Fishing Gear Classification from Vessel Trajectories and Velocity Profiles: Database and Benchmark

Pietro Melzi(✉), Juan Manuel Rodriguez-Albala, Aythami Morales, Ruben Tolosana, Julian Fierrez, and Ruben Vera-Rodriguez

BiDA-Lab, Universidad Autonoma de Madrid, Madrid, Spain
pietro.melzi@uam.es, juanmanuel.rodrigueza@alumni.uam.es

Abstract. International Organizations demand to take care of our oceans and their ecosystems since they are of incalculable value to humanity. The illegal fishing activity does irreparable damage to these ecosystems and these organism are pushing to detect and combat illegal fishing activities. Fishing vessels are equipped with a radio frequency beacon that emits their GPS position and other information relevant to the Automatic Identification System (AIS). The GPS positions can be used to infer the vessel trajectories and detect illegal fishing activities. In this study we present a new database (https://github.com/BiDAlab/TrFGdb) including trajectories representing 5 different fishing gears, and analyze them as in a problem of time sequence analysis. We extract global and local features from the trajectories of vessels, and propose several supervised learning algorithms to classify the kinematics of vessels according to different fishing gears. Compared to previous works, we highlight the importance of considering trajectories with sampling period in the order of minutes instead of hours, to detect activities carried out in a short time that could help to distinguish fishing gears. A considerable effort has been dedicated to pre-processing the real data at our disposal, to generate a quality dataset with highly reliable labels. The best classification accuracy obtained in this study is 90%. We expect to improve it if more trajectories describing the different fishing gears were available.

Keywords: Fishing Gear Classification · Database · Illegal Fishing

1 Introduction

According to the Food and Agriculture Organization of the United Nations (FAO), illegal, unreported and unregulated (IUU) fishing is a broad term that encompasses a wide variety of fishing activities. IUU fishing exists in all types and extents of fishing, occurs both on the high seas and in areas under national

P. Melzi and J. M. Rodriguez-Albala—These authors contributed equally to this work.

A. Pertusa et al. (Eds.): IbPRIA 2023, LNCS 14062, pp. 629–638, 2023.
https://doi.org/10.1007/978-3-031-36616-1_50

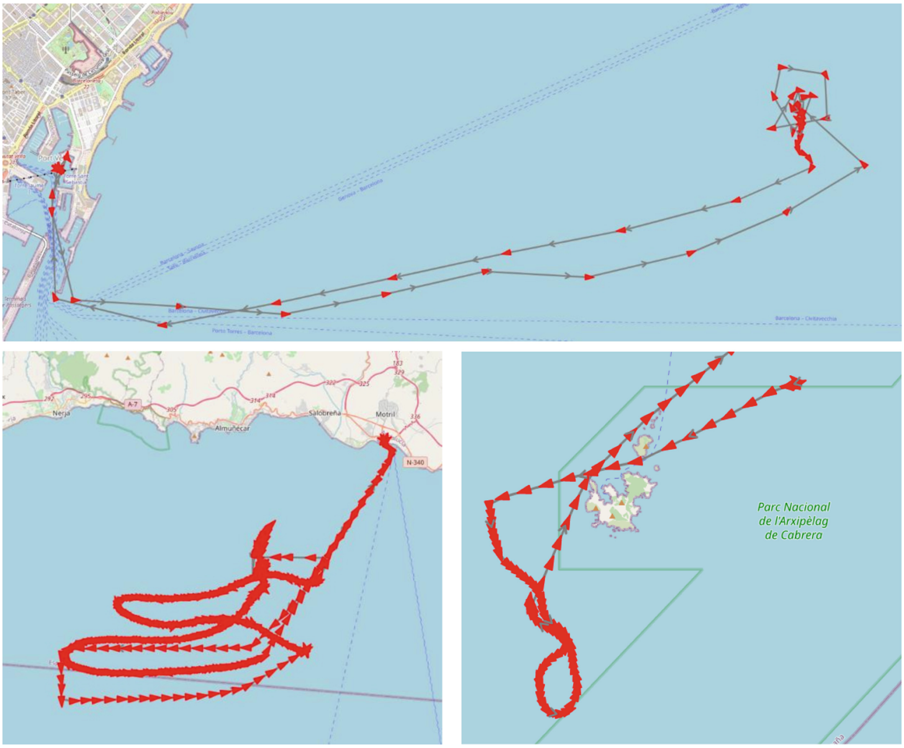

Fig. 1. Example of trajectories for different fishing gears: surroundings (above), long-lines (left), and trawls (right).

jurisdiction, affects all aspects and stages of the capture and use of fish, and can sometimes be associated with the organized crime [6].

IUU fishing represents a great disadvantage and discrimination for fishermen who act responsibly, honestly, and in accordance with the conditions of their fishing authorizations. If IUU fishers operate with vulnerable populations subject to strict moratoriums, healthy levels of those populations cannot be restored, threatening marine biodiversity, the food supply of communities that depend on fishing resources for protein intake, and the subsistence of people related to the sector. According to estimates, IUU fishing involves around 26 million tons of fish per year in the whole world, equivalent to more than 15% of the total annual amount of fish products [5].

With this study, we aim to contribute to the monitoring and surveillance of IUU fishing for social good. We process the records provided by Tragsatec's Management of Agricultural and Fisheries Information Systems, that detail the fishing activities carried out by 357 vessels leaving the ports of Spain. Such records contain information about the geographical position, speed, and direction of vessels over time, together with the description of the fishing gear transported.

The skippers of fishing vessels declare fishing gears at the exit of port, sail in search of schools of fish and begin their fishing operations describing a trajectory with speed usually below 5 knots. As shown in Fig. 1, different fishing gears present trajectories with peculiar characteristics, that we can classify by means of supervised learning algorithms [2,9].

The number of public dataset available to model the trajectories of vessels is limited. In [9] authors presented a database with 1,227 fishing vessels operating in the Indonesian regulatory area, one of the countries with the highest rate of IUU fishing in the world. Each fishing vessel is provided with time series representing its GPS positions over one year, with an approximate resolution of one hour. In such work, features of interest are obtained with unsupervised analysis and combined in classification models, namely Random Forests (RFs) and Support Vector Machines (SVMs). With the sole use of behavioural features, accuracy of 93% is obtained for the following fishing gears: *i)* trawls, *ii)* longlines, *iii)* pole-and-lines, and *iv)* purse seines. The additional use of mean GPS positions increases accuracy to 97%, but this information discriminates region-specific fisheries.

In another work [1], combinations of local and global features are extracted from trajectories of Thai vessels. An accuracy above 90% is obtained for the classification of the following fishing gears: *i)* trawls, *ii)* purse seines, *iii)* longlines, and *iv)* reefer. However, the sampling period of the employed time series is two hours. Such time interval provides insufficient information for purse seine because the speed of fishing vessels suddenly changes, causing lower classification performance for the specific class. The study suggests to reduce the sampling period to 15–30 min, as for the time series considered in our study. A comparison of our database with the ones considered in [9] and [1] is provided in Table 1.

The contributions of this study are the following:

- A new database containing more than one thousand trajectories recorded from 357 fishing vessels with a sampling period of 5 min, to overcome some limitations of previous study based on hourly sampling periods. This database reduce by more than 10 the Nyquist bandlimit of existing databases.
- A novel method based on the fusion of global and local features to classify the trajectories of vessels according to their fishing gear with high reliability.

The remaining of the paper is organized as follows: Sect. 2 describes the database and methods proposed, Sect. 3 presents the results obtained, and Sect. 4 draws the conclusions of this study.

2 Database and Methods

2.1 Database

The data provided by Tragsatec's Management of Agricultural and Fisheries Information Systems, with the authorization of the General Secretariat of Fisheries of the Spanish Ministry of Agriculture, Fisheries and Food, are not originally captured and processed for the purpose of this study. As a consequence,

Table 1. Comparison of databases describing the trajectories of fishing vessels equipped with different fishing gears.

	Our	[9]	[1]
Fishing vessels	357	1,227	32
VMS positions	184,489	5,263,158	184,577
Trajectories	1,045	–	771
Trajectories/classes	209	–	–
Observed days	66	365	–
Sampling period (minutes)	**5 ± 0.83**	**60 ± 15**	**120**
Nyquist Bandlimit (Hz)	**1/600**	**1/7200**	**1/14400**
Classes (Fishing Gear)	Trawls, Purse seines, Trammels, Longlines, Gillnets	Trawls, Longlines, Pole-and-liners, Purse seines	Trawls, Purse seines, Longlines, Reefer

detailed operations of cleaning and data preparation were required. We describe the data curation task that generates a quality dataset with highly reliable labels [10]. The original raw data contain the information described in Table 2. In addition, we consider the expert knowledge provided by Tragsatec's Management and the General Secretariat of Fisheries, about data format and properties of the different fishing gears.

Given the high detail of "Fishing gears" table, we group fishing gears according to the Annex III of Regulation (EU) n° 1379/2013 [3]. The resulting classes of fishing gear that we consider in our study are[1]:

- Trawls: is a fishing practice that herds and captures the target species by towing a net along the ocean floor.
- Purse seines, or surrounding: is a large wall of netting deployed around an entire area or school of fish.
- Trammels: are similar to a gill net but are made up of three layers of netting.
- Longlines: consist of a mainline, gangions, and baited hooks.
- Gillnets: is a wall of netting hanging in the water column, typically made of monofilament or multifilament nylon.

The Tragsatec database presents a Nyquist bandlimit $B = 1/600$. The Nyquist theorem stablish that *"If a function $x(t)$ contains no frequencies higher than B hertz, then it can be completely determined from its ordinates at a sequence of points spaced less than $1/(2B)$ seconds apart."* Thus, the Tragsatec database outperform by 12 and 24 the bandlimit of existing databases [1,9] (see Table 2). This bandlimit is critical when implementing frequency analysis used in time-based feature extraction methods (e.g., Recurrent Neural Networks, Hidden Markov Models, etc.).

[1] https://www.seafish.org/.

Table 2. Description of the data provided by Tragsatec's Management of Agricultural and Fisheries Information Systems and used in this study. "Samples" column refers to the period between 2021-12-15 and 2022-02-19.

Table	Description	Samples	Fields
AIS messages	Messages issued by vessel's AIS beacon	5,007,208	Geom. position, date, hour, speed, course, vessel id
Vessels	Basic data of a vessel	1,647	Vessel id, usual fishing gears
Diary statements	Declarations of when and where vessels start and end their navigation	31,794	Diary id, vessel id, departure date, return date, departure port id, return port id
Fishing gears carried	Declarations of fishing gears carried on board by vessels before going to sea	33,129	Record id, diary id, fishing gear id
Fishing gears	Information about fishing gears	157	Fishing gear id, name, details, code
Ports	Information about ports	31,794	Port id, geometric outline of the port, name

2.2 Data Curation

We rule out diary statements with more than one fishing gear, as we do not know which gear is used at what time. We identify vessel's departures and returns to port by combining two consecutive vessel's GPS positions with the port outline. Due to the variability of the AIS beacon, in some trajectories there is no intersection between vessel's positions and port outline. It may be confused with loss of coverage. Hence, we decided to consider only the trajectories that intersect the outline of a port at their beginning and end, given that the correct use of the AIS beacon provides more reliability. We use the starting and ending time and place of a trajectory, determined from the vessel's GPS positions, to obtain the fishing gear declared in the diary statement.

The messages issued by the AIS beacon do not always have a fixed period of 300 s. We fix the threshold to 350 s, to cover 95.45% (2σ) of AIS messages and detect outliers, according to the empirical rule of 68-95-99.7 (three-sigmas rule) [7]. This threshold represents the maximum time that can elapse between two consecutive messages, which guarantees continuous sampling of GPS positions without loss of coverage and outliers. On the other hand, we significantly reduce the number of diary statements because trajectories with at least a message exceeding the threshold are completely discarded (19,633 of 31,794 diary statements). Additionally, we discard trajectories with a low percentage of AIS messages at fishing speed (lower than 5 kn), or with a total duration minor than 180 min. Such trajectories represent other activities, for instance docking at intermediate ports. The number of valid diary statements decreases to 9,399. Finally, we select 209 diary statements and their trajectories (undersampling) to equally represent the five fishing gears described above.

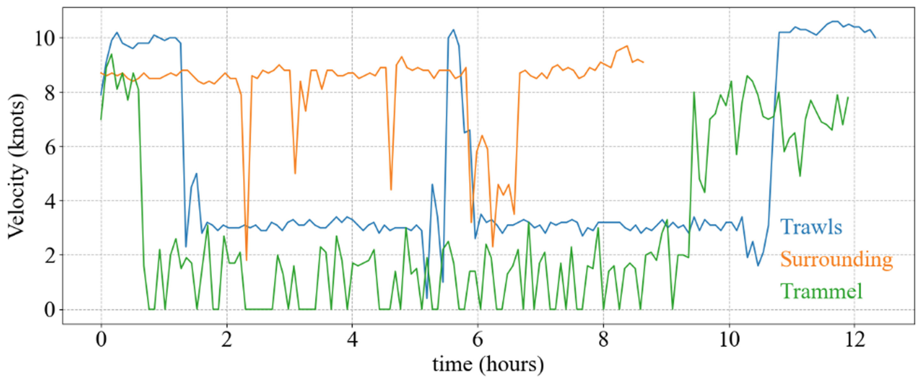

Fig. 2. Velocity profiles of three fishing gears: trawls (blue), surrounding (orange), and trammel (green). (Color figure online)

2.3 Feature Extraction

The course and speed of the vessel are affected by the fishing gear (see Fig. 2). The trajectory over time t of a vessel is described as time sequences of geographical coordinates, longitude(t) and latitude(t). They are analogous to the $x(t)$ and $y(t)$ coordinates of a trajectory over two-dimensional space (x, y) over time t. The literature on modeling trajectories using machine learning approaches is very broad. Among the different applications of these methods, the biometric recognition of dynamic signatures is interesting for the present work because of the high intra-class variability of signers and the low inter-class variability of forgeries. Based on this, we adapt state-of-the-art techniques for dynamic handwritten signature recognition to the kinematics of vessels. Moreover, portions of trajectories representing fishing activities, usually with a speed lower than $5kn$, provide an analogy with the contact of digital pens with electronic tablets. Hence, we establish a relationship of inverse proportionality between the fishing speed and the pressure of the digital pen $p(t)$.

Global Features. A trajectory can be described by an n-dimensional vector, containing features related to its shape and temporal events. In [8], a set of 100 global features is considered, many of them providing high performance in the literature for the task of online signature trajectory recognition. Global features are extracted from discrete time signals of digital pen trajectories: $x(t)$, $y(t)$, and $p(t)$, with $p > 0$ to indicate digital pen down and $p = 0$ to indicate digital pen up. Global features are normalized in $[0, 1]$ with hyperbolic tangent. They can be divided into four categories:

- Time: 25 features related to the duration of the trajectory, events such as raising the digital pen, or local maximums.
- Velocity and acceleration: 25 features obtained from the first and second order temporal derivatives of position-temporal functions.

– Direction: 18 features extracted from the trajectory, for instance the starting and average direction.
– Geometry: 32 features associated with the line or aspect of the dynamic trajectory.

In this study we adapted the extraction of global features from [8] to fishing vessel's trajectories. We considered the following input parameters: $x(t)$ and $y(t)$ representing the GPS vessel's position converted to nautical miles, $p(t)$ analogous to the stylus pressure, with $p > 0$ to indicate vessel at fishing speed, and $p = 0$ to indicate vessel at navigation speed, $TSAMPLE$, *i.e.* the average sampling period, and a time vector indicating the real-time instant of each point, since it is not equally distributed as assumed in the original function. The output was the *100*-dimension vector proposed in [8], with no elimination of any global feature.

Local Features. We adapted the set of local features proposed in [8]. This set of features was an extension of the set described in [4], comprising seven discrete time functions extracted from the trajectory and pressure of the digital pen, their first and second order derivatives, and other for a total of 27 features. A detailed description of the global and local features extracted can be found in [8].

3 Experiments and Results

We evaluated the performance of several supervised learning classifiers according to the mean accuracy obtained with *10*-fold stratified cross-validation (CV). We split our database into training and test sets with a 70:30 ratio. We considered three classifiers for global features, and set optimal parameters with CV: *i)* Support Vector Machine (SVM) with Gaussian kernel, as data are not linearly separable, $C = 100$ and $\gamma = 0.01$, *ii)* Random Forest (RF) with 101 estimators, and *iii)* Multilayer Perceptron (MLP) with hidden layer of size $10,000$, Rectified Linear Unit (ReLU) activation function, max iterations $= 1,000$, and $\alpha = 0.0001$.

For classification with local features, we used a Bidirectional Gated Recurrent Unit (BiGRU) with the following layers: *i)* input layer, *ii)* masking layer, to ignore trajectory positions without information, *iii)* bidirectional layer surrounding a GRU layer, and *iv)* fully connected layer with 5 outputs and softmax activation function. Finally, the possibility of RF and BiGRU fusion at score level was also investigated. It consists in a weighted sum of the scores provided by RF (s_{RF}) and BiGRU (s_{BiGRU}) classifiers ($s_{fusion} = 0.2s_{BiGRU} + 0.8s_{RF}$).

The mean accuracy obtained with *10*-fold CV is reported in Table 3 for the different classifiers. The best accuracy provided by a single classifier was 86.22%, obtained with RF with the Global feature set. The MLP and SVM classifiers showed lower performances with 82.69% and 83.16% respectively. The BiGRU classifier provided 75.6% of accuracy using the Local feature set. However, the best performance is obtained when combining the Global and Local feature set scores. We considered at the same time global and local features by fusing

Table 3. Classification accuracy of the different approaches.

Classifier	Features	Mean Acc. [%]
BiGRU	Local	75.60
MLP	Global	82.69
SVC	Global	83.16
RF	Global	86.22
RF + BiGRU	Fusion	90.13

the scores provided by RF and BiGRU, and obtained an increase of accuracy above 90%. This is a relative error reduction of 28%. We consider it a promising result and expect to increase it with improvements in feature selection and data availability.

In Fig. 3 we provide the confusion matrices obtained for the five fishing gears with the following classifiers: *i)* RF (left), *ii)* BiGRU (center), and *iii)* fusion of

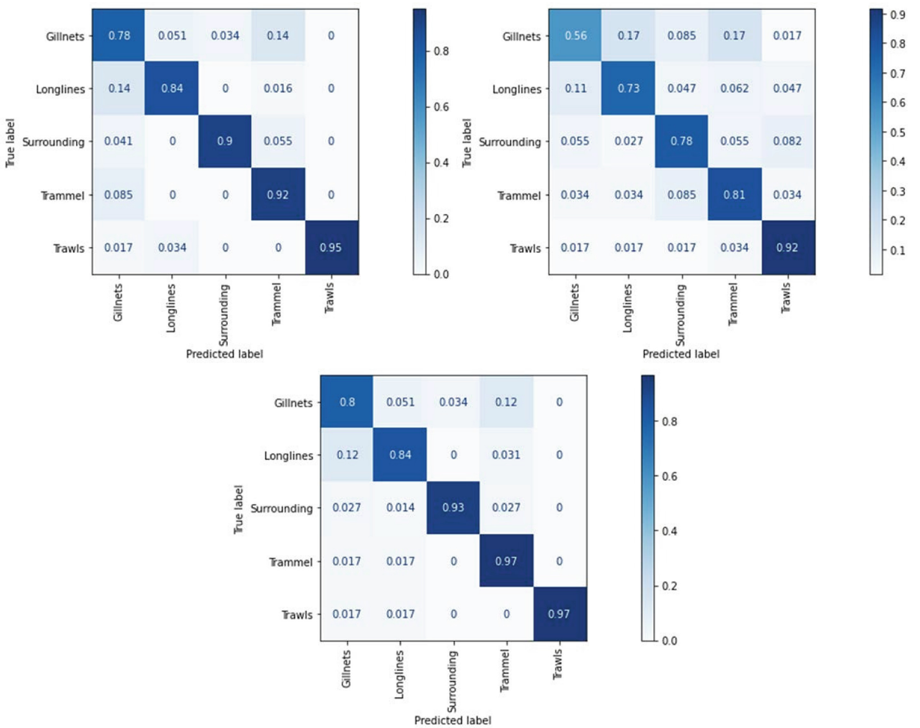

Fig. 3. Confusion matrices obtained for RF based on global features (left), BiGRU based on local features (center), and fusion of RF and BiGRU based on both features (right).

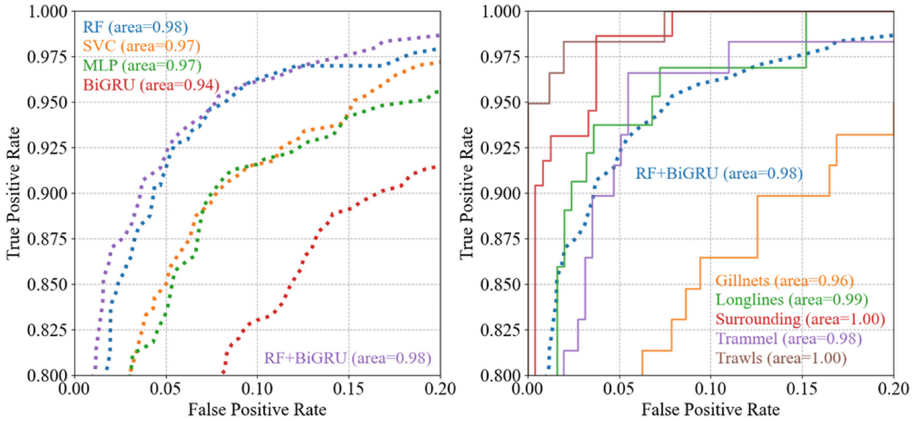

Fig. 4. ROC curves of different classifiers (left) and different fishing gears in the case of fused scheme with RF + BiGRU (right).

RF and BiGRU at score level (right). We observe in Fig. 3 (right) that we obtain accuracy of 93% for the "surrounding" class, greater than the total accuracy of the classifier (90.13%). This is not the case of [1], where time series with two-hours sampling period have been employed. Such sampling interval is too big to classify the "surrounding" gear ("purse seine" in [1]) with an accuracy similar to the other classes.

Finally, in Fig. 4 we report the Receiver operating characteristic (ROC) curves of different classifiers (left) and different fishing gears when RF + BiGRU is considered (right). Trammels consist in a variant of gillnets. This may explain the lower area under the curve obtained for the "gillnets" gear and the error of 12% between "gillnets" and "trammels" provided by the confusion matrix in Fig. 3 (right).

4 Conclusions

This work present a new database to train and evaluate fishing vessel classification from GPS trayectories. The database comprises more than one thousand trajectories recorded from 357 fishing vessels and 5 different fishing gears. This database reduce by more than 10 the Nyquist bandlimit of existing databases and represent a new resource to combat illegal fishing activities.

In this study we carried out the analysis of fishing vessel's dynamic trajectories to classify them according to fishing gear classes. The kinematics of vessels have been modeled according to the global and local features extracted from their trajectory, exploiting the analogy with the problem of dynamic handwritten signature recognition. The results obtained in terms of accuracy with multiple supervised learning classifiers confirm our choice to adapt the feature extraction process proposed in [8] to the analysis and classification of fishing vessel's trajectories.

Acknowledgments. We thank Tragsatec's Management of Agricultural and Fisheries Information Systems and the General Secretariat of Fisheries of the Spanish Ministry of Agriculture, Fisheries and Food for the data and expertise provided to carry out the study. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860813 - TReSPAsS-ETN. This study is also supported by the project INTER-ACTION (PID2021- 126521OB-I00 MICINN/FEDER).

References

1. Chuaysi, B., Kiattisin, S.: Fishing vessels behavior identification for combating IUU fishing: enable traceability at sea. Wirel. Pers. Commun. **115**(4), 2971–2993 (2020)
2. Dunn, D.C., et al.: Empowering high seas governance with satellite vessel tracking data. Fish Fish. **19**(4), 729–739 (2018)
3. European Union: Regulation (EU) no 1379/2013 of the European parliament and of the council of 11 December 2013 on the common organisation of the markets in fishery and aquaculture products, amending council regulations (EC) no 1184/2006 and (EC) no 1224/2009 and repealing council regulation (EC) no 104/2000 (2013). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32013R1379. Accessed 21 Nov 2022
4. Fierrez, J., Ortega-Garcia, J., Ramos, D., Gonzalez-Rodriguez, J.: Hmm-based on-line signature verification: feature extraction and signature modeling. Pattern Recogn. Lett. **28**(16), 2325–2334 (2007)
5. Food and Agriculture Organization of the United Nations: What is IUU fishing? (2022). https://www.fao.org/iuu-fishing/background/what-is-iuu-fishing/en/. Accessed 17 Nov 2022
6. Jarvis, R.M., Young, T.: Pressing questions for science, policy, and governance in the high seas. Environ. Sci. Policy **139**, 177–184 (2023)
7. Leonard, K.: Schaum's Outline of Business Statistics, 4th edn. (2003)
8. Martinez-Diaz, M., Fierrez, J., Krish, R.P., Galbally, J.: Mobile signature verification: feature robustness and performance comparison. IET Biometrics **3**(4), 267–277 (2014)
9. Marzuki, M.I., Gaspar, P., Garello, R., Kerbaol, V., Fablet, R.: Fishing gear identification from vessel-monitoring-system-based fishing vessel trajectories. IEEE J. Oceanic Eng. **43**(3), 689–699 (2017)
10. Miller, R.J., et al.: Big data curation. In: COMAD, p. 4 (2014)