Ecological Informatics 85 (2025) 102953

Contents lists available at ScienceDirect

Ecological Informatics

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

Improving machine learning predictions to estimate fishing effort using vessel's tracking data

J. Samarão [a,d], A. Moreno [a], M.B. Gaspar [a,c], M.M. Rufino [a,b,*]

[a] *Portuguese Institute for the Sea and the Atmosphere (IPMA), Av. Dr. Alfredo Magalhães Ramalho, 6, 1495-65 Lisboa, Portugal*
[b] *Centre of Statistics and its Applications (CEAUL), Faculty of Sciences, University of Lisbon, Portugal*
[c] *CCMAR*
[d] *Nova School of Science and Technology (FCT), Almada, Portugal*

ARTICLE INFO

Keywords:
Fishing effort
Machine leaning
Spatio-temporal high-resolution data
Small scale fisheries

ABSTRACT

Small-Scale Fisheries (SSF) comprise over 80 % of the global fleet and serve as the primary income source for numerous coastal communities. However, these critical fisheries face various threats. To effectively monitor SSF activities and their ecological impacts, it is required precise estimation of fishing effort using high-resolution spatio-temporal data. This information can identify areas with high fishing density, warranting protection of their main fishing grounds against other users (i.e. ocean grabbing), while also signalling potential stock depletion requiring management interventions and preserving the ecosystems from which these fisheries depend on.

In this study, we propose a series of steps to enhance the performance of Machine Learning algorithms in estimating fishing effort. We assessed seven supervised ML algorithms, including Logistic Regression, Ridge Classifier, Random Forest Classifier, K-Neighbours, Gradient Boosting Classifier, LinearSVC, Recurrent Neural Networks and XGBoost, using four case studies, from bivalve dredge and octopus pots and traps fisheries.

First, in a preliminary statistical analysis between common error measures derived from the confusion matrix was decided to use accuracy, precision, and sensitivity as evaluation criteria. We found that a simple moving average applied to speed, employed as a pre-processing technique using ten neighbouring points, showed up to 3 % improvement in results. Random Forest and XGBoost gave the best performances among the models compared (18 % change), using the variables Latitude, Longitude, Speed, Time, and Month (accuracies near 99 %)(61 % change). The proportion of the training/test dataset, showed a minimal impact on accuracy, with changes of less than 8 % when varying the training data percentage between 10 % and 90 %, making 60 % a suitable compromise. Considering the sampling unit to be (1) point-based (randomly selected pings) or (2) boat trip-based (randomly selected boat trips), leaded to changes in accuracy between 2.53 % and 3.99 %, depending on the model. Temporal resolution (ping rate) showed minimal effects on model performance, ranging from less than 2 % for intervals between 30 s (raw data with irregular time series) to 10 min (regular time series). As a post-processing step, it was concluded that replacing isolated data points with neighbouring values, significantly enhanced the detection of fishing events, with improvements ranging from 80 % to 250 %, depending on the model.

In conclusion, this study presents a straightforward procedure for selecting a machine learning method and enhancing its power of classification using simple procedures. These approaches should be applied in all works using machine learning to produce fishing effort maps.

1. Introduction

Global fish production is projected to reach 200 million tons by 2029, placing immense pressure on marine ecosystems and the sustainability of fish stocks. Robust, science-driven fisheries data are essential for developing effective management strategies and informing sound policy decisions. Achieving fisheries sustainability, enabling effective marine spatial planning, and assessing the ecological impacts

* Corresponding author.
E-mail address: marta.m.rufino@gmail.com (M.M. Rufino).

https://doi.org/10.1016/j.ecoinf.2024.102953
Received 3 December 2024; Received in revised form 9 December 2024; Accepted 9 December 2024
Available online 14 December 2024

of fishing activities require precise knowledge of when and where fishing vessels operate—i.e., the estimation of fishing effort (FE)(Coro et al., 2021). Fishing fleets can be categorized into two main segments based on vessel size: Large-Scale Fisheries (LSF), with vessels over 12 m in total length, and Small-Scale Fisheries (SSF), comprising vessels under 12 m. LSF accounts for approximately 60 % of global marine catches, whereas SSF represents 82 % of the 2.86 million motorized fishing vessels worldwide, and in Europe, 84 % *(*Bennett et al., 2015*)*. In Portugal, the pattern is similar: while LSF is responsible for 78 % of the seafood landings, SSF makes up 87 % of the fleet (https://rpubs.com/MRufino/SSF_EU_Map).

SSF fishers have minimal resources and often lack of political influence, thus making them more susceptible to most threats, like climate change or ocean grabbing, i.e. *dispossession or appropriation of use, control or access to ocean space or resources from prior resource users, rights holders, or inhabitants, which without appropriate governance processes, might comprise human security or livelihoods, or produce impacts that impair social–ecological well-being (*Bennett et al., 2015*)*. Monitoring and mapping the fishing effort of SSFs, has been done using remote sensing in very particular cases (Clarke et al., 2019) or fishers' questionaries, although generally it is done by tracking devices located on the boats. This provides critical information for policymakers, enabling sustainable ocean governance. Defining SSF essential operating zones and integrating them with Marine Protected Areas (MPAs) and other industries could ensure the long-term viability of SSFs. Such measures align conservation goals with socioeconomic development by fostering sustainable resource use.

However, estimating fishing effort in SSF requires high spatio-temporal resolution vessel tracking (i.e., minutes or seconds, ICES, 2023) and overall, there is an urge to increase fishing effort temporal resolution also in LSF (Rufino et al. in prep). Thus, it is essential to improve and standardise methodological approaches associated with fishing effort estimation, namely towards the required temporal resolution, and on the approaches to treat this kind of data, in face of recent amends to the Council Regulation (EC) No 1224/2009 Article 9 (Brussels, 30.5.2018, Regulation of the European Parliament and of the Council amending Council Regulation (EC) No 1224/2009) which states that all vessels including those below 12 m' length must have a tracking system implemented until 2030. This implies that a substantial amount of data will be available and requiring proper treatment on the following years. With this aim, recent efforts have been done towards developing a common framework to monitor fishing effort of SSF and static gears within the EU, namely through the Workshop on Geo-Spatial Data for Small-Scale Fisheries and within the ICES Working Group on Spatial Fisheries Data 1 and 2 (ICES, 2022; ICES, 2023).

Previous works have addressed the estimation of fishing effort in SSF, either by applying classical statistical methods typically used for VMS in LSF or, more recently, using machine learning methods (Behivoke et al., 2021; ICES, 2022; ICES, 2023; Souza et al., 2016; Syed and Weber, 2018; Torres-Irineo et al., 2021). Considering the large amount of data generated, several previous works explore the use of machine learning algorithms to estimate fishing effort, from tracking data, namely Data Mining (DM), and Multi-Layered Filtering Strategies compared with Hidden Markov Models (HMM) to classify trawlers and other fleets using AIS data (Souza et al., 2016), Conditional Random Fields were used to classify long-liners (Hu et al., 2016), Partition-wise Gated Recurrent Units (pGPRUs) and Recurrent Neural Network (RNN) to classify trawler fishing activities (Jiang et al., 2017), artificial Neural networks (multi-layer perceptron network)(Russo et al., 2011) to identify metiers on VMS data, Random Forest (Behivoke et al., 2021; Mendo et al., 2023; Torres-Irineo et al., 2021), Extra Trees, Random Forests, XGBoost, Bagging, Cat Boost, KNeighbors, LGBM, HistGBoost, Adaptive RF and MLP, to classify trawlers fishing activities in the Northern Adriatic Sea using AIS data (Brandoli et al., 2022). On the same area, ML methods were used to classify the gear being used, and the evolution of fishing effort through COVID pandemics, also based in AIS data (Coro et al., 2022). Rufino et al. (2023) develops a framework to treat this kind of data and compared methods to identify fisher's behaviour using statistical approaches and ML and evaluated pre- and post-processing approaches, whereas Mendo et al. (2023) proposes a protocol for the pre-processing steps.

Several isolated approaches have been used to improve ML classifications on previous works, but not in a systematic way, as in a framework. For example, Behivoke et al. (2021) and Rufino et al. (2023) advocated the application of a moving average to smooth speed tracking data, but did not statistically evaluated how many neighbouring values should be used to calculate the average (often called the order of the moving average). The most common aspect being compared are a plethora of ML algorithms (Rodriguez and Danhiez, ICES, 2022), but the proportion of train/test data has also been assessed in previous works (Brandoli et al., 2022; Desdhanty and Rustam, 2021). O'Farrell et al. (2017a) improved detection accuracy of VMS tracking data by sampling the data using a window-labelling method instead of point base, which avoids underestimation of 33 % of fishing effort.

Nevertheless, to the authors best knowledge, none of these studies assessed the effect of how small changes in the ML methodologies, pre and post processing of the data can improve the final predictions. Thus, the aim of the current work is to develop a framework of a series of steps that can be used to improve the predictions of supervised Machine Learning (ML) methods to estimate fishing effort, namely (graphical abstract in Fig. 1):

1) Error measures to be used to assess the methods;
2) Application of a moving average in speed and how many points should be used to estimate it (pre-processing);
3) The algorithm to be used: we tested 8 methodologies;
4) Assessment of the number of variables in the modelling procedure;
5) Data splitting. Percentage of train & test data. Often the models are developed and tested with a smaller percentage of training due to the reduced size of the validated dataset, but then when applied to the complete dataset (often very large), most data will be used for

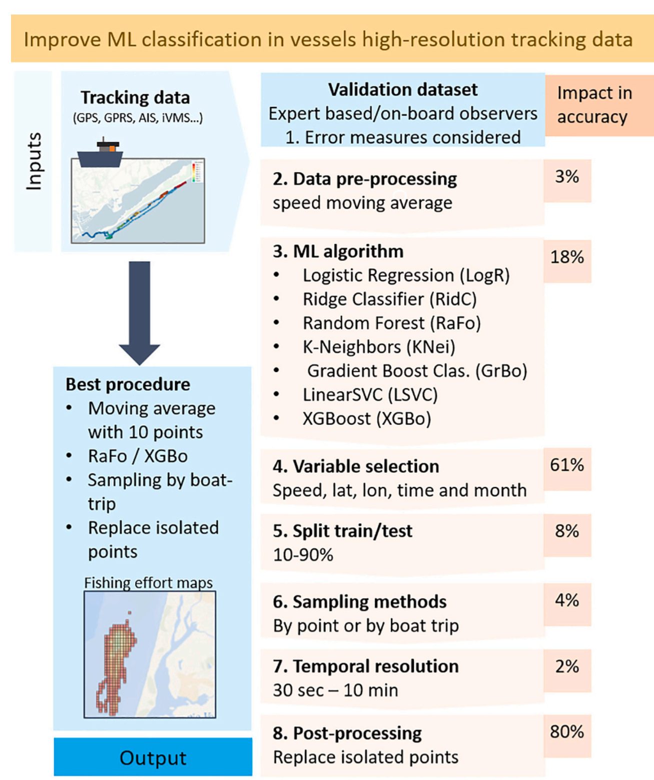

Fig. 1. Graphical Abstract.

training, thus the method selected should give similar results across different split percentages;
6) Sampling procedure, meaning whether the selection of train/test data is split by point or by boat trip;
7) Temporal resolution of the dataset, i.e. should we use the highest resolution possible or should we aggregate the data into larger intervals;
8) Effect of post-processing of replacing isolated points by neighbouring observations and fishing event detection;

2. Materials and methods

2.1. Case studies

Real time trackers (produced by the Portuguese company Robot) were installed in all fishing boats operating with bivalve dredges in the three main fishing grounds (Northwest, Southwest, and South) and in some SSF operating with octopus pots and traps (operating on the South) (project MONTEREAL, Mar2020). All Portuguese bivalve fleets are currently subject to mandatory tracking, since 2017, whereas only a few boats from the octopus' fisheries in the Algarve were equipped with GNSS (Global Navigation Satellite System, e.g. GPS) (although this number is increasing). From this dataset, 170 boat trips (from 48 boats) from 2017 were selected by a randomly stratified method (by month, gear, and zone) and validated by experts (Miguel Gaspar and Marta Rufino, IPMA), giving 272,054 locations to work on. Interactive plots of speed vs time and maps of the tracks were given to the independent experts, who registered the exact time when the haul started and finished in a fishing trip, as well as the gear used. Further details on the validation procedure can be found in Rufino et al. (2023).

The trackers deliver daily files with information of latitude and longitude, time, process (moving or stop), speed and bearing, every 30 s. The data received was very noisy and a preliminary data treatment was carried out to remove all points where the signal was not proper and points within the port area (defined as the points located within the port's areas/open bays/navigation paths polygons, manually defined). Table 1 shows the summary of the validated trips used in the current work. Further details of the dataset can be found in Rufino et al. (2023).

Additionally, the data is anonymised to prevent any ethical compliance, as the objective of the framework is to estimate the fishing effort and improve SSF management.

1. Error measures to be used to assess the methods

Several performance measures are commonly calculated to evaluate ML models, using the observed/predicted confusion matrix, namely accuracy, precision, sensitivity (recall), F1-score and AUC (see SUPPLEMENT 1 for further detail). These error measures were calculated for all models tested, for both sampling procedures, and compared following Rufino et al. (2023). Spearman index of correlation between the results given by each error measure was calculated and a correlation matrix plotted. Cluster analysis using correlation index was overlaid in the correlation matrix of the error measures to determine the main groups of highly collinear measures. Additionally, a PCA biplot was produced to visualize the patterns between the error measures relative to the ML methods and sampling procedures considered. All these analysis and plotting were done using R (r-project). The resulting plots can be found in Supplement 2.

Table 1
Summary of the validated boat trips.

Zone/ Fishery	Bivalve Dredges	Octopus Traps & Pots	Trammel Nets	Total
Northwest	46	–	–	46
Southwest	47	–	10	57
South	35	42	–	79
Total	128	42	10	170

2. Data pre-processing: speed moving average

The speeds estimated from positioning data tends to exhibit high irregularity. Additionally, certain fishing activities may share similar speeds with non-fishing activities, and vice versa. Consequently, an evaluation was conducted to assess the impact of applying a moving average (rolling mean) on the performance of the models. This process involved employing diverse rolling means, computed with varying numbers of neighbouring points for a given point. The calculation involves selecting n points around a point p, using both preceding and succeeding n points to estimate the mean speed (rolling mean) that becomes the new speed of point p. Additionally, along the current work, when n equals 0, it indicates the use of the standard speed directly obtained from GPS devices (raw data), whereas, for other values of n, the speed moving average was applied.

3. ML algorithms

Machine learning (ML) methods were used to classify each location given by the trackers into fishing/not-fishing. Seven supervised ML algorithms were evaluated: Logistic Regression (LogR), Random Forest Classifier (RaFo), K-Neighbor Classifier (KNei), Gradient Boosting (GrBo), LinearSVC (LSVC), Ridge Classifier (RidC) and XGBoost Classifier (XGBo). Additionally, Recurrent Neural Networks (RNN) were applied as a comparison with the ML algorithms. These models were selected as they have been reported to achieve high accuracies in previous works (> 85 %) (Behivoke et al., 2021; Brandoli et al., 2022; ICES, 2022; Jiang et al., 2017; Souza et al., 2016; Syed and Weber, 2018; Torres-Irineo et al., 2021). Furthermore, these methods have been applied in other fields of science, obtaining similar scores as the more recent methods, such as neural networks. An example could be Subasi and Erçelebi (2005) achieving a specificity of 90.1 % with LogR/RidC to classify if a patient is epileptic or "normal" using electroencephalograph signals. Likewise, RaFo, GrBo, and XGBo have been applied in the fields of medicine such as liver cancer and diabetes detection, with promising results (>80 % accuracies) (Desdhanty and Rustam, 2021; Xu and Wang, 2019). Nonetheless, RNN have been applied to predict phylogenetically distinct protein families (Liu, 2017), speech recognition (Hori et al., 2018), and even for learning spatio-temporal sequences (Wang et al., 2023). Additionally, LSVC and KNei have been successfully applied in classification for object-based image analysis (Tzotsos and Argialas, 2008). Moreover, the selection of algorithms such as LogR, KNei, LSVC, and RidC provide a more consistent interpretation of the results contrary to RF, GrBo, XGBo, and RNN which are known as "black-box" models, being challenging to understand how the predictions were made.

Nonetheless, a brief description of the methods used is given below, whereas the mathematical details can be found elsewhere (Farnham et al., 2021; Pedregosa et al., 2012; Zhang et al., 2023). Note that for this description, the term 'label' was used for points in the training data set, whereas the term 'classification' was used for the predicted classification of the new observations, i.e., test data set.

Logistic Regression (LogR): is a statistical technique used to determine the likelihood of an event happening, based on the given input data. Let's consider a scenario where we want to determine the probability of a boat being classified as 'fishing' in each coordinate. In LogR, we utilize a special mathematical function called the sigmoid function to carry out this calculation. The sigmoid function takes the input data and transforms it into values ranging from 0 to 1. These transformed values represent the probability of the event occurring. For instance, if the sigmoid function gives a value of 0.8 for a particular point, it means there is an 80 % chance that the boat is fishing in that point.

Random Forest Classifier (RaFo): In a Random Forest, the algorithm creates a "forest" of many decision trees. Each decision tree is like a flowchart that asks a series of questions about the features of the trip and eventually assigns it to one of the categories (fishing or not fishing). The "random" in Random Forest comes from the fact that each decision tree is slightly different. The algorithm randomly selects a subset of the input data and a subset of the features for each tree. The final model results of an assemble of all trees produced. Thus, it creates a diverse set of decision trees that can capture different patterns and reduce overfitting (making the model too specialized for the training data).

K-Neighbor Classifier (KNei): This algorithm stores all the information given in the training data set (labels) and makes the predictions using only the stored data. This is why it is so-called lazy learning, because it does not actually 'learn' from the data set, and only compares the new observations with the stored ones. Thus, when a new point from a trip needs to be classified, the algorithm looks at its neighbouring points (K nearest neighbours) in the feature space (stored data). Moreover, it calculates the similarity or distance between the new point and its K nearest neighbours using a distance metric (e.g., Euclidean distance). The distance metric measures how far apart two points are in the feature space. The KNei considers the labels of the K's nearest neighbours and assigns the most frequent label among these to the new point.

Gradient Boosting (GrBo): This method starts with an initial weak model, often a simple decision tree, and uses this model to make predictions, which are not very accurate initially. The algorithm then focuses on the examples where the model performs poorly and tries to improve on those. It does this by training a new weak model that specifically targets the mistakes or errors made by the previous model. The new weak model is trained to predict the difference between the actual target value and the predictions made by the previous model. This difference is called the residual error. The algorithm, then adjusts the weights or parameters of the new model to minimize this residual error.

LinearSVC (LSVC): The algorithm tries to find a straight line (in the case of 2D) or a hyperplane (in higher dimensions) that best separates the fishing and non-fishing events (similar to a linear discriminant analysis). The goal is to find the largest distance from this discriminating line to the observations of each class, i.e., the margin. The optimal hyperplane is found by creating a linear decision boundary that assigns each point to one of the two classes. It does this by calculating the weights or coefficients for each feature (x and y in our example) to determine the position and orientation of the decision boundary. The algorithm aims to minimize classification errors while maximizing the margin.

Ridge Classifier (RidC): The algorithm aims to find a linear decision boundary that best separates the data points into their respective categories (similar to a logistic regression). The decision boundary is a hyperplane that splits the feature space into two regions, one for each category. The Ridge Classifier uses Ridge regularization, also known as L2 regularization, during the training process. This regularization technique adds a penalty term to the model's objective function, which encourages smaller and more balanced weights for the features. By doing so, Ridge regularization helps prevent overfitting and reduces the sensitivity of the model to noisy or irrelevant features.

XGBoost Classifier (XGBo): or Extreme Gradient Boosting is a powerful machine learning algorithm that combines multiple weak predictive models, usually in the form of decision trees. It follows an additive training process, where each subsequent model focuses on reducing the errors made by the previous models, similar to Gradient Boosting.

The key idea behind XGBoost is to gradually enhance the overall prediction accuracy by training weak models sequentially. During this training process, it assigns weights or scores to each example based on their difficulty of prediction. Initially, all examples are given equal weights. However, as the models are trained, these weights are adjusted to emphasize examples that are more challenging to classify correctly. By doing so, XGBoost can effectively learn from the mistakes made in previous models and improve its predictions. To prevent overfitting, the model constructs decision trees that are shallow, meaning they have only a few levels. Shallow trees are less prone to memorizing noise in the data and are more generalizable. Additionally, it incorporates various techniques to regularize the models and prevent them from becoming overly complex. These techniques include applying penalties for model complexity and utilizing subsampling of the data, where only a subset of the examples is used during each iteration.

Recurrent Neural Networks (RNNs): These are a type of artificial neural network designed to handle sequential data by maintaining memory of past inputs. RNNs utilize recurrent connections to feed the output of a neuron back to itself or to other neurons in the network, allowing it to capture temporal dependencies in the data. Taking into consideration the current dataset, we have a series of data points with information about location, speed, and other pertinent variables. Every sequence denotes a portion of the trip, with every data point signifying a certain time interval, which depending on whether the boat was seen fishing during that segment, we assign a label to each sequence. The RNN's purpose is to process these data point sequences and identify the temporal relationships among them. The RNN learns to identify patterns in the data that point to fishing activity by processing each data point in the sequence while preserving its internal state.

2.2.

Model optimization was performed in two steps. The first step is the Random Grid Search whereas an initial grid of the hyperparameters that would affect the model performance is produced and over a total of five iterations, the hyperparameters are randomly selected and tested (Pedregosa, F. et al. 2011). Furthermore, in the second step which is implemented within the first one, a Cross-Validation procedure will control overfitting (which concept can be read in Supplement 3). In this procedure, the data was divided into K-Folds (K = 10), and for a total of K iterations, K-1 folds are used to train the model and one to validate it, whereas in each iteration, the folds used to train and validate the model are changed. As all observations were used in the train set, as in the validation set, it will assure that the model has most patterns in the data, and it is not picking up too much of the noise, or in other words, it is low on bias and variance. The cross-validation was stratified along the process, which means that all labels were equally characterized and distributed through the K sets.

4. Variable selection

One of the most important steps in these analyses, is the variable's selection, as it is the correlation between the variables that will explain and distinguish the different events in the fishing trips. Thus, the first step was to select which variables should be included in the model and how many. Time was considered as numeric (i.e., seconds) and month was extracted to incorporate seasonal variability in the analysis. The speed variable given by the GPS showed a high correspondence to the one calculated manually, i.e., using the coordinates and the distance between successive points. Therefore, having in mind, the computer burden, it was decided to use speed variable given by the GPS directly and not calculate a new one. Several other variables were tested but not used in the final models (e.g., depth (GEBCO), distance to the coast, habitat/sediment type and boat features like engine power) (not included for brevity), in face of the good results obtained with the dataset using only the variables given by the GPS, as this would require a minimum posterior processing. Thus, combinations of two to five variables were used, including, Latitude, Longitude, Speed, Time, and Month, giving 26 possible combinations. These were evaluated for all methods considered expect for the Recurrent Neural Networks.

5. Split train/test and sampling method

For model evaluation, different percentages of the train/test data sets have been used. As data availability for Small-Scale Fisheries is scarcer due to certain regulations not being applied to these fisheries, it is crucial to observe what is the impact of training models with fewer observations. Furthermore, the validation data set is often reduced, as the number of verified boat trips in these types of data sets is typically low as it relies on on-board observers, which are in short supply. Therefore, to access how the number of trips influence the models' performance a range of percentages was considered to train the models, with the lowest being 10 % (17 trips) and incrementally increasing by 5 % the number of trips until reaching 90 % (153 trips). The process of training the models was performed five times to obtain a measure of the associated variability, avoiding any misleading conclusion.

6. Sampling methods

When applying ML methods, the data should be split into a training and test data sets, by randomly sampling the initial data. For this procedure, two steps were evaluated, i.e. (i) the way the data is sampled and (ii) the percentage of training/test on the model's performance.

Sampling can be done by (1) point, that is randomly selecting a certain number of points (pings) on all data sets or by (2) boat trip, where the selection of the training data is done by complete boat trips. It is expected that if we train the model using random points from various boat trips, it can be overly optimistic, but this effect has not been measured yet. All the analysis was done for both situations, i.e., sampling by point or by boat trip, to evaluate its effect on the results.

7. Temporal resolution

There are several devices to monitor the boats (AIS, VMS, GPS) which dependently of the model and data acquired, provide different temporal resolutions (10 s, 30 s, 5 min, so on). Thus, it is desirable that an algorithm keeps its performance despite the interval between pings. The model performance was evaluated across eight-time intervals, by resampling the data over regular intervals of 30 s, 1, 2, 4, 5, 6, 8, and 10 min. Raw data was mostly every 30 s, but it was not regular, and therefor was labelled as zero (0) in the figures.

8. Post-processing: number of isolated points replaced and fishing events detection

Typically, ML methods give scattered results. However, the fisher's behaviour is continuous in time during the fishing operation. Throughout the project, after classifying several boat trips it was encountered sequences of points that had some misclassified points. Therefore, a methodology was used to do final refinements consisting of analysing the predicted values from the models' and selecting n values to verify before and after a certain point.

For example, the model predicted a vector $\mathbf{x}$, with k values containing 0's and 1's, where 0 means travel activities and 1 fishing activity, the algorithm will switch between events only

$$if\, x_{i-n,\dots,i-1}! = x_i\, and\, x_{i+1,\dots,i+n}! = x_i$$

(if all points behind and after are different from the current one –Fig. 2: 1st case)

Or

$$if\, x_{i-n,\dots,i-1}! = x_i\, and\, x_{i+n}! = x_i$$

(if all points behind are different and the last point is different –Fig. 2: 2nd case)

Or

$$if\, x_{i+1,\dots,i+n}! = x_i\, and\, x_{i-n}! = x_i$$

(if all points after are different and the first point behind is different –Fig. 2: 3rd case).

where n is the number of points to verify.

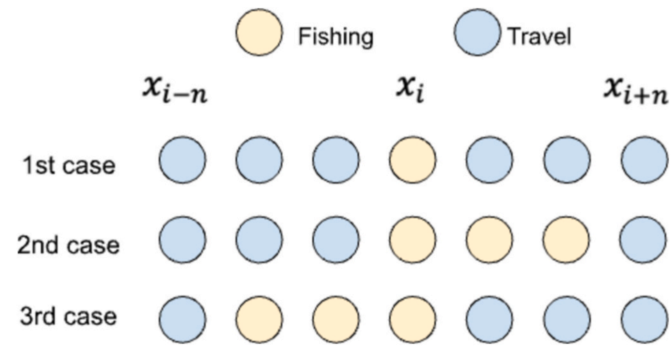

Fig. 2. Schematic representation of the post-processing algorithm used ($n = 3$). The first row represents the case of one isolated point (which would then be re-coded as 'travel'). The second and third cases detect if after or behind a sequence of points there is a probability of having a sequence of the opposite class (if so then x_i would then be re-coded as 'travel').

If one of the conditions above are matched, then the event is switched.

After using all the steps presented before, a last assessment was done based on the number of fishing events retrieved. As the vessels carry "grey boxes" containing fixed GPSs devices, which are continuously recording, implying that the transmitter is never off, it was defined that a fishing trip starts when the boat leaves the harbour and finishes when it returns, and during this period there might be more than one fishing event (e.g., more than one hauling/setting events). The number of fishing events retrieved is therefore another important parameter to be optimized during the procedure.

To find the number of events during a trip (to apply in not validated datasets), we applied a time-based filter to the total duration of fishing events for each fishing method (métier). Essentially, we examined the time gap between the end of one fishing activity and the start of the next.

To find the number of fishing events in other trips, we developed an algorithm to detect all fishing sequences. It calculates the time difference between the last point in one sequence and the first point in the next, considering a fixed 30-s interval between data points (pings). If this time difference exceeds a certain threshold, we consider it a separate fishing event; otherwise, it's part of the same event. To fine-tune the value of this threshold, we utilized Python's SciPy Minimize function (Virtanen et al., 2020). The objective was to find the optimal threshold that minimizes the error between the actual number of events and the number defined by our algorithm. We evaluated thresholds ranging from 1 to 20 min. The number of fishing events detected by each algorithm was therefore considered as an additional evaluation criterion for the different machine learning algorithms considered.

3. Results

The results of the six error measures calculated for all models and both sampling strategies (by point and by boat trip), were highly correlated between each other (Spearman correlation coefficient > 56 % for all pairs), as it can be expected by the equations used to calculate each of them (SUPPLEMENT 1). Nevertheless, cluster analysis identified three main groups of measures. The first cluster group was composed of precision (which would represent the percentage of times that fishing points were assigned as fishing) and specificity. From these two, precision was selected as it was considered that the correct assignment of the percentage of positive cases (i.e., when the boat was fishing) was more important than the negative cases (i.e., when the boat was not fishing). The second group showed only sensitivity/recall, which represents the percentage of points of all the true fishing activities that were right classified, which is a very important measure, and therefore used in further analysis. The third group included accuracy, AUC, and F1 score. From these three measures, accuracy was used as it is less correlated with the remaining measures. This measure represented the percentage of correct classification, and it was more straightforward to understand and most commonly used in previous works. The PCA biplot showed that the first axes separated the models showing better performances overall from the remaining ones, i.e., KNei and LSVC method, for both sampling strategies (Supplement 2). The second axis discriminated sampling strategy mostly, with sampling by boat trip being overall more related to higher Precision/Specificity values and sampling by point being more related to higher Sensitivity/recall and Accuracy/F1-score values. This indicated that the method used produces higher differences in the outcomes than the sampling procedure.

All ML algorithms showed excellent performances (>81 % accuracy using 5 variables) (Fig. 3). Moreover, Recurrent Neural Networks (RNNs) were tested with the variables Latitude, Longitude, Speed, and Time only. As these take a long time to optimise to achieve similar results, we did not test all combination of features. Nonetheless, these models achieved an impressive accuracy of 96 %, obtaining similar results to the tree models (Supplement). However, when estimating its precision only 60 % was obtained, illustrating an overestimation of fishing activities. Furthermore, sample by point was generally better then sampling by boat trip, for all methods considered. The methods showing better performances were Random Forest classifier and XGBo for sampling by point (accuracies approximately of 99 %), and GrBo for sampling by boat trip (93 % accuracy), followed also by RaFo (92 %).

Sampling by boat trip had a difference in accuracy of 7.99 % less for RaFo, 7.71 % less for XGBo, but 2.9 % more for GrBo than sampling by point (Fig. 3). Therefore, the remaining analysis were done using these three methods only, RaFo, XGBo and GrBo.

The model that produced worst results was LinearSVC (LSVC), it was the only model that had performances below 70 %, for both sampling approaches.

Accuracy increased with the number of variables used in the modelling procedure, in all cases except KNei, thus, the remaining models were always produced using the 5 variables (latitude, longitude, speed, time and month). The improvement due to the inclusion of the variables varied between 82.65 % and 94.82 % for sampling by point, and 77.90 % and 91.81 % for sampling by boat, according to the methods used (excluding KNei).

The method showing the most similar performance across the different percentages used to train/test data was RaFo, for sampling by point (varying between 96.69 % to 99.58 %), whereas for sampling by boat trip, it was GrBo (varying between 53.69 % to 94.29 %)(Fig. 4). The change in accuracy of the results from using 10 % of the data for training the algorithm up to 90 % of the data was below 0.7 % for sampling by point and 7.9 % for sampling by boat trip. In this last case, sampling by boat, produced worse results when less than 30 % of the data in the training set was used. Thus, in the remaining work, all models were trained using 60 % of the data, for conservative purposes.

Random forest (RaFo) was the method showing higher computation time, i.e., up to 30 s when 90 % of the data was used for training, followed by GrBo and XGBo. Computation time was similar in both sampling strategies (Supplement 4). Comparing the performance score with the computation time, XGBo had a favourable trade-off as it took a maximum of approximately six seconds to be trained and reached performances close to 99 % for sample unit by point, and above 90 % for sample unit by boat. Regarding the optimal model in each approach, RaFo, despite achieving higher performances, required more computation time for both training and prediction. GrBo took nearly the same amount of time as the Random Forest per sample unit by boat, with the difference from RaFo not being significantly substantial. However, when considering this time in the context of the total number of observations (272054), thirty seconds represents a considerable duration.

For the three best models (RaFo, XGBo, and GrBo) the different temporal resolutions resulted in a change of −1.9 % in accuracy for sample by point and 1.1 % in sampling by boat trip for ping intervals between 30 s (raw data, irregular time-series) up to 10 min (regular time series)(Fig. 5). Fig. 5 illustrates that for sampling by point, the increase in temporal resolution corresponded to a decreasing tendency in performances, but the opposite was observed for sampling by boat. *Re*-sampling over different time resolutions, as it was done in the current work, implies a decrease in the data variability, which may in part have caused these observed fluctuations from raw data of 30 s to 10 min regular data. Because the priority would be to avoid extra data processing, the remaining models were applied using the raw data only, as

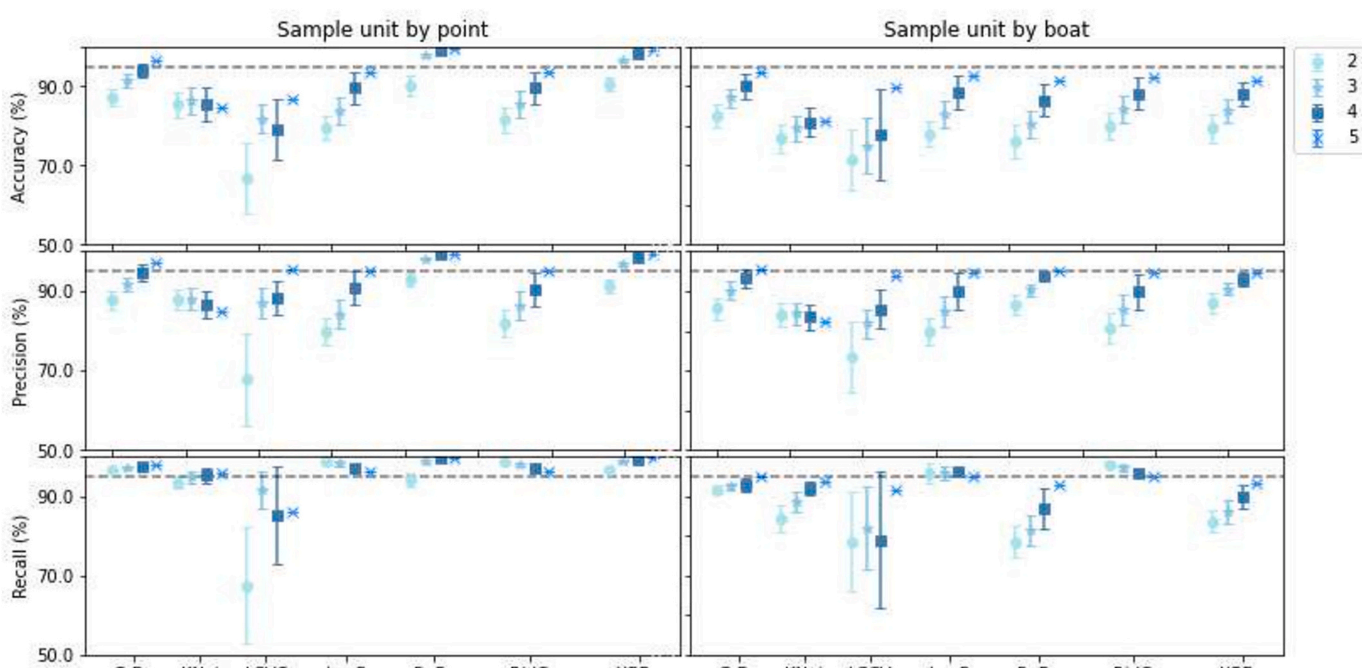

Fig. 3. Average performance of each method (and respective standard deviation error), by the number of variables included where the colors illustrate how many variables were used (all possible combinations of latitude, longitude, speed, time, and month). The dashed grey line illustrates the 95 % level.

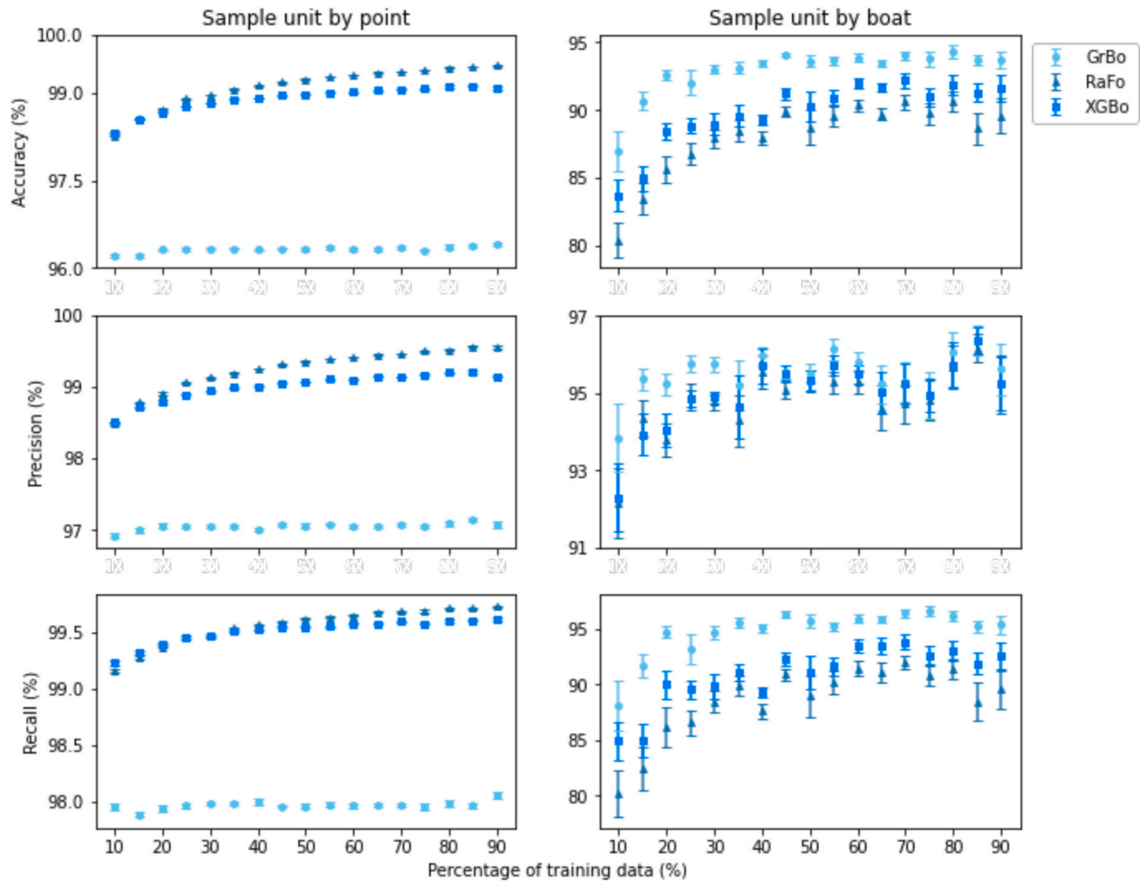

Fig. 4. Behaviour of the ML models in face of the percentage of training/testing data split (10–90 %), both for sampling by point and by boat trip;

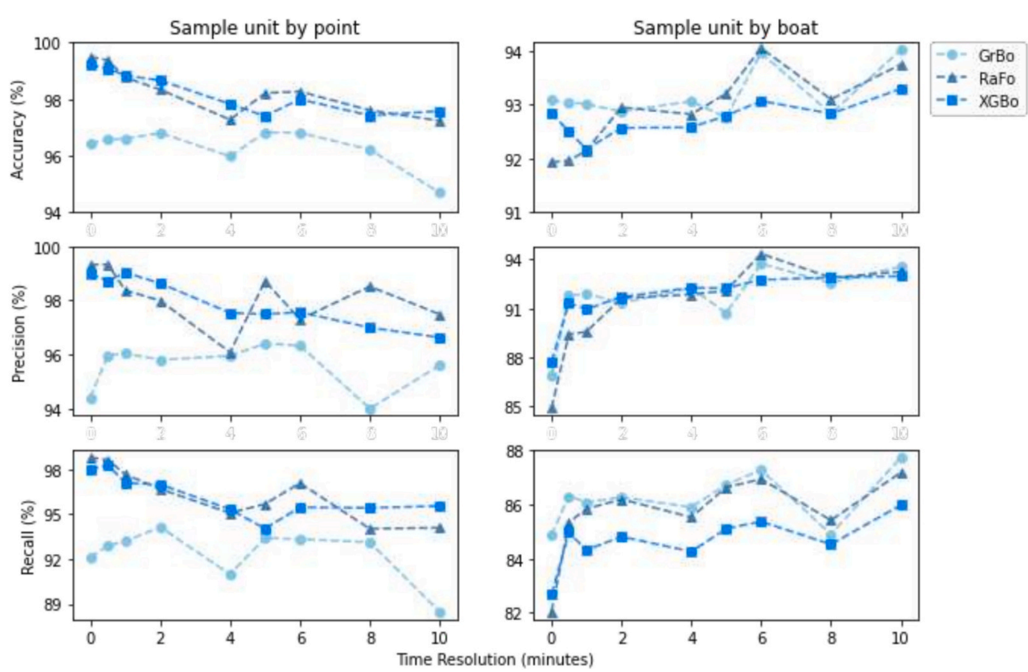

Fig. 5. Performance of the ML algorithms by changing the temporal resolution.

this work is focused on improving methods' performance, and estimate the variability associated with each step considered.

Applying a moving average (MA) to 'speed' prior to the analysis, improved the results given by the three best models' performance for both sampling procedures, showing a change of accuracy of 0.6–3.39 %, from using the raw data (zero points used in the MA) up to calculating a moving average of speed using the 20 neighbouring points (10 min) (Fig. 6). The patterns observed in the change of performance were

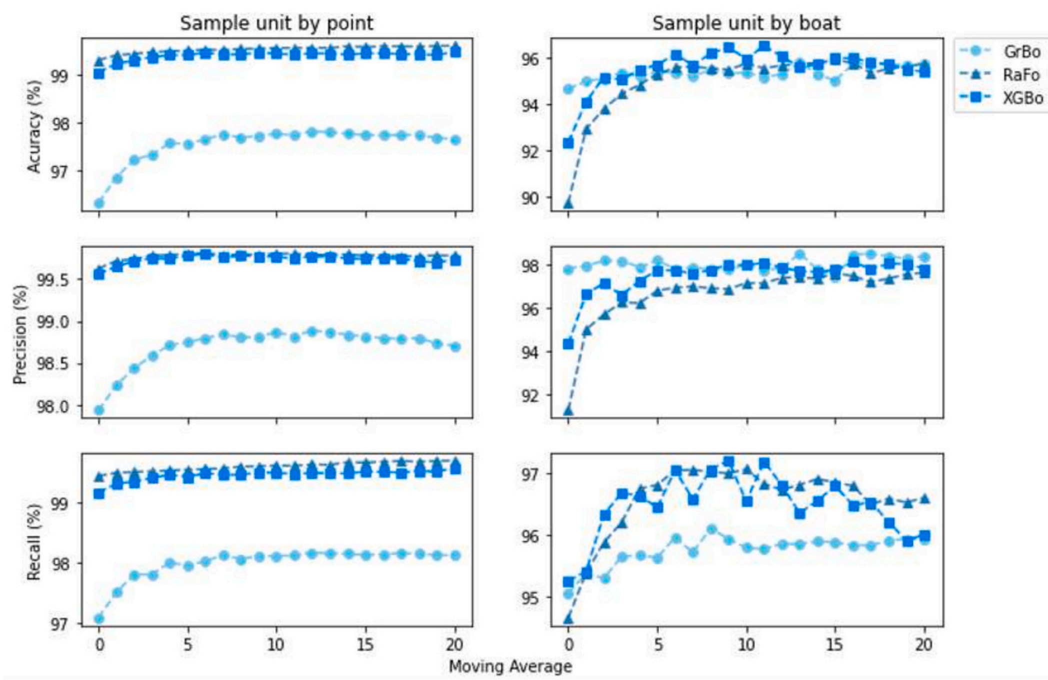

Fig. 6. Performance of the ML algorithms by applying different moving averages.

similar for both sampling by point and by boat trip. In all cases, the curves were stable at about 10 points (5 min) used in the speed moving average, thus, this value was selected for the final model.

After taking into consideration the pre-processing steps that would produce the best results according to what was previously analysed, the following figures (Fig. 7 and Fig. 8) illustrate the variance explained by the predictive models (VEvc) between the real number of events and the predicted ones, when applying the post-processing algorithm to rectify isolated points. Also, the figures on the right side of the VEvc show how many events were overestimated or underestimated by the algorithm.

To determine distinct fishing events, we calculated the minimum time interval required per métier and trip between two fishing events and estimated the total number of events after applying the post-processing algorithm.

These results suggest that RaFo and XGBo for both sample approaches were the models that produced a better detection of fishing events within each boat trip. Random Forest had the highest VEvc (0.95) when replacing isolated points using the two neighbours and XGBo when using ten (0.78), using sample by point. For sample by boat, RaFo had presented a higher VEvc by analysing ten neighbours (0.60) and

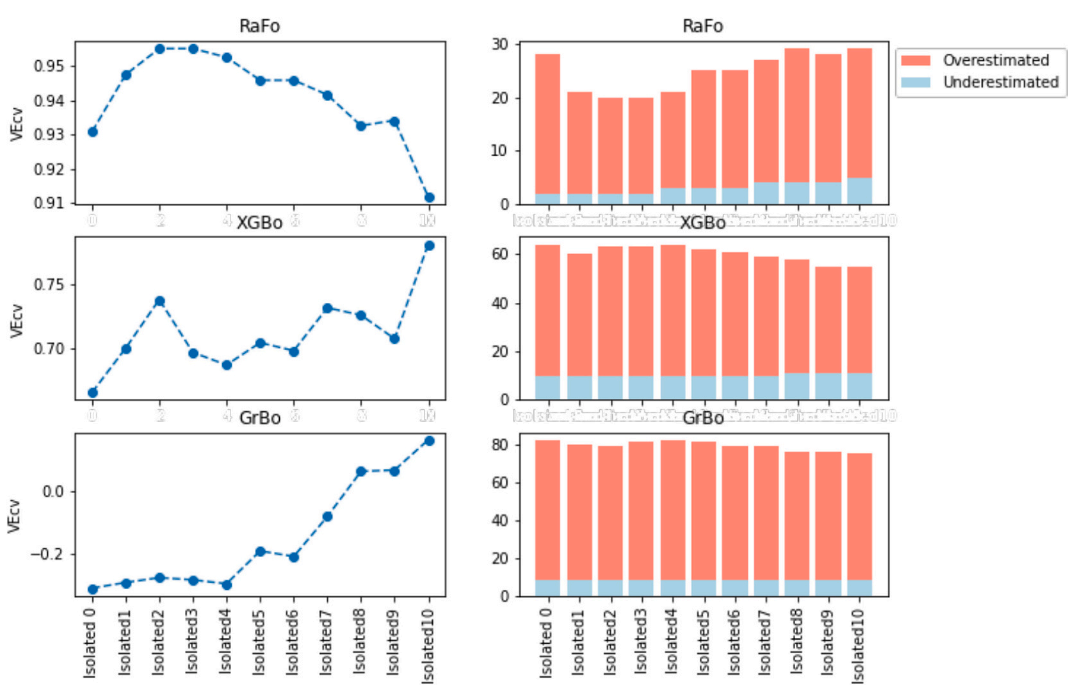

Fig. 7. Performance of the application of the post-processing algorithm and representation of the number of events that were overestimated and underestimated with and without it (sampling unit by point).

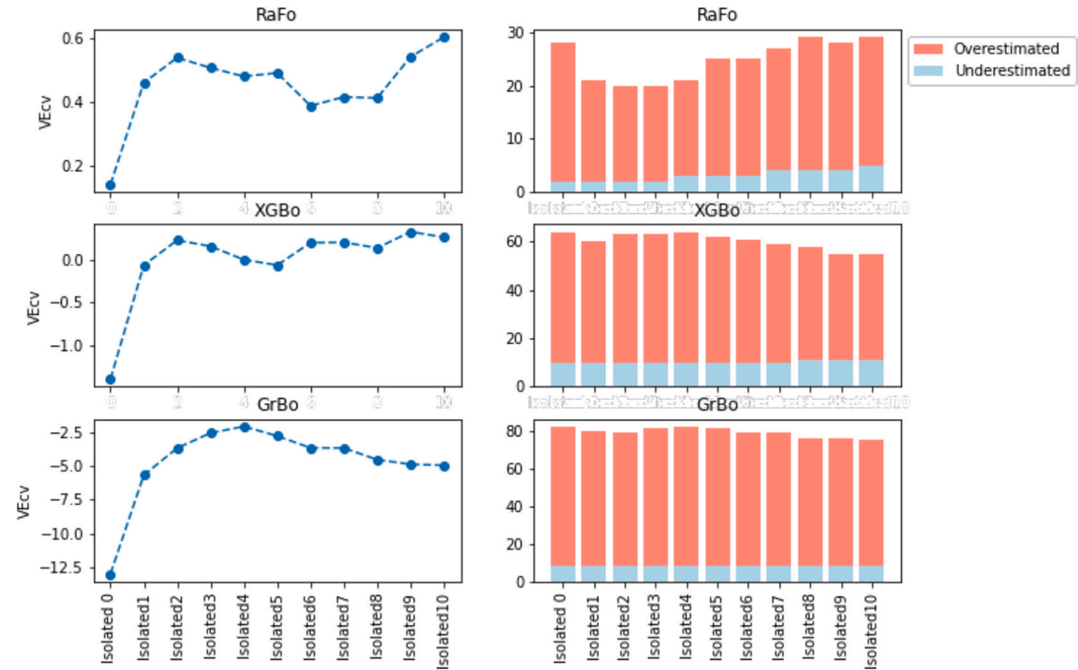

Fig. 8. Performance of the application of the post-processing algorithm and representation of the number of events that were overestimated and underestimated with and without it (sampling unit by boat).

XGBo nine (0.32).

Further, despite most of the methods' results producing a higher number of trips that either had overestimated or underestimated events a higher VEvc suggests that these estimations are mostly wrong by one or two events.

Moreover, the decisions considered in these series of steps were compared with the worst-case scenario that could be chosen, in order to quantify the potential effect of each step (Fig. 9). The three models illustrated to be affected mostly by the amount of data used to train, by the variables, and by the post-processing algorithm. The least impactful steps were the time resolution and the speed moving average. However, depending on the model these steps could improve the performance by 2 % to 5 %.

The best model in the current case-study was Random Forest Classifier with sampling by point, using the five variables Latitude, Longitude, MovingAverage10, Time and Month, with no application of the time resolution and with a post-processing of 2 isolated points with an accuracy of 99.07 %.

RaFo emerged as the model with superior performance. Subsequently, a final test was conducted to assess the model's efficacy on trammel net fishing trips data that the model had never encountered and a gear that was not part of the training set. The results proved remarkably favourable. When using the sampling unit by point, the model achieved an accuracy of 92.15 %, precision of 86.27 %, and recall of 99.98 %. Meanwhile, employing the sampling unit by boat yielded corresponding metrics of 92.48 % accuracy, 86.77 % precision, and 98.98 % recall.

4. Discussion

In the current work, we propose a framework of a series of simple pre- and post-processing procedures to improve the predictions of

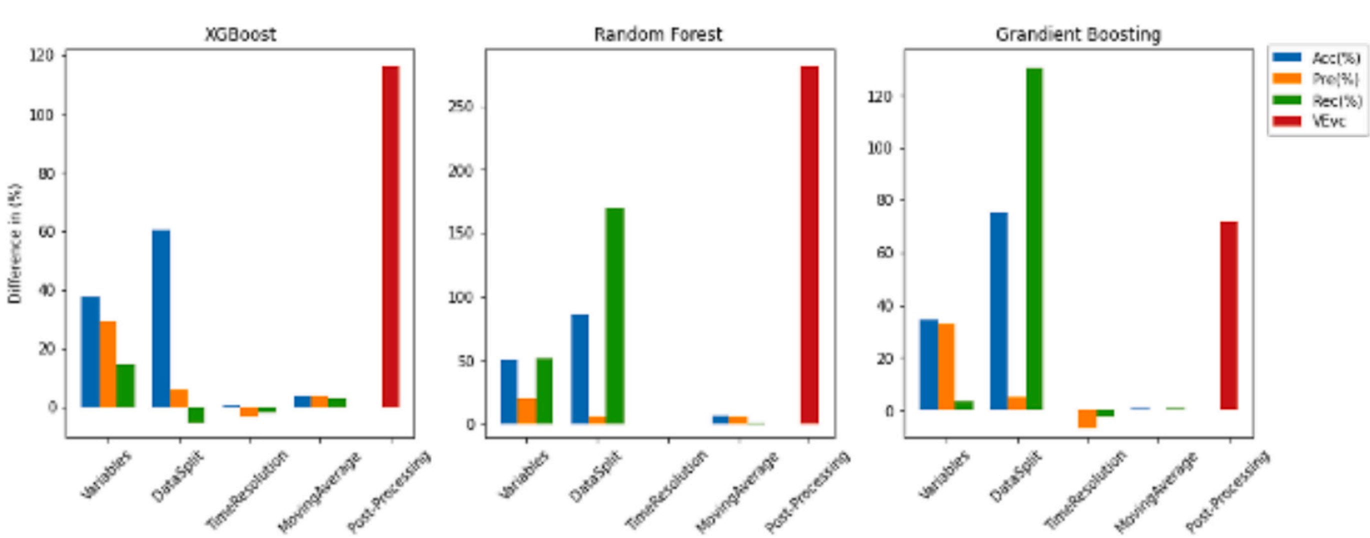

Fig. 9. Evaluation of how impactful each selected step was compared with the worst choice that it could have been made.

machine learning algorithms (ML) to classify fishing boat behaviour, from tracking data. The framework was illustrated using four case studies of small-scale fisheries (SSF) operating on the Portuguese coast: bivalve dredges from three different fishing grounds and octopus pots & traps. Eight main steps were considered, namely: (1) the error measures to evaluate the algorithm (based on the confusion matrix), (2) number of points to be used in the speed moving average (pre-processing), (3) the algorithm to be used, (4) the number of variables to include in the model; (5) the percentage used in train/test data, (6) the data sampling procedure (for test/train split, by point or by boat trip), (7) the temporal resolution (intervals between consecutive points), (8) number of isolated points to be reclassified (post-processing) and number of fishing events identified. Using this procedure, the best model for the case studies was successfully applied to a new different metier (nets), not used in the algorithm training, to evaluate its performance on new datasets. It was recommended that for each new metier, these steps are carried out as a preliminary analysis, before implementing an ML procedure.

(1) error measures to evaluate the algorithm (based on the confusion matrix)

To evaluate the models, we statistically selected the metrics accuracy, precision, and recall, which are also the most used when the task is classification (Rodriguez's work in ICES, 2023; Henriques et al., 2023; Behivoke et al., 2021; Souza et al., 2016). Accuracy is known to be biased for unbalanced data, as it is very easy for an algorithm to define that all the observations are just from one of the classes, thus achieving values above 80 %, for example. Each metric has its issues, but also reflect different aspects where the models might be failing. Therefore, it is recommended to have more than one metric to evaluate the models. In other works, the same metrics were applied but for the negative class, i.e. specificity and precision for the negative class (Brandoli et al., 2022; Souza et al., 2016), however, here we give more emphasis to accurately estimate the positive class, i.e. when the boat is fishing. To the authors best knowledge, no previous works analysed which metrics should be used for model's comparison. However, we have shown that the best method changes according to the classification metric used, as different aspects of the data are being accounted. It is thus crucial to use a good combination of metrics properly justified. Furthermore, for future comparation among works, standards on what metrics to use, should be defined, alternatively.

(2) number of points to be used in speed moving average (pre-processing)

The fact that speed raw data (either obtained from GPS or AIS) is very irregular (scattered/conspicuous), showing high variability with numerous sudden ups and downs, and outlying boats behaviours, evidenced the importance of testing the effect of applying a moving average (rolling mean) to speed prior to the algorithm. The question than arising is, how many points should be used to calculate this moving average? We proposed an approach to test the effect of the number of points in model performance. In our case study, we concluded that from applying a filter to speed by calculating a moving average using 10 points, would provide a good compromise between resolution and performance. A similar approach was used by Behivoke et al. (2021) who advocated the application of a moving average (MA) implementation using 5–10 points, before applying Random Forest. Those were all included in the model that had a total of 24 features, reaching accuracies scores of 74–89 %. We did not use more than one moving average to train the models as the performances were already close to 99 % of accuracy. Therefore, there was no need to add more complexity to the model. Future work can include different ways to process the speed variable, as temporal resolutions differ among works. Taking the previous example, if the shortest trips last 60 min, and temporal resolution is 10 min, the moving average would be applied only to 6 points, which may not make any sense.

(3) the algorithms to be used

The results of the machine learning algorithms obtained in the current work, slightly outperformed the ones found in most previous works, to the authors best knowledge (e.g., current work 99.07 % accuracy; 94.3 % Julien Rodriguez ICES, 2023; 91 % Mendo et al., 2023; 89 % Behivoke et al., 2021, 95 % Torres-Irineo et al., 2021; 89 % Jiang, X., et al. 2017; 83 % O'Farrell et al., 2017a; 89 % Hu, B. et al. 2016). Nevertheless, the accuracies reported were all above 80 %, making ML approaches among the top performing methods, for this type of data.

Within these works, Random Forest was one of the methods showing higher performances (Mendo et al., 2023, Rufino et al., 2023, Julien Rodriguez ICES, 2023, Behivoke et al., 2021, O'Farrell et al., 2017a) and from the statistical methods used the commonest are probably the Hidden Markov Models (HMM) (Henriques et al., 2023; Marine Institute, Ireland ICES, 2022; Mendo, 2019; Souza et al., 2016). Moreover, Rufino et al. (2023), compared the results given by statistical methods and machine learning and found similar performances between the best statistical approaches and Random Forest, showing that both alternatives should be considered.

Rodriguez and Danhiez (ICES, 2022) presented the efficacy of five machine learning algorithms for distinguishing more than 10 fishing gears and similarly to the current work, XGBoost demonstrated an impressive accuracy of 94.89 %, positioning it as a robust methodology for identifying fishing activity, regardless of the gear employed. In parallel, Rodriguez (ICES, 2023), modelled an algorithm to identify hauling, setting, and non-fishing activity in vessels operating nets, and found that Random Forest was the best method with an accuracy of 91.5 % and when integrated with Geo-computation further enhanced Random Forest's accuracy to 94.3 %. This aligns with the current work's findings, reinforcing that Random Forest remains among the best methodologies for accurately identifying fishing activities. Furthermore, this coincides with Farnham et al., 2021, acknowledging that while Random Forest stands as a powerful and simplistic method, the increasing popularity and award-winning success of XGBoost in Kaggle competitions (a platform where machine learning practitioners and data scientists participate to solve real-world problems by developing and refining predictive models using provided datasets) highlight its growing influence. Despite being a slightly less-performing alternative than Random Forest, XGBoost's robust performance underscores its credibility as a viable choice in the classification of fishing behaviour.

Other alternative methodologies have been applied in previous works to identify fishing activity. Brandoli et al. (2022) explored Conditional Random Fields (CRFs) to discern fishing and non-fishing activities within Longines fisheries. By modelling the conditional probability distribution to fishing activity detection, the authors achieved 89.2 % accuracy. Despite the novelty of this approach and its limited availability, the provided GitHub repository with code makes it an intriguing avenue for future exploration. Data mining approaches have also been used, namely, Lavielle's algorithm, achieving accuracies of 83 %, which fell slightly behind the current work's performance (Souza et al., 2016). Notably, some vessels in those studies represented Large-Scale Fisheries (LSF), prompting the need for future validation of this methodology on high resolution SSF tracking.

Machine Learning showed a high performance to identify fishing activity, namely when using RaFo and XGBo which gave the best results. These are not complex methodologies and are available for everyone through Python sckitlearn (and in R, TidyModels), not requiring high computational resources. Yet alternative methods should be considered and evaluated also. In the current work we have tested Recurrent Neural networks which allow to analyse sequences (but were not equally successfully, as these were overestimating fishing activities), and in the future can be used to discriminate patterns between gears and fishing behaviours without losing the temporal information. These were also applied in previous works (Jiang et al., 2017). However, this type of analysis can take a lot of resources to optimise, without achieving higher performances than the ones currently obtained with simpler methods in the current work. Furthermore, this work focuses on improving the ML results by modelling the data itself with pre- or post-processing steps. Therefore, for the sake of reproducibility, these models were not considered for further optimization. Future works can incorporate other more sophisticated methods into the framework proposed.

(4) the number of variables to include within the model

There are several variables to potentially consider when applying ML algorithms to estimate fishing effort. Within the methods used the accuracy varied between 1 and 20 % when changing the number of variables from 2 to 5, with the models using 4 and 5 variables showing higher performances. However, the importance of the variables can vary between different metiers, and so, this aspect should be always evaluated in future works.

In our case studies, the speed variable contributed with a percentage of 74.54 % to accurately distinguish fishing events from non-fishing events. Similar results were obtained in previous works, highlighting the importance of speed in the classification (Marine Institute, Ireland ICES, 2022; Behivoke et al., 2021; O'Farrell, Shay et al. 2017; Souza et al., 2016; Hu et al., 2016). These works demonstrated that despite using different combination of variables, speed and time were always the most important variables to define fishing related activities. However, speed and time by themselves are not reliable enough to identify fishing activity, with less than 90 % accuracy in the current work which is in agreement with Behivoke et al. (2021), who demonstrated that a threshold based on speed only, achieved accuracies between 66 %–89 %. This might happen as for some fisheries the movement of setting the gear in the ocean is unpredictable and may have similar behaviours as non-fishing activities.

Furthermore, according to the method used, different types of variables can change and improve the algorithms' performance, such as differential coordinates (Hu et al., 2016, 89.2 % accuracy) and oceanographic variables (Sea Surface Temperature and bathymetry - Torres-Irineo et al., 2021, 95 % specificity). However, depending on the metiers, these variables might not be important. For example, in the current work variables such as bathymetry and distance to the coast did not strongly influence the algorithm performance (not included for brevity), probably due to the fishery operating very near coast. For any Metier, the variables to be used in the modelling process should always be evaluated previously. Oceanographic features, can strongly influence the target species distribution, thus are indirectly related to the fishing activity. Nevertheless, relying on external features may be more adequate for building a global model that could be used in other regions. Yet, it requires high-resolution oceanographic data availability, which is scarce in most coastal areas (<100 m depth) where typically SSF operate.

(5) the percentage of train/test data

There are not many publications that cover how different algorithms' performance in foretelling fishing activity is affected by the percentage of training/test data. One may believe that to implement the best model, there is always a requirement for a vast amount of data. However, there are situations when the explanatory power of the variables used to train the models is enough to reveal the underlying trends in the data. Therefore, having less data is not always a bad omen. This study examined how the models would perform from 10 % of the data to 90 % of the data, attaining accuracy levels of 80 % to 99 %. As for an operationalization of the models, models might be trained with 90 % of the data, tested with de remaining 10 % of the validation data and then, applied to the rest of the non-validated dataset (i.e., all fishing trips), thus it is essential to make this evaluation. The current work demonstrated that independently of the percentage of the training set and with the variables used, it is possible to achieve high performances. However, it is essential to always check for overfitting, which is very common, by implementing techniques as cross-validation, hyperparameter optimization, and regularization (Farnham et al., 2021; Zhang et al., 2023).

Brandoli et al. (2022) conducted a comparable evaluation, analysing various regression models to estimate catch per unit effort (CPUE). Among these approaches, Random Forest showed a logarithmic curve, achieving accuracies near 94 %. The performance evolution of this model resembled that observed in the present work. The authors also employed Gradient Boosting and XGBoost, both exhibiting similar behaviours to the performances observed in our study. Other algorithms, including Multi-Layer Perceptrons and Catboost, were tested, and a consistent pattern emerged across all - increasing the data used for model training led to improved performance. However, some models used by Brandoli et al. (2022) displayed a tendency to overfit, in contrast to our findings.

Despite increasing tendencies on the performance by incrementing the percentage of observations in the training dataset, it is noteworthy to mention that it depends on the case-study. In a similar work with cancer cells, comparing RaFo and XGBo results along different training/test size, Desdhanty and Rustam (2021) reported that the performances oscillated between 10 %–90 %, achieving a better score using 60 % of the data to train the model.

In addition to analysing the impact of the amount of data required for a model to perform well, it is crucial to keep in mind that onboard observers are very expensive and typically only cover a small portion of the fleet, whereas logbooks are created by fishermen and their quality depends on their willingness to accurately record every significant aspect of the fishing activity, which is not always the case. Therefore, studies such as Rufino et al. (2023) where a framework is provided to manually (expert based) validate boat trips could be used to increase or create a validated dataset. Furthermore, as mentioned by May Petry et al. (2020), new advanced tools such as Generative Adversarial Networks could be used to synthesize new behaviour data to compensate for the lack of validated data in some cases in future works.

(6) sampling methods (by point or by boat trip)

A desirable algorithm should have high performances, whether its task is to classify boat trips segments or random points of a boat trip. The current work explores two types of approaches: data sampling by point and by boat trip. Some works that implemented Hidden Markov Models (HMMs) cite that trip segments were used as sample unit which implies a similar approach to the one used when training RNNs (Marine Institute, Ireland ICES, 2022; Souza et al., 2016). Similar to the current work, Henriques et al. (2023) used the sample unit by boat trip approach to train the models, with the objective of classifying each point into one of the four different underlying hidden states of a fishing trip (steaming, deployment, hauling, and slow navigation) reaching accuracies above 90 %. These results are similar to the ones obtained in the current study (approximately 96 %).

In the current work, both sampling approaches showed a high score, however, these provide scattered classifications by the ML algorithms. The models could benefit from a different sampling strategy to avoid these results, such as sampling by segment. In this case, all the points within the segment would be classified as the labelled segmented given the segment classification, as opposed to labeling the data by point. Future works should attest this hypothesis and further evaluate the optimal segment length.

(7) the temporal resolution (intervals between consecutive points)

For broad applicability across various scenarios such as different gears, fisheries, and locations, the chosen method must exhibit consistent performance across diverse temporal resolutions. The necessary temporal resolution, defined by the time interval between successive location points, holds significance, particularly concerning the examination of shorter trips. For example, if the shorted trips last 60 min, and temporal resolution is 10 min, methods need to work with only 6 points, which would be hardly feasible in most statistical approaches. This, however, is an advantage to ML methods which do not require a minimum number of points to work, in spite of most cases the algorithms tend to underfit (Supplement 3), and some authors recommend that they should only be applicable in situations with thousands of cases (Farnham, B. et al., 2021; Zhang et al., 2023). While sampling by point did not exhibit improvement with increasing temporal resolution, XGBoost showcased enhanced performance with a higher ping rate, achieving an accuracy of approximately 98 %. In contrast, Gradient Boosting (GrBo) excelled with increased ping rates for sampling by boat, attaining an accuracy of around 94 %. Although GrBo outperformed in the sampling-by-boat approach, XGBoost still achieved an impressive accuracy of about 93 %. Nevertheless, Random Forest and GrBo produced the greatest results when sampling by point and boat, respectively, using lesser ping rates (30 s, 1 min, and 1:30 min).

Behivoke et al. (2021) utilized GPS devices, inferring regular 60-s intervals, and achieved accuracy scores of 74–89 % with Random Forest. Conversely, other studies employing the same device had larger intervals, ranging from 1 to 10 min (Marine Institute, Ireland ICES, 2022). Additionally, investigations conducted by DTU Aqua (ICES, 2022) involved information provided in 10-s intervals through EM validation. Although these latter works do not furnish specific results, the present project serves to exemplify the potential efficacy of the developed algorithms across diverse temporal resolutions. Henriques et al. (2023) reconstruct fishing tracks to have a timely consistence of datapoints, using an algorithm that relies on the Catmull–Rom approach, a modification of the hermit cubic spline algorithm. This interpolation was used to generate data points at 1-min intervals, achieving accuracies above 90 %, reinforcing the use of lesser ping rates.

The findings in this study align with the observations of Mendo (2019), supporting that 1-min intervals prove to be the most effective for identifying fishing events. Similarly, Julien Rodriguez, in ICES (2023), reinforces this idea, specifically to distinguish complete hauling and setting operations and compute accurate metrics for passive gears such as soak time and net lengths, a 1-min interval is optimal for small-scale fisheries (SSFs), while vessels over 15 m benefit from a slightly longer 2-min interval. This concurrence suggests that lower temporal resolutions, such as 1-min intervals, consistently yield better results in the identification of fishing activities.

Moreover, O'Farrell et al. (2017a, 2017b) provides an alternative for shorter trips or trips with hourly ping rates. The authors tested the performance of RaFo by exploring labelling the tracks by point and by window. For the first approach, an accuracy of 66 % was obtained. The authors cite that by point it may exhibit inconsistency in instances of the shorter trip durations, as there is no guarantee that a ping was recorded when the gear is deployed, consequently losing temporal information concerning to the initiation of fishing activities. Further, the second labelling methodology was proposed as a possibility to overcome this issue, whereas a window algorithm was implemented to define a record as fishing if the gears were deployed at any time during the hourly ping window surrounding that ping. By applying this procedure, it increased the balanced accuracy by 17 % (83 %). The authors provided an interesting alternative for larger temporal intervals but confirming that smaller intervals would be more effective for detecting fishing activity.

Nevertheless, there is a crucial requirement for a model capable of handling a diverse array of temporal resolutions. Despite the great performance obtained by the machine learning algorithm through the different temporal intervals, subsequent research endeavours should focus on analysing an algorithm with the capability to distinguish between fishing and non-fishing activities irrespective of the temporal resolution employed.

(8) number of isolated points to reclassify (post-processing) and number of fishing events classified

Moreover, tracking data can potentially profit from the time series structure, as well as the spatial features within the data to avoid sparse classifications, as demonstrated in the current work. ML algorithms were able to classify fishing activity with high accuracies but give scattered results with many isolated points misclassified. However, these misclassified points are determinant when discriminating between fishing events or mapping fishing activity. Future works should explore algorithms specifically developed for time series data as an alternative to avoid these issues as illustrated by Jiang, X., et al. (2017) who have implemented Recurrent Neural Networks with an accuracy of 89 %.

Besides evaluating the steps applied to the data before choosing and optimise the machine learning models, in the current work, we propose a post-processing approach to be also evaluated in the procedure, also. ML methods do not recognise that fishing trips are continuous time series, which makes the algorithms classify some isolated points as non-fishing during fishing periods, and the opposite. These points can be mathematically changed after the algorithm classification. In this case, we changed these points based on its neighbor's classification, revealing that the optimal number of neighbours to analyse would be two (the two points before and after). Henriques et al. (2023), proposed a similar approach to decrease the false positives, where these isolated datapoints would be labelled as the same state of the previous and next datapoints, if there were up to two different labelled datapoints from the previous and succeeding five datapoints. However, the improvements of applying this post-processing procedure were not reported in the author's work, neither it's number optimized. Souza et al. (2016), reduced the false positives (non-fishing points considered fishing activities) using two algorithms combined, namely, First-Passage Time algorithm (FPT) and Utilization Distribution algorithm (UD). The first algorithm tries to find areas where the patterns appear in a trajectory and the second one uses a kernel method clustering algorithm in the coordinate parameters. Applying these post-processing approaches offered an extra 1 % to 2 % accuracy improvement and a reduction of non-fishing activity false alarm comparing to the expert labels. These results are very similar to what was achieved to detect the number of events, whereas the methodology used improved the variance explained by the model (VEvc) in 2.5 % (0.9307 to 0.9549). The algorithms tested by the former author's seems promising, yet the methodology proposed offered similar improvements than these combined algorithms.

After reducing the isolated points on classified boat tracks, it was evaluated the number of fishing events estimated in individual trips, as a measure to decide which algorithm would be more adequate. In the current work, we identified those by applying a temporal threshold between fishing events estimated a priori, which was evaluated as we were optimising the post-processing procedure, and thus, obtaining the same VEvc score of 0.95. However, this methodology may not be the most optimal solution as it depends on the isolated points removed in the previous step and if all the points were well classified by the ML algorithm. Nonetheless, Random Forest, which was the algorithm with the most events identified, had only 19 events misidentified (two underestimated and seventeen overestimated) and it is noteworthy mention that a higher VEvc indicates that the number of events that were underestimated or overestimated, most were by one or two events.

Subsequent efforts should aim to refine the distinction between various fishing events, rather than relying on thresholds, which can vary significantly for each métier (zone and gear) and trip, a more effective approach could involve the implementation of machine learning, neural networks algorithms, or statistical methods to enhance the detection of these events or even changing the sampling strategy as it was mentioned previously. Moreover, Burgos et al. (2013) estimated how many hauls per vessel were being carried, which can directly provide an estimation of the number of events occurred in a single trip. The authors' predicted a number of hauls of 479, underestimating by 19 hauls the real number of hauls (498). The authors' algorithm, based on consecutive records, defined a haul when no speed increase beyond 4 knots lasting more than 6 min was observed, with records filtered to exclude speeds exceeding 4 knots. If the time interval between records exceeded 6 min, it signaled the end of a haul, with the subsequent series of records indicating the start of a new haul. The methodology proposed by Burgos et al. (2013) is interesting but requires experts of each fishery to provide their feedback on these events; however, in the case of a model that predicts the fishing activity independently of the gear used, this methodology might not be feasible. Julien Rodriguez, ICES (2023) and Henriques et al. (2023) implemented a methodology that distinguish hauling, setting, and non-fishing activities for static gears (e.g., nets and pots & traps). The author's work allows to indirectly estimate how many fishing events are occurring on one single trip as they classify hauling and setting events. However, there is no reports about estimating the number of events as we did in the current work. The author's work had accuracies above 90 %, which suggests that the identification of the number of fishing events could potentially have similar results to the ones achieved in this work.

(9) global considerations and conclusions

Considering all the options mentioned that were optimized using the proposed procedure, in the current case study the best model was Random Forest Classifier with all variables included (Latitude, Longitude, Speed, Time, and Months) without changing the temporal resolution and smoothing the speed information by applying a moving average of ten points. Besides these pre-processing steps, an additional post-processing algorithm is included to rectify isolated points. Further, it was decided that we should use the sample unit by boat, as the sample unit by point can be optimistic, because instead of using the full trips to train the models it uses random points from different trips, which is unrealistic for the type of data that we are working with (time-series).

Further work is required to test the methodology implemented in other fisheries and zones with more detail, which should be possible to do with an improved EU based data example.

One of the issues of using ML methods, is that the model may need to be trained for a new dataset if there are big differences between the original data that was used to train the model. Therefore, it might be hard to apply these algorithms. This could be a weakness of the methods scrutinized in the current work, especially as the coordinates that were used in the model, vary with the fishing grounds of each country. However, the model is adequate to handle a considerable amount of data as it was shown in Supplement 4. RaFo took less than 30 s to be trained on 90 % of the data and XGBoost less than 10 s. Nonetheless, to get an appropriate estimation of the algorithm scalability we should model the time, noting that it is not a simple proportion, but a more complex relationship between the number of observations, variables, and the number of parameters to be optimized. Regardless, it can be expected that for one million observations, that RaFo would take approximately 122 s to be trained on this amount of data and even less for XGBo, less than 3 s to predict more than 200,000 observations (Supplement 4). Aside from the machine learning modelling, all the steps evaluated (pre and post processing) also did not consume substantial computational resources, which makes this framework adequate to be implemented in larger datasets.

Therefore, this makes the framework suitable to apply in another regions, independently of the temporal resolution, métier, and tracking device. However, regarding the tracking devices, it is important to refer that AIS devices can comprise the fishermen privacy as the AIS system will leave the data available for everyone to access and buy it (a detailed comparison of tracking systems can be found in Quincoces, 2021). In contrast, when the vessels carry GNSS devices, only the company authorized to monitor these vessels has the ownership of the data.

Nonetheless, the use of coordinates as explanatory variables, might make it challenging to apply these modes in other regions, it worked well in other metiers not found in the training dataset, operating within the studies areas, as it was illustrated at the end of the results section.

Our findings strengthen that boat track devices, combined with expert validation, would improve the reliability of fishing efforts indicators in small-scale fisheries in Portugal, and contribute to more efficient management. Identical results were found in the several studies already cited along the paper.

Additionally, taking into consideration the inclusion of socio-economic factors to our modelling framework may provide a more thorough comprehension of the dynamics of fishing effort. Important factors influencing fishing behaviour include variables like fuel prices, vessel capacity, fish market prices, and fishing regulations. This information can be obtained from different sources such as landing profiles (species caught information - weight in kg and price per kg) or in the European Fleet Registry (for boats characteristics). By including these socioeconomic factors in our analysis, we may be able to better understand the regulatory pressures and financial incentives that drive fishing activity.

Souza et al. (2016) employed different algorithms for various gears, indicating significant diversity between gears and the distinct behaviours of each métier. In future research, it is essential to analyse the performance of creating individual models for each gear or métier and compare these results with a generalized model trained using all gears available in the dataset, as implemented in the current work.

Another essential aspect to consider is that our algorithm was only trained with data from 2017, and over the years changes related to the fishing habits, or even in the oceans could make the location of the activities change, for example. This issue could bring future misclassifications within our model, that only has data from 2017. Therefore, along the years we should search for a data drift that would indicate us that the model needs to be re-trained or use techniques such as online learning, so the model readjusts its parameters as new input variables arrive.

Finally, the estimation of soaking time was not considered. However, there are a few works that have been developed to estimate it (Henriques et al., 2023; Mendo et al., 2023). These authors were able to efficiently match setting events to hauling events and, therefore, estimate the soaking time (R-squared >90 % and accuracy >90 %, respectively). Nevertheless, future work should include testing different approaches from the ones used in these works (Buffers, time restrictions, and expert validation and Hidden Markov Model, respectively), such as neural networks or machine learning techniques, as it could be more efficient and less computationally expensive.

5. Conclusions

In this work, we develop a framework with a series of simple sequential steps that can improve the power of machine learning algorithms classification predictions, to estimate fishing effort from vessel's tracking data. The selection a better model, involved choosing the most promising options from each section based on their superior performances. To illustrate the procedure, we applied it to three case studies in Portuguese Small-Scale Fisheries fleets. The framework proposed do not imply a large computational burden.

The application of a moving average to boat speed, using 10 values was optimal, improving the model's performance by 0.6–3.39 % in accuracy. In our case study, with 4–5 variables only, namely latitude, longitude, speed, time, and/or months, most of the tested models reached high-performance scores, close or above 90 % when using both samplings by point and by boat trip. Even with only 10 % of the observations used to train the algorithm, the performance scores exceeded 95 % for sampling by point and surpassed 80 % for sampling by boat trip. Furthermore, using >60 % of the data for model training did not substantially enhance performance. Models achieved good performances using 30 s up to 10 min temporal resolution, and the decrease of time interval improved the model when sampling by boat was used, but decrease when using sample by point. As post-processing, replacing single values by the two closest neighbours of a point, improved the estimation of the number of fishing events.

In the current project this model could reach an accuracy of 99.07 % when predicting if a boat was fishing or non-fishing, and a VEvc of 0.95 when estimating the number of fishing events in a boat trip through this sequential simple steps. Therefore, we conclude that the framework proposed should be used in all boat's tracking data analysis using ML methods to classify fishers' behaviour.

CRediT authorship contribution statement

J. Samarão: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis. **A. Moreno:** Project administration, Funding acquisition. **M.B. Gaspar:** Resources, Project administration, Funding acquisition. **M.M. Rufino:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Conceptualization.

Acknowledgments

The authors acknowledge all active participants in WGSFD and WKSGEO, in particular Tania Mendo for valuable discussions on this subject. Marta M. Rufino is funded by a DL57 *contract (junior researcher)* awarded by IPMA within the project "Real-time monitoring of bivalve dredge fisheries" (MONTEREAL, MAR-01.03.02-FEAMP-0022), funded by the Fisheries Operational Programme (MAR 2020) and co-financed by the European Maritime and Fisheries Fund (EMFF 2014–2020) and by the project "Small scale fisheries and harvest monitoring program" (MOPPA) funded by the Fisheries Operational Programme (MAR-014.7.2-FEAMPA-0005). João Samarão received a research grant (ref: IPMA-2022-015-BII) awarded by IPMA within the framework of the project PESCAPANHA. The authors also acknowledge five anonymous referees for their useful comments on the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2024.102953.

Data availability

The data used in the current work cannot be provided due to confidentiality reasons. Annotated detailed python scripts can be found in https://github.com/joaosamarao/IMLP-SSF

References

Behivoke, F., Etienne, M.-P., Guitton, J., Ranaivoson, E., Léopold, M., 2021. Estimating fishing effort in small-scale fisheries using GPS tracking data and random forests. Ecol. Indic. 123, 107321. https://doi.org/10.1016/j.ecolind.2020.107321.

Bennett, Nathan James, Govan, Hugh, Satterfield, Terre, 2015. Ocean grabbing. Mar. Policy 57, 61–68. ISSN 0308-597X,. https://doi.org/10.1016/j.marpol.2015.03.026.

Brandoli, B., Raffaeta, A., Simeoni, M., Adibi, P., Bappee, F., Pranovi, F., Rovinelli, G., Russo, E., Silvestri, C., Soares, A., 2022. From multiple aspect trajectories to predictive analysis: a case study on fishing vessels in the Northern Adriatic Sea. GeoInformatica 26. https://doi.org/10.1007/s10707-022-00463-4.

Burgos, C., Gil, J., Olmo, L., 2013. The Spanish blackspot seabream (Pagellus bogaraveo) fishery in the Strait of Gibraltar: spatial distribution and fishing effort derived from a small-scale GPRS/GSM based fisheries vessel monitoring system. Aquat. Living Resour. 26, 399–407. https://doi.org/10.1051/alr/2013068.

Clarke, L.J., Hill, R.A., Ford, A., Herbert, R.J., Esteves, L.S., Stillman, R.A., 2019. Using remote sensing to quantify fishing effort and predict shorebird conflicts in an intertidal fishery. Eco. Inform. 50, 136–148.

Coro, G., Ellenbroek, A., Pagano, P., 2021. An open science approach to infer fishing activity pressure on stocks and biodiversity from vessel tracking data. Eco. Inform. 64, 101384.

Coro, G., Tassetti, A.N., Armelloni, E.N., et al., 2022. COVID-19 lockdowns reveal the resilience of Adriatic Sea fisheries to forced fishing effort reduction. Sci. Rep. 12, 1052. https://doi.org/10.1038/s41598-022-05142-w.

Desdhanty, V.S., Rustam, Z., 2021. Liver cancer classification using Random Forest and Extreme Gradient Boosting (XGBoost) with genetic algorithm as feature selection. In: 2021 International Conference on Decision Aid Sciences and Application (DASA), Sakheer, Bahrain, pp. 716–719. https://doi.org/10.1109/DASA53625.2021.9682311.

Farnham, B., Tokyo, S., Boston, B., Sebastopol, F., Beijing, T., 2021. Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow Concepts, Tools, and Techniques to Build Intelligent Systems second Edition.

Henriques, N., Russo, T., Bentes, L., Monteiro, P., Parisi, A., Magno, R., Oliveira, F., Erzini, K., Gonçalves, J., 2023. An approach to map and quantify the fishing effort of polyvalent passive gear fishing fleets using geospatial data. ICES J. Mar. Sci. 80. https://doi.org/10.1093/icesjms/fsad092.

Hori, T., Cho, J., Watanabe, S., 2018. End-to-end speech recognition with word-based Rnn language models. In: 2018 IEEE Spoken Language Technology Workshop (SLT), Athens, Greece, pp. 389–396. https://doi.org/10.1109/SLT.2018.8639693.

Hu, B., Jiang, X., Souza, E., Pelot, R., 2016. Identifying Fishing Activities from AIS Data with Conditional Random Fields. https://doi.org/10.15439/2016F546. October.

ICES, 2022, February 2. Workshop on Geo-Spatial Data for Small-Scale Fisheries (WKSSFGEO). ICES Scientific Reports. https://ices-library.figshare.com/articles/report/Workshop_on_Geo-Spatial_Data_for_Small-Scale_Fisheries_WKSSFGEO_/19248947/1 (Accessed 24 October 2023).

ICES, 2023, May 9. Workshop on Small Scale Fisheries and Geo-Spatial Data 2 (WKSSFGEO2). ICES Scientific Reports. https://ices-library.figshare.com/articles/report/Workshop_on_Small_Scale_Fisheries_and_Geo-Spatial_Data_2_WKSSFGEO2_/22789475/2 (Accessed 24 October 2023).

Jiang, X., Liu, X., Souza, E., Hu, B., Silver, D., 2017. Improving point-based AIS trajectory classification with partition-wise gated recurrent units, pp. 4044–4051. https://doi.org/10.1109/IJCNN.2017.7966366.

Liu, X., 2017. Deep Recurrent Neural Network for Protein Function Prediction from Sequence. ArXiv. /abs/1701.08318.

May Petry, L., Soares, A., Bogorny, V., Brandoli, B., 2020. Challenges in Vessel Behavior and Anomaly Detection: From Classical Machine Learning to Deep Learning, pp. 401–407. https://doi.org/10.1007/978-3-030-47358-7_41.

Mendo, T., 2019. Effect of temporal and spatial resolution on identification of fishing activities in small-scale fisheries using pots and traps. ICES J. Mar. Sci. 76. https://doi.org/10.1093/icesjms/fsz073.

Mendo, T., Glemarec, G., Mendo, J., Hjorleifsson, E., Smout, S., Northridge, S., Rodriguez, J., Mujal-Colilles, A., James, M., 2023. Estimating fishing effort from highly resolved geospatial data: focusing on passive gears. Ecol. Indic. 154, 110822. https://doi.org/10.1016/j.ecolind.2023.110822.

O'Farrell, Shay, Sanchirico, James N., Chollett, Iliana, Cockrell, Marcy, Murawski, Steven A., Watson, Jordan T., Haynie, Alan, Strelcheck, Andrew, Perruso, Larry, 2017a. Improving detection of short-duration fishing behaviour in vessel tracks by feature engineering of training data. ICES J. Mar. Sci. 74 (5), 1428–1436. https://doi.org/10.1093/icesjms/fsw244.

O'Farrell, Shay, Sanchirico, James N., Chollett, Iliana, Cockrell, Marcy, Murawski, Steven A., Watson, Jordan T., Haynie, Alan, Strelcheck, Andrew, Perruso, Larry, 2017b. Improving detection of short-duration fishing behaviour in vessel tracks by feature engineering of training data. ICES J. Mar. Sci. 74 (5), 1428–1436. https://doi.org/10.1093/icesjms/fsw244.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., Louppe, G., 2012. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12.

Quincoces, I., 2021. Research for PECH Committee – Workshop on Electronic Technologies for Fisheries - Part I: Transmitted Positional Data Systems. European Parliament, Policy Department for Structural and Cohesion Policies, Brussels. https://www.europarl.europa.eu/RegData/etudes/STUD/2021/652254/IPOL_STU(2021)652254_EN.pdf#page22.

Rufino, M.M., Mendo, T., Samarão, J., Gaspar, M.B., 2023. Estimating fishing effort in small-scale fisheries using high-resolution spatio-temporal tracking data (an implementation framework illustrated with case studies from Portugal). Ecol. Indic. 154, 110628. https://doi.org/10.1016/j.ecolind.2023.110628.

Russo, T., Parisi, A., Prorgi, M., Boccoli, F., Cignini, I., Tordoni, M., Cataudella, S., 2011. When behaviour reveals activity: assigning fishing effort to métiers based on VMS data using artificial neural networks. Fish. Res. 111 (1–2), 53–64.

Souza, E., Boerder, K., Worm, B., 2016. Improving fishing pattern detection from satellite AIS using data mining and machine learning. PLoS One 11, e0158248. https://doi.org/10.1371/journal.pone.0158248.

Subasi, Abdulhamit, Erçelebi, Ergun, 2005. Classification of EEG signals using neural network and logistic regression. Comput. Methods Prog. Biomed. 78 (2), 87–99. ISSN 0169–2607,. https://doi.org/10.1016/j.cmpb.2004.10.009.

Syed, S., Weber, C., 2018. Using machine learning to uncover latent research topics in fishery models. Rev. Fish. Sci. Aquac. 26, 1–18. https://doi.org/10.1080/23308249.2017.1416331.

Torres-Irineo, E., Salas, S., Euán, J., Palomo, L., Quiñones, D., Coronado, E., Joo, R., 2021. Spatio-temporal determination of small-scale vessels' fishing grounds using a vessel monitoring system in the southeastern Gulf of Mexico. Front. Mar. Sci. 8, 643318. https://doi.org/10.3389/fmars.2021.643318.

Tzotsos, A., Argialas, D., 2008. Support vector machine classification for object-based image analysis. In: Blaschke, T., Lang, S., Hay, G.J. (Eds.), Object-Based Image Analysis, Lecture Notes in Geoinformation and Cartography. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-77058-9_36.

Virtanen, Pauli, Gommers, Ralf, Oliphant, Travis E., Haberland, Matt, Reddy, Tyler, Cournapeau, David, Burovski, Evgeni, Peterson, Pearu, Weckesser, Warren, Bright, Jonathan, van der Walt, Stéfan J., Brett, Matthew, Wilson, Joshua, Millman, K. Jarrod, Mayorov, Nikolay, Nelson, Andrew R.J., Jones, Eric, Kern, Robert, Larson, Eric, Carey, C.J., Polat, İlhan, Feng, Yu, Moore, Eric W., VanderPlas, Jake, Laxalde, Denis, Perktold, Josef, Cimrman, Robert, Henriksen, Ian, Quintero, E.A., Harris, Charles R., Archibald, Anne M., Ribeiro, Antônio H., Pedregosa, Fabian, van Mulbregt, Paul, SciPy 1.0 Contributors, 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods 17 (3), 261–272.

Wang, Y., et al., 2023. PredRNN: a recurrent neural network for spatiotemporal predictive learning. IEEE Trans. Pattern Anal. Mach. Intell. 45 (2), 2208–2225, 1 Feb. https://doi.org/10.1109/TPAMI.2022.3165153.

Xu, Z., Wang, Z., 2019. A risk prediction model for type 2 diabetes based on weighted feature selection of Random Forest and XGBoost ensemble classifier. In: 2019 Eleventh International Conference on Advanced Computational Intelligence (ICACI), Guilin, China, pp. 278–283. https://doi.org/10.1109/ICACI.2019.8778622.

Zhang, Aston, Lipton, Zachary C., Li, Mu, Smola, Alexander J., 2023. Dive into Deep Learning. Cambridge University Press. https://D2L.ai.