fig. 1. Geographical location of Atlantic bluefin tuna sampling areas. Black and grey circles indicate sampling sites for wild and captive-reared specimens, respectively. 1, South Adriatic Sea; 2, South Tyrrhenian Sea; 3, North Ionian Sea (Gulf of Taranto); 4, Ionian Sea around Malta; 5, Puerto de Mazarrón and Cartagena, Spain; 6; Malta; 7, Vibo Marina, Italy; 8, Drvenik and Uglyan Island, Croatia.

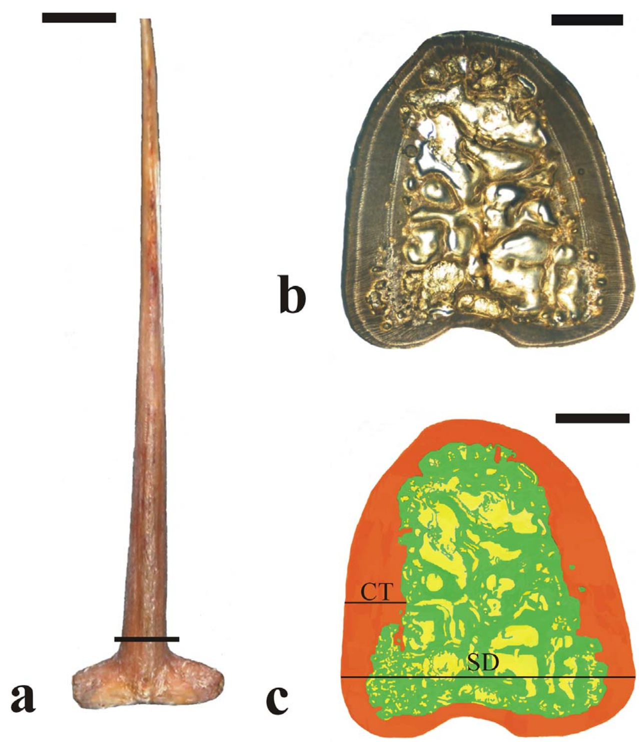

Fig. 2. a) Front view of the first ray of the first dorsal fin from a captive-reared Atlantic bluefin tuna, fork length = 166 cm. The line indicates the sectioning level above the condyle. b) Cross section of the spine showing its structure consisting of an external compact bone and an internal zone undergoing a progressive bone resorption. c) Schematic rendering of the spine cross section showing the parameters for the CAD (Computer Aided Detection) system. The total surface (*TS*) corresponds to the whole spine section surface. The solid part surface (*SS*) corresponds to the red-colored spine surface (compact bone). The reabsorbed part surface (*RS*) is the sum of the totally reabsorbed bone surface (yellow-colored) and the remodeling bone surface (green-colored). *CT*, compact bone thickness. *SD*, spine diameter. Magnification bars = 1.8 cm in (a) and 2 cm in (b) and (c).

reared, based on the Random Forest (RF) classifier. This is a well-known ensemble supervised method, consisting of a set of stochastic classifiers that operate by constructing k decision trees opportunely trained on the available data. Each tree develops an independent own classification for each new unit and the final classification is obtained by a majority vote of the individual trees (Breiman, 2001).

The RF is characterized by a generalization error that converges as the free parameter k becomes larger (Breiman, 1996). Generally, a few hundred to several thousand trees are used, depending on the size and nature of the training set. An optimal number of trees can be found by observing the out-of-bag error, which is the mean prediction error on each training sample x_i, using only those trees that did not have x_i in their bootstrap sample (in RF, bootstrap samples, i.e. random samples with replacement, have the same numerosity of the training data set) (Gareth et al., 2013). The training and test error tend to level off after some number of trees have been fit.

The use of a random selection of features to train each tree allows achieving performances comparable to those obtained by other classifiers, such as AdaBoost, but more robust with respect to noise (Díaz-Uriarte and de Andrés, 2006). This methodology allows managing a large number of features without the need to reduce them, and maintains its efficiency even in case of missing data and in the presence of outliers. Thus, it is inferred that it is a classification algorithm suitable for biological data analysis (Strobl et al., 2008).

In the present case, the classifier was trained to distinguish captive-reared from wild ABFT specimens using the above-mentioned six variables (*AGE, FL, TS, RS, SD, CT*). In order to assess the effectiveness of the RF method, its results were compared with those obtained by the classical Logistic Regression (LR), a generalized linear model to predict a logit transformation of the probability of presence of the characteristic of interest (Hosmer et al., 2013).

Subsequently, the subset of the available variables that showed the highest discriminating capability was identified by means of univariate statistical tests independent from the chosen classification technique. This way, the CAD implementation was simplified in order to become suitable for applicative purposes. As for the use of ROC curves in feature selection, several papers have been published (Caruana and Freitag, 1994; Landgrebe and Duin, 2007; Mamitsuka, 2006; Marrocco et al., 2008). A feature selection method to reduce the dataset size is to detect variables whose AUC value is statistically higher than 0.5, that is unable to discriminate the two (Hanley and McNeil, 1983). Moreover, the suitability of the selected variables was corroborated by the Wilcoxon-Mann-Whitney test (Mann and Whitney, 1947). To this end, a second RF was trained by using the selected subset of variables and its results were again compared to those of the corresponding LR.

2.2.2. Evaluation of classification performances

The performances of both the RF and LR classifiers were evaluated *accuracy*-, *sensitivity*- and *specificity-wise* by means of the Area Under the Curve (AUC) value of the *Receiver Operating Characteristic* (ROC) curves (Bradley, 1997). In particular, once the threshold between 0 and 1 was assigned to each animal by the classifiers, the outputs were provided as binary classification. This was obtained after the identification of the best cut-off that maximized the difference between true positives (*sensitivity*) and false positives (1—*specificity*) by means of the Youden test (Youden, 1950), which maximizes the difference between true positives and false positives.

Specificity and *sensitivity* were the percentages of the properly classified captive-reared and wild ABFT, respectively; *accuracy* was the percentage of properly classified animals with respect to the overall animals in the dataset.

The validation of the models was performed by 10-fold cross-validation (McLachlan et al., 2004). In 10-fold cross-validation, the original sample is randomly partitioned into 10 equal size subsamples. Out of the *k* subsamples, a single subsample is retained as the validation data for testing the model, and the remaining 9 subsamples are used as training data. The cross-validation process is then repeated 10 times (the folds), with each of the 10 subsamples used exactly once as the validation data. The 10 results from the folds are averaged to produce a single estimation (McLachlan et al., 2004) (Fig. 3).

All the measures of performance, namely *accuracy*, *sensitivity* and *specificity*, were expressed as percentages, *p*. In general terms, it is possible to construct confidence intervals for all the performance measures in order to estimate *accuracy* (Witten et al., 2013). For a sufficiently large sample size, as in the present case, the estimator of the performance measure is normally distributed.

3. Results

When using all six variables, namely *AGE, FL, SD, TS, RS, CT*, the RF was trained by fixing a number of trees equal to 100, i.e. the value above which the error of classification of out-of-bag tended to stabilize at less than fluctuations of the random nature (Fig. 4). The ROC curves of classifications obtained by RF and LR are shown in Fig. 5. The two methods showed high classification *accuracy* in terms of AUC value, equal to 0.98 and 0.91 respectively, with an error of 0.01 and 0.02 (cf. Hanley and McNeil, 1982). Fig. 6 shows the confusion matrices of the binary classification for RF and LR. The RF classifier shows a higher performance than the LR technique, i.e. overall *accuracy* of classification = 92.5 ± 3.3% and 82.2 ± 4.8% for RF and LR respectively. Moreover, RF *sensitivity* in the detection of wild specimens was 93.0 ± 3.2% and its *specificity* in recognition of captive-reared specimens was 92.1 ± 3.4%.

At the 1% significance level, the features that showed AUC value higher than 0.5, hence most discriminating, were *AGE, FL, TS* and *RS*. At the 5% significance level, however, *SD* and *CT* were also significantly discriminating (Table 1). The results of the selection method proposed were corroborated by the non-parametric Wilcoxon-Mann-Whitney test (Table 2).

The ROC curves associated to the RF and LR classifiers implemented using the four most discriminating parameters, namely *AGE, FL, TS, RS*, are reported in Fig. 6. The AUC value for the LR decreased to 0.87 with an error of 0.02, indicating that it is still a moderately accurate classifier. The RF classifier showed to be highly accurate (AUC = 0.99 with an error of 0.01), better than that implemented by means of the whole set of variables. In particular, when selecting the best cut-off for each technique, the RF classifier showed an overall *accuracy* higher than that of the LR (95.3 ± 2.6% vs. 79.0 ± 5.1%) (Fig. 7); it also proved to be age/size independent, as shown by the lack of any significant correlation between the individual age class RF classifier accuracy and age class (correlation coefficient, $r = -0.158$; $df = 11$; $P > 0.05$).

The *sensitivity* and *specificity* for the RF classifier were 93.5 ± 3.1% and 96.7 ± 2.2%, respectively, which indicated a slightly higher capability of the classifier to identify captive-reared from wild specimens. By training the RF classifier just on the four most significant variables, the predictive power for both captive-reared and wild specimens increased to 95.1 ± 2.7% and 95.6 ± 2.6%, respectively (Fig. 8).

Considering a 95.3% overall accuracy of the RF classification model trained on the four most significant variables (Fig. 8i), the estimated minimum sample sizes necessary to achieve 1% significance and 5% precision levels was 119 specimens (Table 3).

4. Discussion

The increase of the spine bone resorption rates in captive-reared ABFT was clearly shown and quantitatively described by Santamaria et al. (2015). In the present work, the level of spine bone resorption, assessed through measures taken on spine sections, proved to be a suitable tool to distinguish wild from captive-reared specimens when tested by the RF 'machine learning technique'. Thanks to it, the most discriminating variables (i.e. age, fork length, total spine section surface and spine section reabsorbed surface) were sorted out to achieve an effort-effective applicability of the Computer Aided Detection (CAD) needing short computation times (in the order of a few minutes) in cases of few variables, as it is ours. Incidentally, as shown in Santamaria et al. (2015), the difference between spine bone resorption rates in wild and captive-reared individuals (lower in the former than in the latter) increases dramatically with fish ageing. This notwithstanding, the RF 'machine learning technique' proved to satisfactorily discriminate captive from wild individuals throughout their whole age/size range, as confirmed by the lack of any significant correlation between the accuracy of RF classifier and age. Moreover, the under-representation in either fish group (wild and captive-reared) of a few age classes did not affect the present RF technique thanks to the presence, in both groups data sets, of data for contiguous age classes, which gives high stability to the system. (This was proven by additional RF technique tests where the age classes under-represented in either fish group were removed from both groups; the additional tests results did not differ to any appreciable extent from the overall results.)

To sum up, the RF model performed better than the traditional LR technique. The observed differences in the outputs of the two methods are most probably due to their inherent characteristics. The decision boundaries of LR are defined by rigid linear functions of the input variables and therefore create a greater distortion in the classification, whereas the RF classifier is less sensitive to the correlation between the input variables and, thus, can provide more accurate estimates (Breiman, 2001). Moreover, the LR model shows a lower discriminative ability as the number of variables decreases.

Estimating the minimum sample size is a most relevant element when applying the RF classifier to solve actual feature selection problems, as in the present case. The estimated minimum sample size, i.e. 119 specimens, would allow the correct identification of the fish batch origin (either wild or captive-reared) at the 1% significance level and a maximum semi-width of the estimated confidence interval of 5%.

The Atlantic bluefin tuna aquaculture activity in the Mediterranean is likely destined to a rapid expansion thanks to technological developments in the fields of reproduction, larval rearing and grow out,

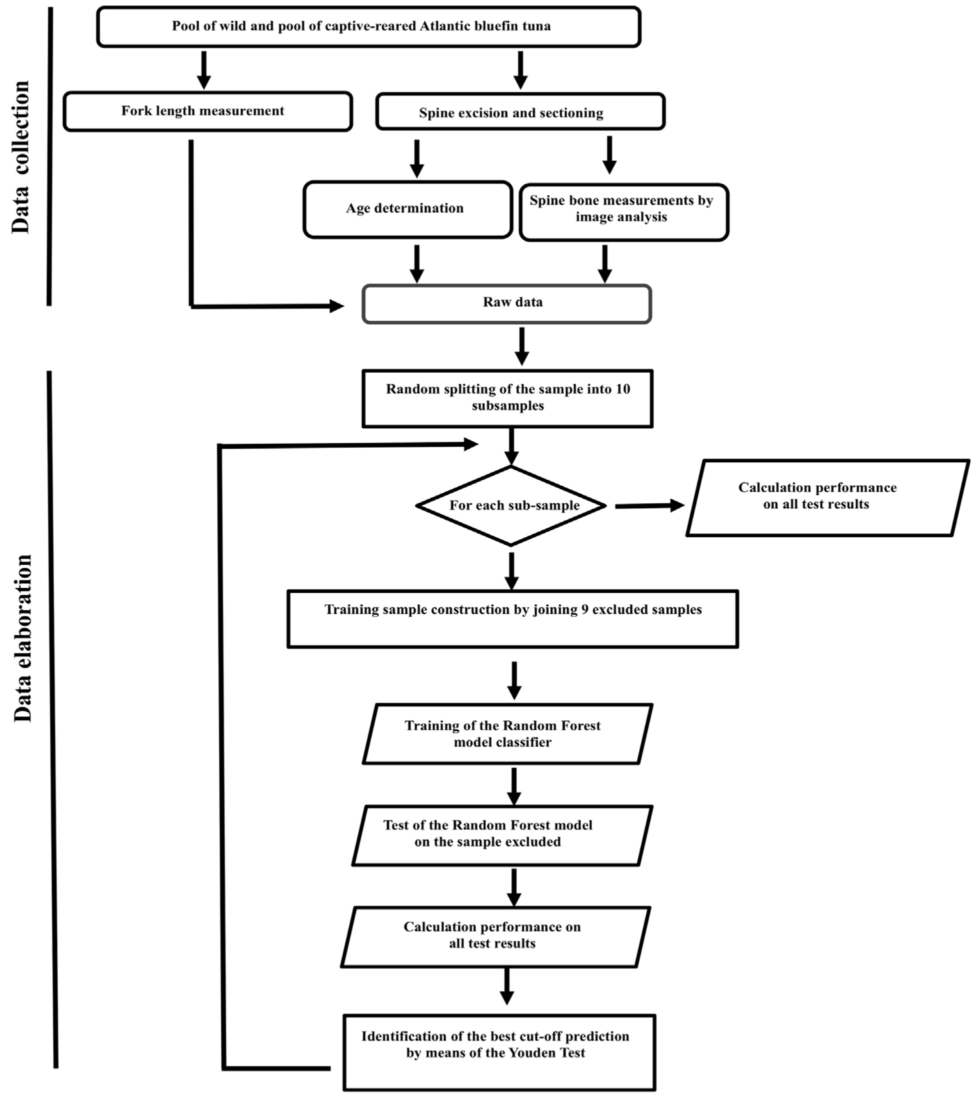

Fig. 3. Flow chart of the procedure to discriminate captive-reared from wild Atlantic bluefin tuna.

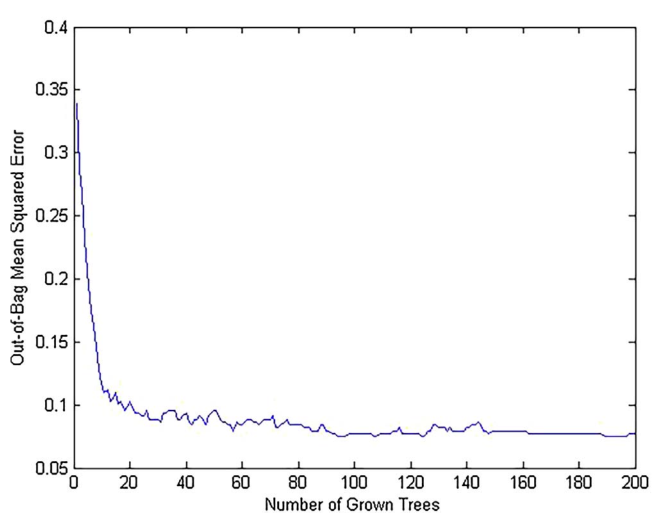

Fig. 4. Changes of the out-of-bag error according to the number of trees. The out-of-bag error is calculated as the mean prediction error on each training sample x_i, using only those trees that did not include x_i in their bootstrap sample.

which are fueled by EU and national grants. Hence the ABFT aquaculture activity will likely place on the market increasing product amounts. Obviously, fully aquaculture-produced fish, i.e. from eggs spawned by captive-reared individuals, will be available on the market in addition to ICCAT-controlled products. At the present, it is not readily possible to distinguish, among marketed ABFT, the wild product from that coming from either short or long term captivity in tuna farms following capture in the wild. Also, no method has been developed to date to identify the individuals entirely produced in captivity. The lack of any standardized procedure for detecting the ABFT origin will likely allow in the next future the fraudulent introduction, in the controlled market, of fish illegally caught to be sold as aquaculture products. This would bring, in turn, a further threat to an already heavily over-exploited fish stock.

The Atlantic bluefin tuna are generally reared in captivity in tuna farms for just a few months before slaughtering (Mylonas et al., 2010). In addition to material coming from this kind of farming activity, a part of the fish used in the present study belonged to broodstock reared for scientific purposes in the framework of different research projects (Corriero et al., 2007; De Metrio et al., 2010; Mylonas et al., 2007; Zupa et al., 2014, 2013), which had been kept in captivity for longer periods (1–3 years). (In fact, the analysis of spine bone resorption by Santamaria et al. (2015) did not aim at disclosing the relationship

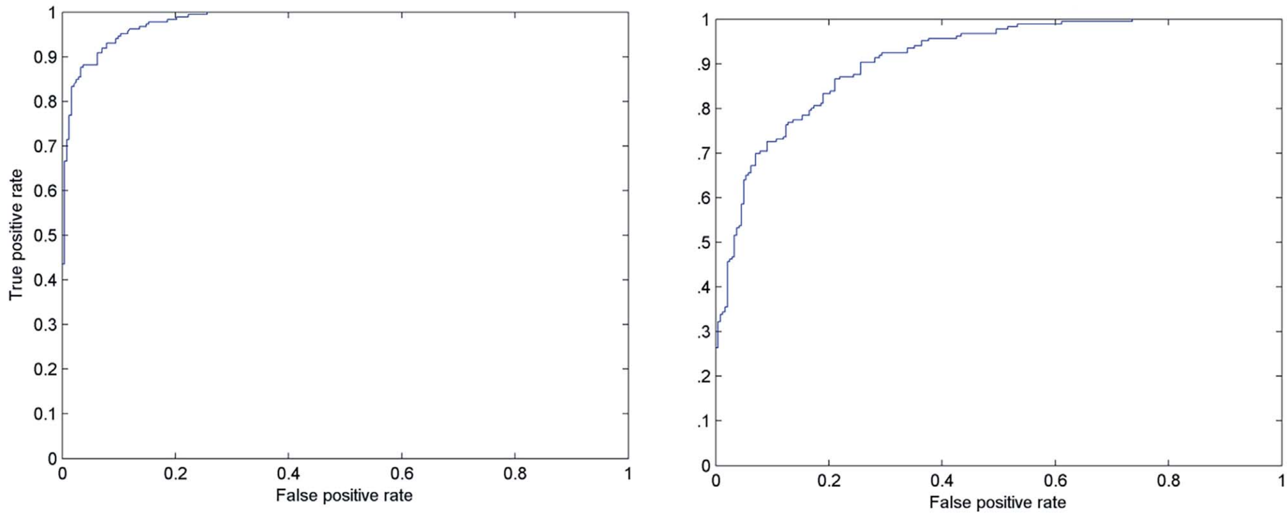

Fig. 5. ROC curves for RF (left) and LR (right) classifiers implemented with all the six available parameters.

between resorption rate and captivity time extent). Nevertheless, even the fish commercially reared in captivity for just few months showed rates of spine bone resorption comparable to those observed in the animals reared experimentally for longer periods. In other words, the sharp enhancement of spine bone resorption in captive-reared ABFT becomes visible after only few months of captivity.

To conclude, the CAD developed in the present work allowed distinguishing wild from captive-reared ABFT, thanks to the different rates of spine bone resorption. This technique might represent a useful investigation tool anytime the suspicion arises that illegally caught fish are sold as aquaculture products. It just requires the record of fish length, the collection of the first spine of the first dorsal fin (the spine customarily used for age determination studies), and a comparatively simple microscopic examination.

Acknowledgements

The data used in the present work come from Atlantic bluefin tuna sampled within the following research projects: EU 5th Framework Programme contract Q5RS-2002-01355 "Reproduction of the bluefin tuna in captivity – a feasibility study for the domestication of *Thunnus thynnus* (REPRODOTT)"; Italian Strategic Project of the Apulian Region PS_085 "Set up of an integrated system for bluefin tuna farming" (ALLOTUNA). EU 5th Framework Programme grant number 212797 "From capture based to SELF-sustained aquaculture and Domestication of bluefin tuna, *Thunnus thynnus*" (SELFDOTT).

The present study is a joint work carried out by all the authors. F.C. and A.C. designed the application of the Random Forest analysis to the ABFT spine bone resorption data. F.C. and A.F. carried out the development of CAD model and the statistical elaborations. N.S.

Table 1
AUC value and standard error of means (SEM) for each parameter used for the RF classification. Significance levels: * = $p < 0.05$; ** = $p < 0.01$.

	Fork length	Age	Spine diameter	Total surface	Reabsorbed surface	Compact bone thickness
AUC value	0.60**	0.59**	0.56*	0.61**	0.66**	0.44*
SEM	0.03	0.03	0.03	0.03	0.03	0.03

Fig. 6. Classification matrices for RF (left) and LR (right) classifiers implemented with all the six available parameters. Green boxes (a) and (e) show the absolute value and the 1% confidence interval relative to the percentage of correctly classified captive-reared and wild specimens, respectively. Red boxes (b) and (d) show the misclassified wild and captive-reared specimens, respectively. Grey boxes (c) and (f) show the confidence intervals of the predictive power of each of the two classes. Grey boxes show specificity (g), sensitivity (h) and the blue box accuracy (i). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Wilcoxon-Mann-Whitney test for each parameter used for the RF classification. Significance levels: * = $p < 0.05$; ** = $p < 0.01$.

	Fork length	Age	Spine diameter	Total surface	Reabsorbed surface	Compact bone thickness
Score test	−3.66**	−3.24**	−2.21*	−3.81**	−5.50**	−2.15*

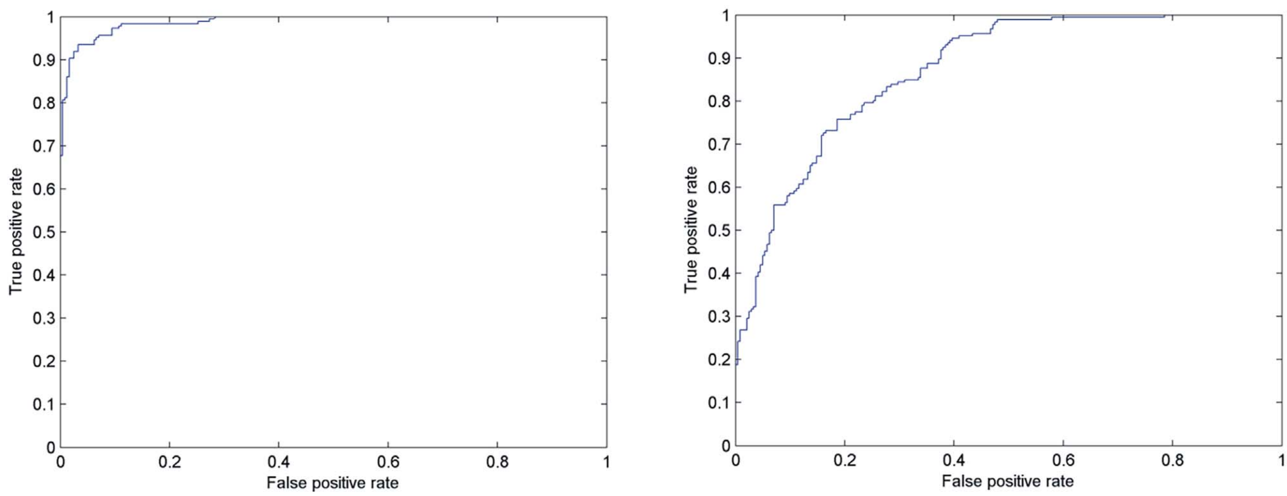

Fig. 7. ROC curves for RF (left) and LR (right) classifiers implemented with the four sorted out parameters.

Fig. 8. Classification matrices for RF (left) and LR (right) classifiers implemented with the four sorted out parameters. Green boxes (a) and (e) show the absolute value and the 1% confidence interval relative to the percentage of correctly classified captive-reared and wild specimens, respectively. Red boxes (b) and (d) show the misclassified wild and captive-reared specimens, respectively. Grey boxes (c) and (f) show the confidence intervals of the predictive power of each of the two classes. Grey boxes show specificity (g), sensitivity (h) and the blue box accuracy (i). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Sample size for different values of the significance level, α, and of the level of precision *d* of estimates.

	Percentage points	Significance level α		
		0.10	0.05	0.01
Half-width of the confidence interval d/2	3	135	191	330
	5	48	69	119
	10	12	17	30

measured spine diameter (SD) and compact bone thickness (CT). All the authors wrote the text.

Comments and suggestions of two anonymous reviewers improved the present paper.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fishres.2017.05.008.

References

Berkovich, N., Corriero, A., Santamaria, N., Mylonas, C.C., Vassallo-Aguis, R., de la Gándara, F., Meiri-Ashkenazi, I., Zlatnikov, V., Gordin, H., Bridges, C.R., Rosenfeld, H., 2013. Intra-pituitary relationship of follicle stimulating hormone and luteinizing hormone during pubertal development in Atlantic bluefin tuna (*Thunnus thynnus*). Gen. Comp. Endocr. 194, 10–23.

Bradley, A.P., 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Patte-rn Recogn. 30, 1145–1159.

Breiman, L., 1996. Bagging predictors. Mach. Learn. 24, 123–140.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Caruana, R., Freitag, D., 1994. How useful is relevance? In: Proceedings of the 1994 AAAI Fall Symposium on Relevance. AAAI Press Menlo Park. pp. 21–25.

Corriero, A., Karakulak, S., Santamaria, N., Deflorio, M., Addis, P., Desantis, S., Cirillo, F., Fenech-Farrugia, A., Vassallo-Agius, R., de la Serna, J.M., Oray, Y., Cau, A., De Metrio, G., 2005. Size and age at sexual maturity of female bluefin tuna (*Thunnus thynnus* L. 1758) from the Mediterranean Sea. J. Appl. Ichthyol. 21, 483–486.

Corriero, A., Medina, A., Mylonas, C.C., Abascal, F.J., Deflorio, M., Aragón, L., Bridges, C.R., Santamaria, N., Heinisch, G., Vassallo-Agius, R., Belmonte, A., Fauvel, C., Garcia, A., Gordin, H., De Metrio, G., 2007. Histological study of the effects of treatment with gonadotropin-releasing hormone agonist (GnRHa) on the reproductive maturation of captive-reared Atlantic bluefin tuna (*Thunnus thynnus* L.). Aquaculture 272, 675–686.

Corriero, A., Medina, A., Mylonas, C.C., Bridges, C.R., Santamaria, N., Deflorio, M., Losurdo, M., Zupa, R., Gordin, H., de la Gandara, F., Belmonte Rìos, A., Pousis, C., De Metrio, G., 2009. Proliferation and apoptosis of male germ cells in captive Atlantic bluefin tuna (*Thunnus thynnus* L.) treated with gonadotropin-releasing hormone agonist (GnRHa). Anim. Reprod. Sci. 116, 346–357.

Cort, J.L., 1991. Age and growth of the bluefin tuna. ICCAT Coll. Vol. Sci. Pap., vol. 35. pp. 213–230.

Díaz-Uriarte, R., de Andrés, S.A., 2006. Gene selection and classification of microarray data using random forest. BMC Bioinform. 7, 3.

De Metrio, G., Bridges, C.R., Mylonas, C.C., Caggiano, M., Deflorio, M., Santamaria, N., Zupa, R., Pousis, C., Vassallo-Agius, R., Gordin, H., Corriero, A., 2010. Spawning induction and large-scale collection of fertilized eggs in captive Atlantic bluefin tuna (*Thunnus thynnus* L.) and the first larval rearing efforts. J. Appl. Ichthyol. 26, 596–599.

Gareth, J., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning. Springer, Berlin, pp. 316–321.

Hanley, J.A., McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143, 29–36.

Hanley, J., McNeil, B., 1983. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 148, 839–843.

Hosmer, D.H. Jr., Lemeshow, S., Sturdivant, R.X., 2013. Applied Logistic Regression. John Wiley & Sons, Hoboken.

Landgrebe, T.C., Duin, R.P., 2007. Efficient multiclass roc approximation by decomposition via confusion matrix perturbation analysis. IEEE Trans. Pattern Anal. Mach. Intell. 30, 810–822.

Luque, P.L., Rodriguez-Marin, E., Landa, J., Ruiz, M., Quelle, P., Macias, D., Ortiz De Urbina, J.M., 2014. Direct ageing of *Thunnus thynnus* from the eastern Atlantic Ocean and western Mediterranean Sea using dorsal fin spines. J. Fish. Biol. 84, 1876–1903.

Mamitsuka, H., 2006. Selecting features in microarray classification using roc curves. Pattern Recogn. 39, 2393–2404.

Mann, H.B., Whitney, D.R., 1947. On a test of whether one of two random variables is stochastically larger than the other. Ann. Math. Stat. 18, 50–60.

Marrocco, C., Duin, R.P.W., Tortorella, F., 2008. Maximizing the area under the roc curve by pairwise feature combination. Pattern Recogn. 41, 1961–1974.

McLachlan, G.J., Do, K.A., Ambroise, C., 2004. Analyzing Microarray Gene Expression Data. John Wiley & Sons, Hoboken.

Megalofonou, P., De Metrio, G., 2000. Age estimation and annulus-formation in dorsal spines of juvenile bluefin tuna, *Thunnus thynnus*, from the Mediterranean Sea. J. Mar. Biol. Assoc. UK 80, 753–754.

Micera, E., Zupa, R., Zarrilli, A., Camarda, A., Moramarco, A.M., Acone, F., De Metrio, G., Corriero, A., 2010. A rapid latex agglutination test for gender identification in the Atlantic bluefin tuna, *Thunnus thynnus* (Linnaeus). Aquac. Res. 41, 1396–1401.

Miyake, M., Miyabe, N., Nakano, H., 2004. Historical trends of tuna catches in the world. FAO Fisheries Technical Paper No 467. FAO, Rome, pp. 74.

Mylonas, C.C., Bridges, C.R., Gordin, H., Belmonte Ríos, A., García, A., De la Gándara, F., Fauvel, C., Suquet, M., Medina, A., Papadaki, M., Heinisch, G., De Metrio, G., Corriero, A., Vassallo-Agius, R., Guzmán, J.M., Mañanos, E., Zohar, Y., 2007. Preparation and administration of gonadotropin-releasing hormone agonist (GnRHa) implants for the artificial control of reproductive maturation in captive-reared Atlantic bluefin tuna (*Thunnus thynnus thynnus*). Rev. Fish. Sci. 15, 183–210.

Mylonas, C.C., de la Gandara, F., Corriero, A., Belmonte Rios, A., 2010. Atlantic bluefin tuna (*Thunnus thynnus*) farming and fattening in the Mediterranean Sea. Rev. Fish. Sci. 18, 266–280.

Pousis, C., De Giorgi, C., Mylonas, C.C., Bridges, C.R., Zupa, R., Vassallo-Agius, R., de la Gandara, F., Dileo, C., De Metrio, G., Corriero, A., 2011. Comparative study of liver vitellogenin gene expression and oocyte yolk accumulation in wild and captive atlantic bluefin tuna (*Thunnus thynnus* L.). Anim. Reprod. Sci. 123, 98–105.

Pousis, C., Santamaria, N., Zupa, R., De Giorgi, C., Mylonas, C.C., Bridges, C.R., de la Gándara, F., Vassallo-Agius, R., Bello, G., Corriero, A., 2012. Expression of vitellogenin receptor gene in the ovary of wild and captive Atlantic bluefin tuna (*Thunnus thynnus* L.). Anim. Reprod. Sci. 132, 101–110.

Rosenfeld, H., Bridges, C.R., Heinisch, G., Corriero, A., Vassallo-Agius, R., Medina, A., Belmonte, A., Garcia, A., Fauvel, C., De Metrio, G., Meiri-Ashkenazi, I., Gordin, H., Zohar, Y., Mylonas, C.C., 2012. GnRHa-mediated stimulation of the reproductive endocrine axis in captive Atlantic bluefin tuna, *Thunnus thynnus*. Gen. Comp. Endocr. 175, 55–64.

Santamaria, N., Bello, G., Corriero, A., Deflorio, M., Vassallo-Agius, R., Bök, T., De Metrio, G., 2009. Age and growth of Atlantic bluefin tuna, *Thunnus thynnus* (Osteichthyes: Thunnidae), in the Mediterranean Sea. J. Appl. Ichthyol. 25, 35–45.

Santamaria, N., Bello, G., Pousis, C., Vassallo-Agius, R., de la Gándara, F., Corriero, A., 2015. Fin spine bone resorption in Atlantic bluefin tuna, *Thunnus thynnus*, and comparison between wild and captive-reared specimens. PLoS One 10 (3), e0121924.

Strobl, C., Boulesteix, A.L., Kneib, T., Augustin, T., Zeileis, A., 2008. Conditional variable importance for random forests. BMC Bioinform. 9, 307.

Ticina, V., Katavic, I., Grubisic, L., 2007. Growth indices of small northern bluefin tuna (*Thunnus thynnus:* L.) in growth-out rearing cages. Aquaculture 269, 538–543.

Tortonese, E., 1975. Fauna d'Italia. Osteichthyes (Pesci Ossei), vol. 2. Edizioni Calderini, Bologna, pp. 358–362.

Witten, I.H., Frank, E., Hall, M.A., 2013. Data Mining: Practical Machine Learning Tools and Techniques. Elsevier, Amsterdam.

Youden, W.J., 1950. Index for rating diagnostic tests. Cancer 3, 32–35.

Zupa, R., Fauvel, C., Mylonas, C.C., Santamaria, N., Valentini, L., Pousis, C., Papadaki, M., Suquet, M., de la Gándara, F., Bello, G., De Metrio, G., Corriero, A., 2013. Comparative analysis of male germ cell proliferation and apoptosis in wild and captive Atlantic bluefin tuna (*Thunnus thynnus* L.). J. Appl. Ichthyol. 29, 71–81.

Zupa, R., Santamaria, N., Mylonas, C.C., Deflorio, M., de la Gándara, F., Vassallo-Agius, R., Pousis, C., Passantino, L., Centoducati, G., Bello, G., Corriero, A., 2014. Male germ cell proliferation and apoptosis during the reproductive cycle of captive-reared Atlantic bluefin tuna *Thunnus thynnus* (Linnaeus). Aquac. Res. 45, 1733–1736.