SoftwareX 34 (2026) 102579

Contents lists available at ScienceDirect

SoftwareX

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

Original Software Publication

Predicting fishing vs. not-fishing in small-scale fisheries: A sample vessel tracking dataset and a reproducible machine learning approach

Pamela Lattanzi [a,b,c,*], Claudio Vasapollo [a], Joao Samarão [d,e], Alessandro Galdelli [c,f], Tania Mendo [g], Marta Rufino [h,i], Luca Bolognini [a,c], Anna Nora Tassetti [a,c]

[a] National Research Council, Institute of Marine Biological Resources and Biotechnologies (CNR-IRBIM), Ancona, Italy
[b] Department of Biological, Geological and Environmental Sciences (BiGeA), University of Bologna, Bologna, Italy
[c] NBFC, National Biodiversity Future Center, Palermo, Italy
[d] Infratrading Lda, Lisboa, Portugal
[e] Nova School of Science and Technology, Almada, Portugal
[f] VRAI Lab, Dipartimento di Ingegneria dell'Informazione, Università Politecnica delle Marche, 60131 Ancona, Italy
[g] School of Geography and Sustainable Development, University of St. Andrews, KY16 9AL, St. Andrews, UK
[h] Divisão de Modelação e Gestão de Recursos Pesqueiros, IPMA—Instituto Português do Mar e da Atmosfera, Lisboa, 1495-165, Portugal
[i] Faculty of Sciences, Centre of Statistics and its Applications (CEAUL), University of Lisbon, Lisbon, 1749-016, Portugal

ARTICLE INFO

Dataset link: Data for Predicting Fishing vs. Not-Fishing in Small-Scale Fisheries: a Sample Vessel Tracking Dataset and a Reproducible Machine Learning Approach (Original data)

Keywords:
Small-scale fisheries
High-resolution tracking data
Prediction
Passive gears
Activity status
Machine Learning classification

ABSTRACT

Predicting fishing activity from vessel tracking data is crucial for quantifying fishing effort. This study addresses this challenge by classifying the fishing versus non-fishing activity status of small-scale vessels using passive gears, with a suite of different algorithms, ranging from basic statistics (Logistic Regression - *LoRe*) to Machine Learning (Decision Trees - *Dtree*, Random Forests - *RaFo*, and Extreme Gradient Boosting - *XGBo*). Results demonstrate that the Machine Learning (ML) ensemble significantly outperformed *LoRe*, especially with *XGBo* and *Dtree* achieving comparable high accuracy and robustness across training, validation, and test sets. By employing SHAP (SHapley Additive exPlanations), we demonstrate that the vessel speed *(SPEED)* and course variations (*course_diff*), the hour of the day (*hours*), and the distance from the coast *(distance_from_coast)* or the bathymetric depth *(depth)*, are the primary mechanistic drivers for discerning fishing operations in passive-gear small-scale fisheries (SSF). We provide a fully reproducible workflow and a unique, high-resolution dataset of manually labelled tracking data to address the critical scarcity of validated resources in this field. This framework provides a timely, scalable solution for high-resolution tracking analysis, directly addressing the technical needs arising from upcoming EU mandates (Control Regulation 2023/2842) for small-scale vessel monitoring. The shared code and data enable researchers to evaluate model transferability and generalisation, providing a standardised approach to harmonise fishing effort estimation across diverse geographic contexts. Finally, the provided code is structured as an accessible framework for fisheries scientists with limited ML experience, offering a practical foundation for implementing automated activity classification.

Metadata

Nr	Code metadata description	Metadata
C1	Current code version	v2.0
C2	Permanent link to code/repository used for this code version	https://github.com/PamelaLattanzi/softwarex_paper.git https://doi.org/10.5281/zenodo.18407451
C3	Permanent link to reproducible capsule	-
C4	Legal code license	GNU General Public License v3.0
C5	Code versioning system used	Git
C6	Software code languages, tools and services used	Python

(*continued on next page*)

* Corresponding author.
E-mail address: pamela.lattanzi@irbim.cnr.it (P. Lattanzi).

https://doi.org/10.1016/j.softx.2026.102579
Received 13 November 2025; Received in revised form 29 January 2026; Accepted 20 February 2026
Available online 28 February 2026

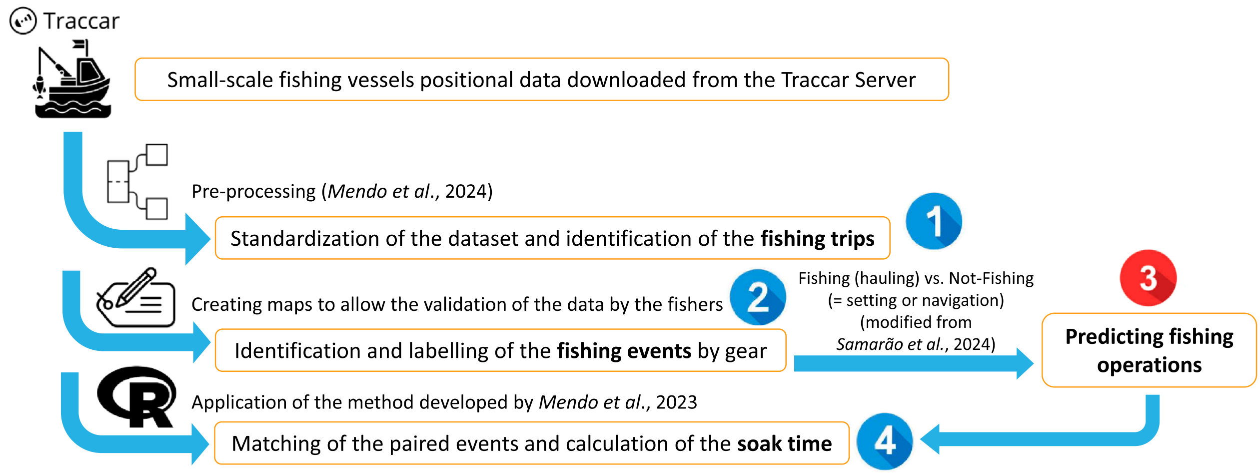

Fig. 1. The proposed pipeline for automated high-resolution vessel tracking data analysis. The process begins with the download of positional data from the storage server (in this case, Traccar - https://www.traccar.org/), followed by data standardisation and trip identification using the code provided by [26]. Next, fishing events are defined based on expert knowledge (distinguishing fishing from not_fishing), and subsequently validated by fishers, who visually inspect vessel tracks on printed maps and annotate the corresponding fishing gear. Finally, the code described in this paper is applied to automatically classify fishing operations, enabling the matching of paired setting/hauling events and the calculation of soak time, with additional routines available from [23].

(*continued*)

Nr	Code metadata description	Metadata
C7	Compilation requirements, operating environments and dependencies	Python 3.12.6, Jupyter https://github.com/PamelaLattanzi/softwarex_paper/blob/master/requirements.txt
C8	If available, link to developer documentation/manual	https://github.com/PamelaLattanzi/softwarex_paper/blob/master/README.md
C9	Support email for questions	pamela.lattanzi @irbim.cnr.it

1. Motivation and significance

Nowadays, a large amount of geospatial data is available from remote-sensing technologies, such as Synthetic Aperture Radar imagery [1], and ship-reporting systems, including the Vessel Monitoring System (VMS; [2–4]) and Automatic Identification System (AIS; [5–7]). These sources enable large-scale fisheries (vessels over 12m in length) to be tracked and their fishing effort to be assessed [8–12]. In contrast, small-scale fisheries (SSF), defined as vessels under 12 m in length and not using towed gears - as per CE Regulation N° 508/2014 [13] - remain largely underrepresented. This is a major challenge, as SSFs represent 80%+ of the world's fleet. For example, in the Mediterranean and Black Seas, SSF account for 84% of the total fleet and contribute 24% of the total landing value [14].

Commonly, SSF vessels employ passive gears, such as nets and traps, varying widely between target species, market demand, and season, which influences where and when the boats operate. Fishing activity using passive gears generally involves two separate actions: *setting* the gear and, after a certain period, *hauling* it. The timing of retrieval (hauling) can also be influenced by weather. These activities further follow regional regulations that limit the number, length and specifications of the gears and, in some cases, fishing quotas.

Accurately quantifying SSF spatio-temporal fishing effort can provide scientists, policymakers and managers with the means to evaluate the potential impacts on the marine ecosystems and identify possible spatial conflicts with overlapping activities such as aquaculture, other fishing fleets, etc [15–18]. Previous works have compared a plethora of methods to evaluate fishers' behaviour (e.g., evaluate whether SSF vessels are fishing or not) [19–21]. It is clear that SSF require high resolution tracking, with intervals on the scale of minutes to seconds [22], which will produce a large amount of data on the spatio-temporal dynamics of SSF. However, there is still the lack of a complete, publicly available workflow that provides the full analytical code, with a particular focus on hyperparameters tuning for the detection of possible model overfitting[1] (in the case of Machine Learning (ML) applications) and evaluation of predictors importance, and a labelled dataset for testing it.

To address this gap, the present work aims to supply and share a comprehensive and fully replicable approach through a Python code designed for classifying fishing versus non-fishing activity in each geoposition of a given trip. In contrast to previous studies (e.g., [23,24, 19–21]), this approach explicitly documents and evaluates models across separate training, validation, and test sets, providing the used tuning parameters and assessing the influence of the variables used as predictors on models' performances. In addition, an anonymised sample dataset of 846 SSF trips employing passive gears (i.e., gillnets – GNS, trammel nets – GTR, and pots – FPO [25]) is provided (see Supplementary Table S1 for additional information). These trips were recorded with GNSS tracking devices (i.e., Global Positioning System, GPS[2]) and allow direct testing, reproducibility and adaptation of the workflow to other fisheries and contexts.

Given that a potential pipeline for automatically estimating fishing effort from high-resolution passive-gear vessel tracking data requires labelling *fishing* and *non-fishing* operations (a highly time-consuming task if done manually), this study addresses the challenge by automating the process and moving towards the prediction of fishing operations to ensure scalability, particularly for large datasets (Fig. 1). This step is critical, as precisely defining where and when fishing events occur is key to understanding the main grounds exploited in a given area and the implications of their use. In this case, dealing with passive gears, *fishing* activities were identified only as *hauling*, whereas the deployment phase (*setting*) was grouped together with *navigation* under the category "*not fishing*". This was done to avoid overestimating fishing effort by ensuring that *setting* and *hauling* are treated as parts of a single fishing event rather than separate operations.

[1] i.e., when a model learns not only the underlying patterns in the training data but also random noise or dataset-specific details, leading to poor performance on unseen data.

[2] The original ping rate is 60 seconds, but positions were interpolated to obtain a more conservative 30 second-resolution, as suggested by [22].

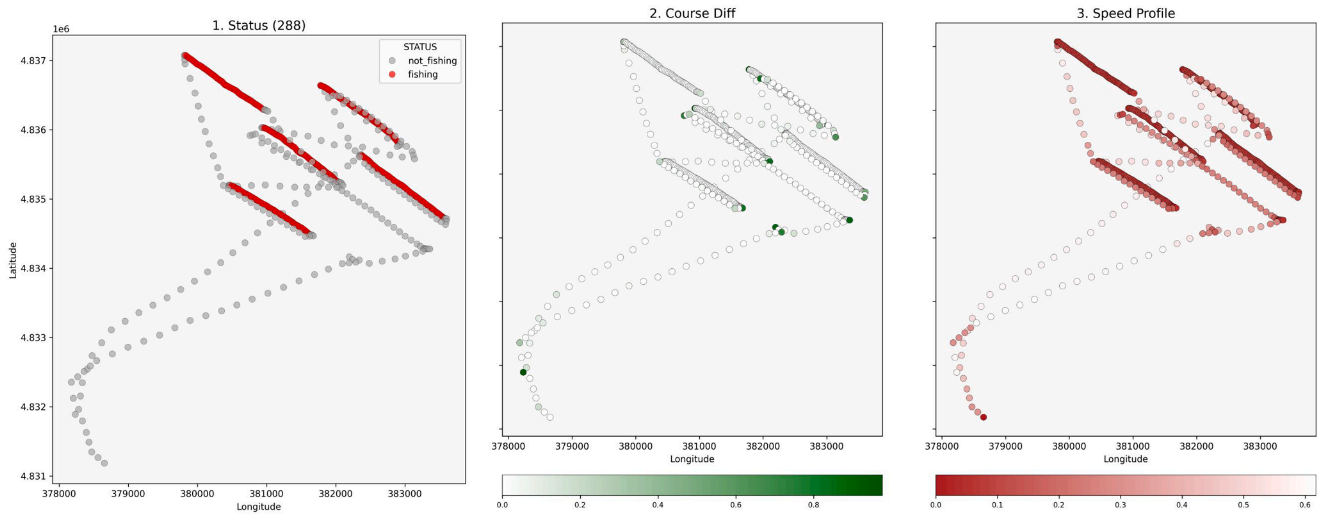

Fig. 2. Exemplification of a fishing trip, with fishing (red) and not-fishing (gray) positions (left panel). Values of course_diff (i.e., difference in course between consecutive points, dark green = high values, middle panel) and SPEED (red = low values, right panel) are mapped too.

2. Software description

2.1. *Software architecture*

The software is designed as a modular Python-based workflow shared via GitHub/Zenodo [27]. It follows a pipeline-oriented architecture that automates vessel tracking data analysis, predicting the activity status for each geo-position within a fishing trip. The core system consists of:

- Data Processing Engine and Model Wrapper: A main Python script ("1_predict_activity_status_parallel.py") that ingests kinematic and temporal features (prioritizing behavioral predictors, such as vessel speed, over location-specific coordinates to ensure geographic transferability) and handles parallelized model execution and cross-validation.
- Visualization & Interpretability Layer: A Jupyter notebook ("2_results_visualization_predict_activity_status_parallel.ipynb") for result inspection and SHAP-based feature importance analysis.
- To facilitate immediate testing, the architecture includes a sample dataset ("anonymized_dataset.rds") and pre-computed output files ("model_performances_on_data_splitting_with_shap.csv" and "dataset_with_predictions.rds"), allowing users to explore the visualization module without re-running the main Python script.

2.2. *Software functionalities*

The software enables a systematic approach to tracking data analysis by automating the benchmarking of multiple algorithms, allowing for a direct performance comparison between traditional statistical methods and advanced ML models. To ensure model reliability, the workflow integrates an automated hyperparameter optimization routine via a *RandomizedSearchCV* interface, which is strictly nested within an evaluation loop to prevent data leakage and ensure unbiased results. Beyond classification, the tool provides built-in routines for extensive evaluation through the generation of diverse performance metrics and diagnostic plots. Finally, the integration of SHAP (SHapley Additive exPlanations) serves as a critical explainability layer, transforming black-box ML decisions into transparent, interpretable insights by quantifying the specific contribution of each feature to the model's predictions.

3. Illustrative examples

To demonstrate the software's utility, the workflow was tested on a case study involving a dataset of vessel tracking data collected from five small-scale fishing vessels operating out of an Italian harbour. These vessels were equipped with GNSS tracking devices that recorded their position every minute (temporal resolution), along with vessel ID and timestamps; the data was subsequently anonymised [29]. The used architecture was described by [30]. This sample dataset, made available in the same repository, allows direct testing of the described code. Further details are provided in the Supplementary Table S1. The distinction between *fishing* and *not_fishing* was made based on the expert opinion and subsequently validated by the fishers, who visually inspected vessel tracks on printed maps and annotated the gear employed.

For classifying the activity status, a statistical model (Logistic Regression – *LoRe*) and a set of ML algorithms (Decision Tree - *Dtree*, Random Forest - *RaFo*, Extreme Gradient Boosting - *XGBo*) were chosen to provide a range of model complexity. Deep Learning algorithms were avoided, as they are highly resource-demanding and usually require large, well-validated datasets to be trained, which are not always available.

The central component of the model development workflow is the nested cross-validation (CV) approach: it involves an outer loop with 5 folds, where the model is trained on four folds (70% of the fishing trips of the input dataset) and validated on the remaining part (20%) to assess its generalisation ability. Within each outer training fold, an inner 3-fold CV is conducted for hyperparameter tuning, using *RandomizedSearchCV* function from the Python *scikit-learn* library [28] to systematically select the most effective settings. This dual-layered strategy ensures that hyperparameter tuning does not influence the final performance evaluation, which is conducted on the holdout test set (10% of the fishing trips). This procedure provides a robust assessment of the model's true predictive power, while the tuning step is crucial to reduce overfitting, a common issue in ML where models memorise training data rather than learning generalizable patterns. The specific settings are detailed in Supplementary Figure S1.

Model performance was evaluated using the following metrics: Accuracy, Precision, Recall (Sensitivity), Specificity, F1 Score and AUC Score. Moreover, a Principal Component Analysis (PCA) was undertaken to understand how each metric is related to the others.

Concerning predictors, a total of 7 variables (i.e., *SPEED, distance_from_coast* OR *depth, course_diff, time_seconds, trip_duration, hours, months*) were selected and tested. Being *depth* and *distance_from_coast*

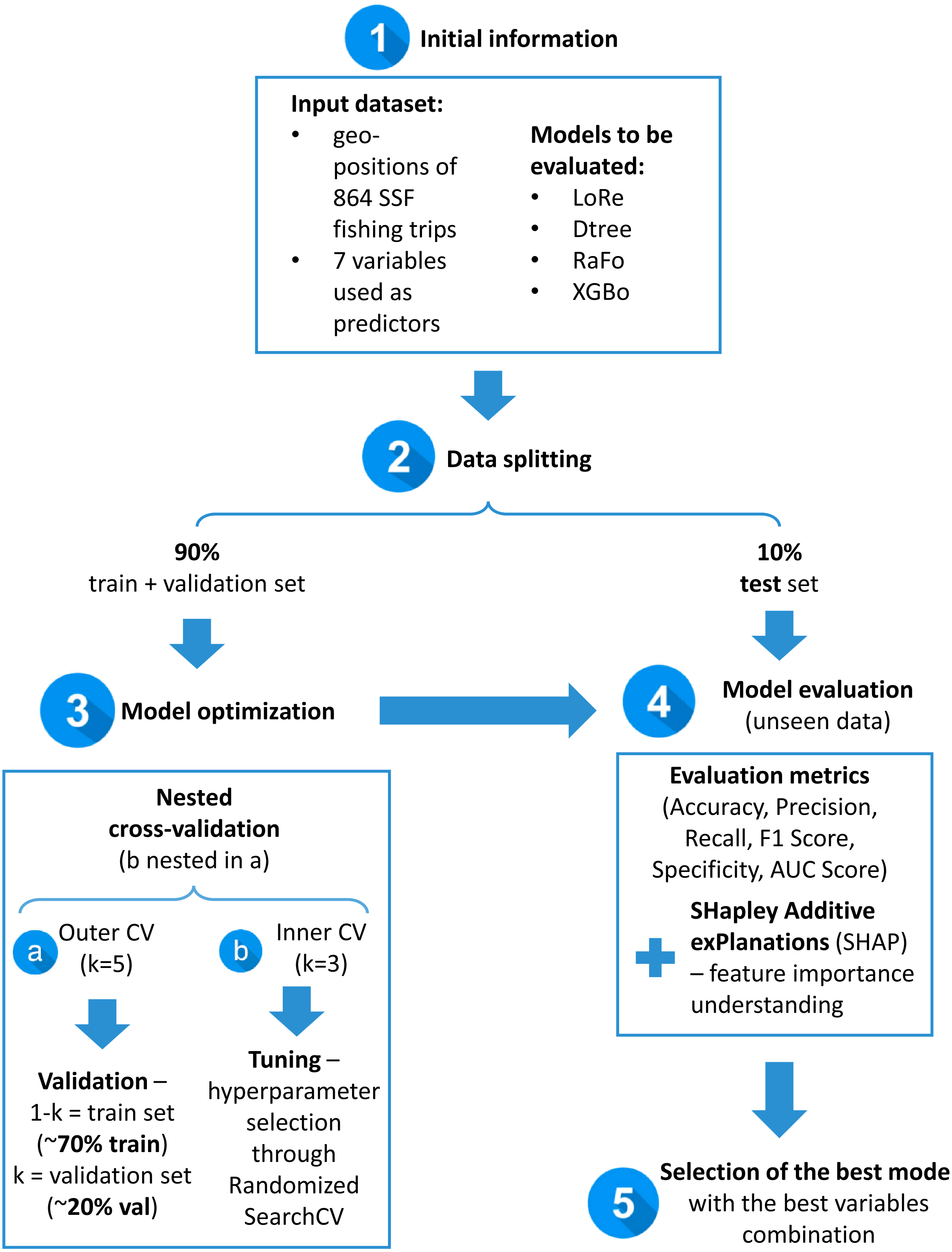

Fig. 3. Workflow of the model development process.

highly correlated (Supplementary Figure S2), the aforementioned models were tested once with *depth* and once with *distance_from_coast*. In this section, authors provide the variable combination including *distance_from_coast*, but similar results were obtained using *depth* and can be visualized in Supplementary Figure S5. To ensure the geographic transferability of the model, authors intentionally excluded longitude and latitude as predictors. Including spatial coordinates often leads to 'location-memorization' (spatial overfitting), where the model learns

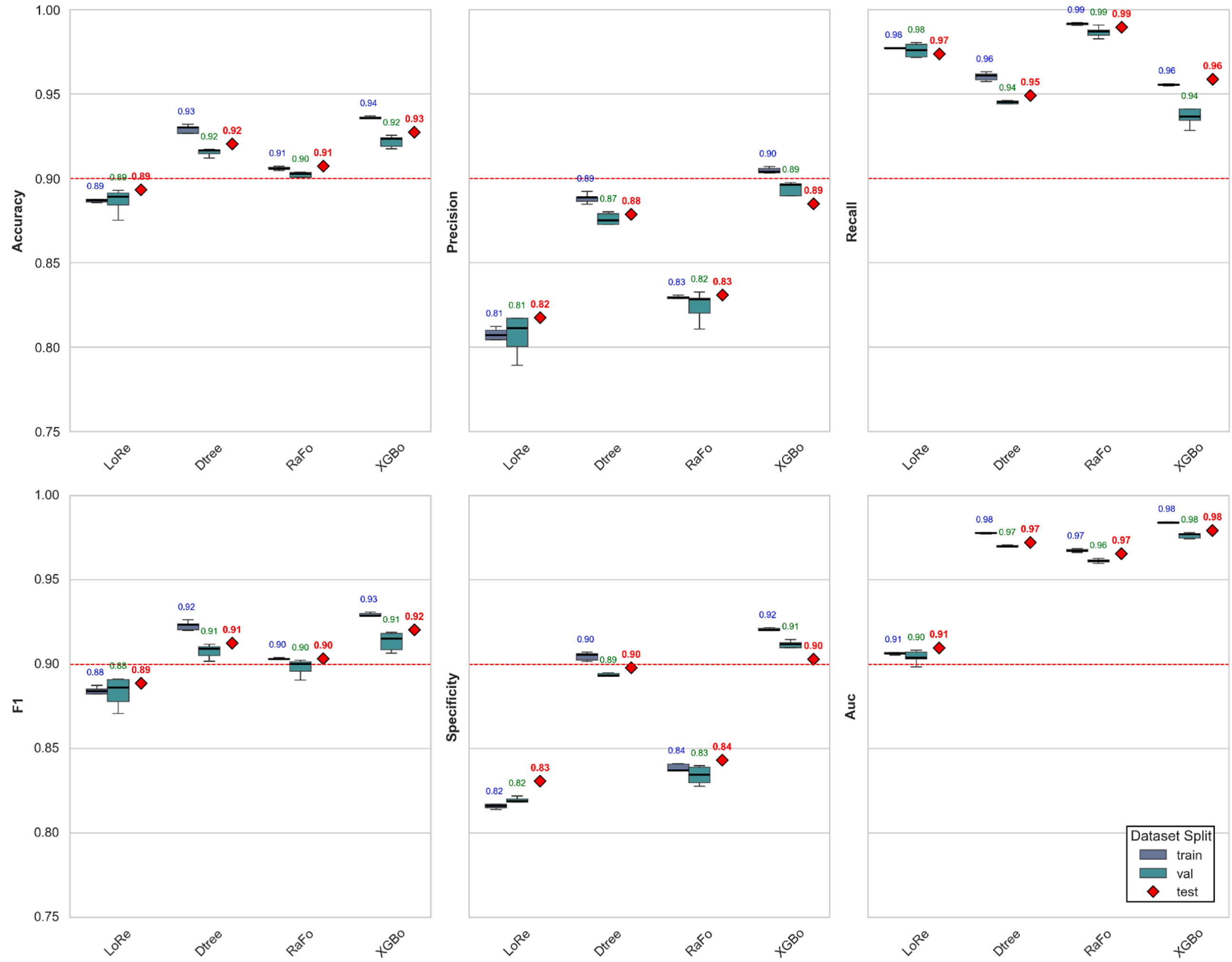

Fig. 4. Performances (i.e., accuracy, precision, recall, specificity, F1 score, AUC score) for the 4 classification models (LoRe, Dtree, RaFo, XGBo), with data splitting into train, validation and test sets. Boxplots display the distribution of training (blue) and validation (green) scores across CV folds. The blue and green numeric labels indicate the mean score for their respective sets. The red diamond and its red numeric label show the final, single score achieved on the independent test set. A 90% threshold (red dashed line) is included for reference.

specific local fishing grounds rather than the universal behavioral patterns of the vessel. By relying solely on kinematic features (e.g., *SPEED, course_diff* – Fig. 2) and environmental variables (e.g., *depth* OR *distance_from_coast*), the model remains applicable to other maritime regions where spatial hotspots may differ. Indeed, the features selected for this study represent the spatial, temporal, and kinematic characteristics of vessel behaviour that define the transition between navigation/setting and fishing/hauling. *SPEED* should be a primary discriminator, as hauling operations in this case study occur at significantly lower velocities compared to setting or transit. To capture the spatial logic of the fishing trip, *distance_from_coast* OR *depth* were included; the former may help identifying the distal points of a journey where hauling typically happens, while the latter serves as a proxy for specific fishing grounds and gear-target species locations. Vessel manoeuvrability is captured via *course_diff* (i.e., difference in course between consecutive points), which highlights the directional shifts inherent in gear handling. Finally, the temporal structure of the operations is defined by *time_seconds, trip_duration, hours*, and *months* (the latter added as suggested by [22]). These variables allow the model to learn the cyclical and rhythmic nature of fishing, such as the preference for sunrise-hauling over sunset-setting, and the mid-trip timing characteristic of hauling events. A further description of each variable was provided in the Supplementary Table S2 and in the “README.txt” of the shared repository.

Besides, SHAP (SHapley Additive exPlanations) values were calculated on the test set for each model using the 7 aforementioned variables, including *distance_from_coast*, to quantify the contribution of individual features to the model’s predictions. A SHAP beeswarm plot was employed for each model to visualize feature importance, providing a detailed view of how each predictor’s value influences the model's decision-making process and highlighting the trade-offs between temporal and spatial drivers. In particular, this allowed for an assessment of how high or low values of selected variables positively or negatively contribute to the classification of fishing activity. Additional outputs, such as the aggregated and global mean absolute SHAP values for each feature were provided as Supplementary Figures S4.

The entire data processing is summarised in the workflow portrayed in Fig. 3.

Hyperparameter tuning ensured that all models remained well-generalized, with performance metrics remaining consistent across training, validation, and test sets, and showing minimal overfitting (Fig. 4). Among the evaluated algorithms, *XGBo* emerged as the most robust, achieving a balanced F1 Score of 92.0% and a peak AUC of 97.9% (Fig. 5). While *RaFo* yielded a slightly higher raw recall (99.0%), its lower precision compared to *XGBo* indicated a tendency to over-predict fishing events. This inverse relationship between recall and other performance metrics is further supported by the Principal Component Analysis (PCA) biplot of the test set results (Supplementary Figure S3), which shows recall following a distinct performance

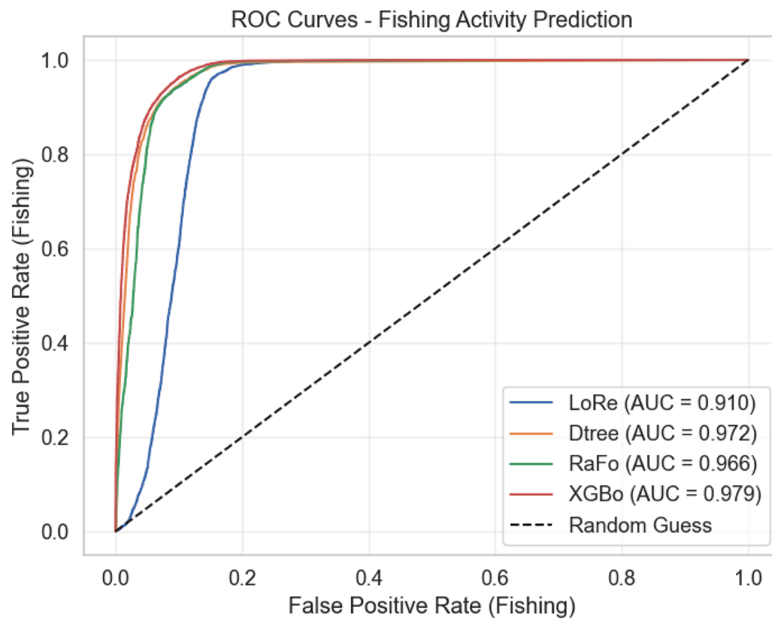

Fig. 5. Receiver Operating Characteristic (ROC) curves for the 4 classification models. The plot illustrates the diagnostic ability of each model by mapping the True Positive Rate (Recall = Sensitivity) against the False Positive Rate (1-Specificity) across all possible classification thresholds. The XGBo model (top red curve) demonstrates superior predictive performance with an AUC of 0.979, indicating a near-perfect separation between fishing and not-fishing activities. The diagonal dashed line represents the performance of a random classifier (AUC = 0.5), serving as a baseline for model efficacy.

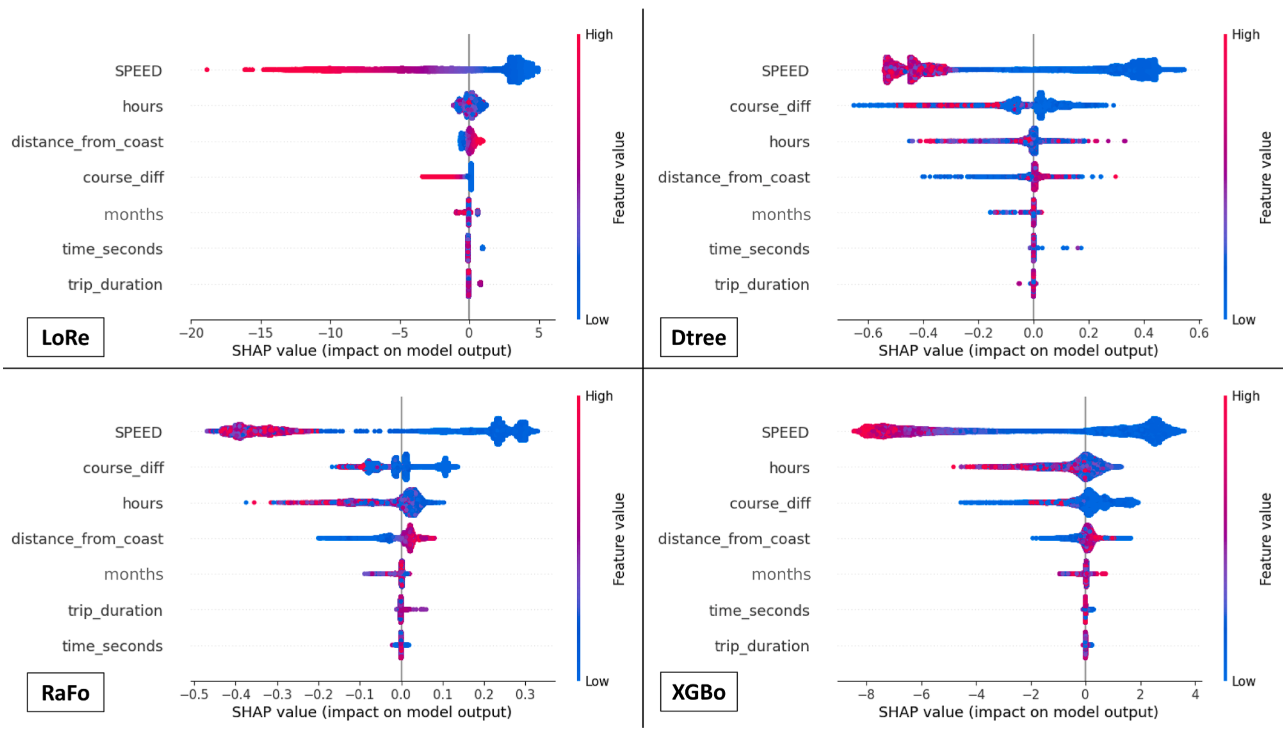

Fig. 6. SHAP beeswarm plots of feature influence on fishing activity classification. Each plot, one for each model, illustrates the distribution of the impacts each feature has on the model output for the test set, arranged in a top-down ranking based on their overall importance. Each dot represents an individual observation; its position on the x-axis indicates whether that feature's value increased (positive SHAP value, on the right) or decreased (negative SHAP value, on the left) the probability of the positive class (fishing_label = 0). The colour gradient represents the feature value, from low (blue) to high (red). For instance, lower speeds (blue) are strongly associated with higher SHAP values (on the right of the plot), indicating a significant contribution toward the fishing classification.

trajectory opposite to the precision-specificity cluster. The competitive performance of the *LoRe*, achieving approximately 89% accuracy compared to the 91-93% seen in complex ensemble methods, can be primarily attributed to the high predictive power of a few key features – most notably vessel *SPEED*, as can be observed from the SHAP beeswarm plots (Fig. 6). Indeed, the SHAP analysis identifies *SPEED* as the primary discriminator, the dominant signal across all models, suggesting that the decision boundary for identifying fishing in this context is largely linear or defined by clear thresholding, which *LoRe* captures effectively. This is further evidenced by its high recall (~97%), indicating that a simpler linear approach is highly efficient at identifying the target class, even if with a slight trade-off in precision (ensuring fewer missed detections at the cost of higher false alarms). Similarly, the marginal performance gap between the *Dtree* and *XGBo* suggests a state of information saturation; since a single well-tuned tree can already capture the bulk of the variance provided by primary features like *SPEED* and *course_diff*, the iterative boosting process of *XGBo* finds few residuals to correct. However, these results are likely influenced by the constrained study area, involving only five vessels with similar operational profiles from a single port. In this specific context, models may inadvertently rely on site-specific temporal patterns rather than generalized behaviours. As this workflow scales to more diverse fleets or geographical regions, the advantages of *XGBo* are expected to become more pronounced. Its sophisticated regularization and learning rate optimization (hyperparameters that can be tuned) allow it to adapt to the complex, non-linear relationships that emerge in heterogeneous datasets where simpler models often fail to generalize.

4. Impact

The workflow was developed to advance research on fishing event recognition, contributing to the automation of activity classification and providing a scalable foundation for high-resolution vessel tracking data analysis. Given the adoption of Regulation (EU) 2023/2842 [31], in Europe, small-scale fishing vessels will be required to install tracking devices onboard from 2028, with a derogation for vessels under 9 m lasting until 31 December 2029; from 2030 onwards, full compliance will be mandatory. Thus, the present work offers a structured and standardised approach for defining fishing events, which is applicable to various contexts, including those beyond SSF.

While many existing studies limit their evaluation to test set performance, this manuscript provides insight on training and validation sets as well. Furthermore, authors investigated predictor influence through SHAP beeswarm plots for each model. This approach moves beyond simple performance metrics to provide a mechanistic understanding of how specific variable values drive the classification of fishing versus non-fishing activity.

Furthermore, the shared sample dataset is unique, as it furnishes validated, labelled data on fishing operations, which is quite difficult to find elsewhere. High-resolution, validated datasets like this are essential for testing method reproducibility and transferability across different case studies and geographic areas. Moreover, the provided code is structured as an accessible framework for fisheries scientists with limited ML experience, offering a practical foundation for implementing automated activity classification. This work follows up on an initial attempt to collect this kind of information made during the ICES Second Workshop on SSF Geo-Spatial Data (WKSSFGEO2) [29] and the ICES Working Group on Spatial Fisheries Data (WGSFD) [32].

5. Conclusions

Automatically distinguishing between fishing and non-fishing is an important move toward a harmonised and automated pipeline for analysing tracking data. By identifying and utilising the right combination of predictors and employing a robust ML model like *XGBo*, this research provides a reliable method for classifying fishing events. Furthermore, the use of SHAP values further enhances interpretability, providing clear insights into the variables driving model predictions. To ensure future robustness, subsequent implementations should prioritize generalization-focused strategies, such as Leave-One-Vessel-Out (LOVO) CV and the application of the workflow to other case studies. While the authors acknowledge the limitations of the furnished data in terms of fleet and spatial representativeness, the present work still provides valuable insights where data scarcity is often the norm.

CRediT authorship contribution statement

Pamela Lattanzi: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Claudio Vasapollo:** Writing – review & editing, Supervision, Methodology, Formal analysis. **Joao Samarão:** Writing – review & editing, Supervision, Software, Methodology, Formal analysis. **Alessandro Galdelli:** Writing – review & editing, Supervision, Software, Methodology, Formal analysis, Conceptualization. **Tania Mendo:** Writing – review & editing, Visualization, Supervision, Methodology, Formal analysis, Conceptualization. **Marta Rufino:** Writing – review & editing, Supervision, Methodology, Formal analysis. **Luca Bolognini:** Writing – review & editing, Resources, Funding acquisition. **Anna Nora Tassetti:** Writing – review & editing, Visualization, Supervision, Resources, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank the small-scale fishing organization and the participant fishers who actively contributed to the data collection within the framework of the completed project ARGOS – ShARed GOvernance of Sustainable fisheries and aquaculture activities as leverage to Protect Marine Resources in the Adriatic Sea, funded by the European Programme Interreg Italy-Croatia – where fishing vessels were equipped with GPS devices.

PL, ANT, AG and LB were supported by the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.4 - Call for tender No. 3138 of 16 December 2021, rectified by Decree n.3175 of 18 December 2021 of Italian Ministry of University and Research funded by the European Union – NextGenerationEU, in the framework of the NBFC - National Biodiversity Future Center (project code CN_00000033, Concession Decree No. 1034 of 17 June 2022 adopted by the Italian Ministry of University and Research, CUP B83C22002930006). This study was for PL a contribution to the research that was conceived within the International FishMed PhD Program "Innovative technologies and sustainable use of Mediterranean Sea fishery and biological resources". Authors sincerely express gratitude to the reviewers for their valuable comments and insightful suggestions, which have significantly improved the quality and clarity of the manuscript.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.softx.2026.102579.

Data availability

Data for Predicting Fishing vs. Not-Fishing in Small-Scale Fisheries: a Sample Vessel Tracking Dataset and a Reproducible Machine Learning Approach (Original data) (GitHub).

References

[1] Guida R, Rodger M, Bissonauth V, Ragoomundun N, Soreefan Z, Hurnath P, Matthews M. Towards satellite-based indicators of economic loss due to illegal unreported and unregulated fishing. IGARSS 2024-2024 IEEE International geoscience and remote sensing symposium. IEEE; 2024. p. 3988–91. https://doi.org/10.1109/IGARSS53475.2024.10642056.

[2] Amoroso RO, Pitcher CR, Rijnsdorp AD, McConnaughey RA, Parma AM, Suuronen P, Eigaard OR, Bastardie F, Hintzen NT, Althaus F, Baird SJ, Black J, Buhl-Mortensen L, Campbell AB, Catarino R, Collie J, Cowan JH, Durholtz D, Engstrom N, Jennings S. Bottom trawl fishing footprints on the world's continental shelves. Proc Natl Acad Sci 2018;115(43). https://doi.org/10.1073/pnas.1802379115.

[3] Bastardie F, Nielsen JR, Ulrich C, Egekvist J, Degel H. Detailed mapping of fishing effort and landings by coupling fishing logbooks with satellite-recorded vessel geo-location. Fish Res 2010;106(1):41–53. https://doi.org/10.1016/j.fishres.2010.06.016.

[4] Doherty PD, Atsango BC, Ngassiki G, Ngouembe A, Bréheret N, Chauvet E, Godley BJ, Machin L, Moundzoho BD, Parnell RJ, Metcalfe K. Threats of illegal, unregulated, and unreported fishing to biodiversity and food security in the Republic of the Congo. Conserv Biol 2021;35(5):1463–72. https://doi.org/10.1111/cobi.13723.

[5] Le Guyader D, Ray C, Gourmelon F, Brosset D. Defining high-resolution dredge fishing grounds with Automatic Identification System (AIS) data. Aquat Living Resour 2017;30:39. https://doi.org/10.1051/alr/2017038.

[6] Natale F, Gibin M, Alessandrini A, Vespe M, Paulrud A. Mapping Fishing Effort through AIS Data. PLOS ONE 2015;10(6):e0130746. https://doi.org/10.1371/journal.pone.0130746.

[7] Tassetti AN, Ferrà C, Fabi G. Rating the effectiveness of fishery-regulated areas with AIS data. Ocean Coast Manag 2019;175:90–7. https://doi.org/10.1016/j.ocecoaman.2019.04.005.

[8] Russo T, D'Andrea L, Parisi A, Cataudella S. VMSbase: an R-package for VMS and logbook data management and analysis in fisheries ecology. PLoS ONE 2014;9(6):e100195. https://doi.org/10.1371/journal.pone.0100195.

[9] Armelloni EN, Tassetti AN, Ferrà C, Galdelli A, Scanu M, Mancini A, Fabi G, Scarcella G. AIS data, a mine of information on trawling fleet mobility in the Mediterranean Sea. Mar Policy 2021;129:104571. https://doi.org/10.1016/j.marpol.2021.104571.

[10] Bernabé P, Gotlieb A, Legeard B, Marijan D, Sem-Jacobsen FO, Spieker H. Detecting intentional ais shutdown in open sea maritime surveillance using self-supervised deep learning. IEEE Trans Intell Transp Syst 2023:1–12. https://doi.org/10.1109/TITS.2023.3322690.

[11] De Souza EN, Boerder K, Matwin S, Worm B. Improving fishing pattern detection from satellite AIS using data mining and machine learning. PLOS ONE 2016;11(7):e0158248. https://doi.org/10.1371/journal.pone.0158248.

[12] Sales Henriques N, Russo T, Bentes L, Monteiro P, Parisi A, Magno R, Gonçalves JM. An approach to map and quantify the fishing effort of polyvalent passive gear fishing fleets using geospatial data. ICES J Mar Sci 2023;80(6):1658–69. https://doi.org/10.1093/icesjms/fsad092.

[13] European Commission. (2014). Regulation (EU) No 508/2014 of the European Parliament and of the Council of 15May 2014 on the European Maritime and Fisheries Fund and repealing Council Regulations (EC) No 2328/2003, (EC) No 861/2006, (EC) No 1198/2006 and (EC) No 791/2007 and Regulation (EU) No 1255/2011 of the European Parliament and of the Council. 66.

[14] FAO. The Regional Plan of Action for Small-Scale Fisheries in the Mediterranean and the Black Sea. Rome: Food and Agriculture Organization of the United Nations; 2021.

[15] Bergh Ø, Beck AC, Tassetti AN, Olsen E, Thangstad TH, Gonzalez-Mirelis G, Søvik G. Analysis of spatial conflicts of large scale salmonid aquaculture with coastal fisheries and other interests in a Norwegian fjord environment, using the novel GIS-tool SEAGRID and stakeholder surveys. Aquaculture 2023;574:739643. https://doi.org/10.1016/j.aquaculture.2023.739643.

[16] Farella G, Tassetti AN, Menegon S, Bocci M, Ferrà C, Grati F, Barbanti A. Ecosystem-based MSP for enhanced fisheries sustainability: an example from the northern Adriatic (Chioggia—Venice and Rovigo, Italy). Sustainability 2021;13(3):1211. https://doi.org/10.3390/su13031211.

[17] Lattanzi P, Pulcinella J, Battaglia P, Di Cintio A, Ferrà C, Di Franco A, Tassetti AN. Bridging the gap in fishing effort mapping: a spatially-explicit fisheries dataset for campanian MPAs. Italy Sci Data 2024;11(1):54. https://doi.org/10.1038/s41597-023-02883-9.

[18] Stelzenmüller V, Lee J, South A, Foden J, Rogers SI. Practical tools to support marine spatial planning: a review and some prototype tools. Mar Policy 2013;38:214–27. https://doi.org/10.1016/j.marpol.2012.05.038.

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

[20] Rufino MM, Mendo T, Samarão J, Gaspar MB. 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 2023;154:110628. https://doi.org/10.1016/j.ecolind.2023.110628.

[21] Samarão J, Moreno A, Gaspar MB, Rufino MM. Improving machine learning predictions to estimate fishing effort using vessel's tracking data. Ecol Inform 2024:102953. https://doi.org/10.1016/j.ecoinf.2024.102953.

[22] Rufino MM, Mujal-Colilles A, Russo T, Stelzenmüller V, Gaspar M, Rodríguez-Gutiérrez J, Mendo T. Implications of low temporal resolution of vessel tracking systems. ICES J Mar Sci 2025;82(8):fsaf123. https://doi.org/10.1093/icesjms/fsaf123.

[23] Mendo T, Glemarec G, Mendo J, Hjorleifsson E, Smout S, Northridge S, James M. Estimating fishing effort from highly resolved geospatial data: Focusing on passive gears. Ecol Indic 2023;154:110822. https://doi.org/10.1016/j.ecolind.2023.110822.

[24] Rodriguez, J. (2023). Iapesca, a R-package for manipulating and interpretating high resolution geospatial data from fishing vessels.

[25] FAO. Fishing gear classification. https://www.fao.org/cwp-on-fishery-statistics/handbook/capture-fisheries-statistics/fishing-gear-classification/en/; 2022.

[26] Mendo T, Mujal-Colilles A, Stounberg J, Glemarec G, Egekvist J, Mugerza E, James M. A workflow for standardizing the analysis of highly resolved vessel tracking data. ICES J Mar Sci 2024;81(2):390–401. https://doi.org/10.1093/icesjms/fsad209.

[27] Lattanzi P, Vasapollo C, Samarão J, Galdelli A, Mendo T, Rufino M, Bolognini L, Tassetti AN. Data and code for predicting fishing vs. not-fishing in small-scale fisheries: a sample vessel tracking dataset and a reproducible machine learning approach [Computer software]. Zenodo 2025. https://doi.org/10.5281/zenodo.18407451.

[28] 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 É. Scikit-learn: machine learning in python. J Mach Learn Res 2011;12(85):2825–30.

[29] ICES. Workshop on small scale fisheries and geo-spatial data 2 (WKSSFGEO2). ICES Sci Rep 2023;5(49):105. https://doi.org/10.17895/ices.pub.22789475.

[30] Tassetti AN, Galdelli A, Pulcinella J, Mancini A, Bolognini L. Addressing Gaps in Small-Scale Fisheries: A Low-Cost Tracking System. Sensors 2022;22(3):839. https://doi.org/10.3390/s22030839.

[31] European Commision. REGULATION (EU) 2023/2842 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 22 November 2023 amending Council Regulation (EC) No 1224/2009, and amending Council Regulations (EC) No 1967/2006 and (EC) No 1005/2008 and Regulations (EU) 2016/1139, (EU) 2017/2403 and (EU) 2019/473 of the European Parliament and of the Council as regards fisheries control. https://eur-lex.europa.eu/eli/reg/2023/2842/oj; 2023.

[32] ICES. Working group on spatial fisheries data (WGSFD). ICES Sci Rep 2025;7(111):42. https://doi.org/10.17895/ices.pub.30734759.