



# Spatio-temporal trajectory data modeling for fishing gear classification

Juan Manuel Rodriguez-Albala<sup>1</sup> · Alejandro Peña<sup>1</sup> · Pietro Melzi<sup>1</sup> · Aythami Morales<sup>1</sup> · Ruben Tolosana<sup>1</sup> · Julian Fierrez<sup>1</sup> · Ruben Vera-Rodriguez<sup>1</sup> · Javier Ortega-Garcia<sup>1</sup>

Received: 22 November 2023 / Accepted: 27 February 2024  
© The Author(s) 2024

## Abstract

International Organizations urge the protection of our oceans and their ecosystems due to their immeasurable importance to humankind. Since illegal fishing activities, commonly known as IUU fishing, cause irreparable damage to these ecosystems, concerned organisms are pushing to detect and combat IUU fishing practices. The automatic identification system allows to locate the position and trajectory of fishing vessels. In this study we address the task of detecting vessels' fishing gears based on the trajectory behavior defined by GPS position data, a useful task to prevent the proliferation of IUU fishing practices. We present a new database including trajectories that span 7 different fishing gears and analyze these as in a time sequence analysis problem. We leverage from feature extraction techniques from the online signature verification domain to model vessel trajectories, and extract relevant information in the form of both local and global feature sets. We show how, based on these sets of features, the kinematics of vessels according to different fishing gears can be effectively classified using common supervised learning algorithms with accuracies up to 90%. Furthermore, motivated by the concerns raised by several organizations on the adverse impact of bottom trawling on marine biodiversity, we present a binary classification experiment in which we were able to distinguish this kind of fishing gear with an accuracy of 99%. We also illustrate in an ablation study the relevance of factors such as data availability and the sampling period to perform fishing gear classification. Compared to existing works, we highlight these factors, especially the importance of using sampling periods in the order of minutes instead of hours.

**Keywords** Fishing gear classification · Database · Illegal fishing · Spatio-temporal trajectory data modeling · AI for social good · Biodiversity · Climate change

---

J. M. Rodriguez-Albala, A. Peña have contributed equally to this work.

---

✉ Juan Manuel Rodriguez-Albala  
juanmanuel.rodriguez@alumni.uam.es

✉ Alejandro Peña  
alejandro.penna@uam.es

Pietro Melzi  
pietro.melzi@uam.es

Aythami Morales  
aythami.morales@uam.es

Ruben Tolosana  
ruben.tolosana@uam.es

Julian Fierrez  
julian.fierrez@uam.es

Ruben Vera-Rodriguez  
ruben.vera@uam.es

Javier Ortega-Garcia  
javier.ortega@uam.es

<sup>1</sup> BiDA-Lab, Universidad Autónoma de Madrid,  
28049 Madrid, Spain

## 1 Introduction

The Food and Agriculture Organization (FAO) of the United Nations,<sup>1</sup> Illegal, Unreported, and Unregulated (IUU) fishing is defined as a “*broad term that encompasses a wide variety of fishing activities*” that violate applicable laws and regulations, either nationally or internationally. IUU fishing practices can be found in all types and extents of fishing, and can sometimes be associated with the organized crime [1]. Hence, activities considered as IUU fishing poses several threats, including environmental, social, and economical challenges. From an environmental perspective, IUU fishing contributes to over-fishing, and may operate with vulnerable populations, ultimately disrupting marine biodiversity and undermining the efforts to accomplish long-term sustainability goals. Furthermore, IUU fishing threatens not only the subsistence of the sector, but the fish stocks and the whole food supply. According to some estimates from FAO,<sup>23</sup> IUU fishing involves around 11–26 million tonnes of fish per year in the whole world, which is equivalent to more than 15% of the total annual number of fish products [2]. In the US, some studies have suggested that the percentage of illegal seafood imports could be as high as 32% [3].

Marine sediments are the largest pool of organic carbon on the planet, and a crucial reservoir for long-term storage. However, disturbance of these carbon stores by bottom trawling can re-mineralize sedimentary carbon to CO<sub>2</sub>, which is likely to increase ocean acidification, reduce the buffering capacity of the ocean, and potentially add to the build-up of atmospheric CO<sub>2</sub>. Thus, protecting the carbon-rich seabed is a potentially important nature-based solution to climate change [4]. Owing to these ecological concerns, and in line with the Regulation 2016/2336<sup>4</sup> on deep-sea fisheries (i.e., the Deep-sea Access Regulation), the European Commission adopted an implementing act on September 2022<sup>5</sup> closing 87 sensitive zones to all bottom fishing gear in the EU waters of the North-East Atlantic. The Deep-sea Access Regulation already banned bottom trawling below 800 meters in 2016 and, with the new act, the Commission implemented Article 9 of that regulation to protect Vulnerable Marine Ecosystems (VMEs) at depths of between 800 and 400 metres [5]. However, given the previous exposition on IUU fishing, one can understand how these practices suppose a challenge to accomplish the goals pursued by these regulations.

Given the critical consequences of IUU fishing, and the lack of resources to reliably identify such activities, some

works have started to develop AI-based systems to detect illegal fishing practices [6–10]. These works mainly focus on detecting IUU fishing through the identification of fishing vessels based on tracking data. Concretely, trajectory data obtained from GPS systems such as Vessel Monitoring Systems (VMS) or Automatic Identification Systems (AIS) are leveraged for this task. These methods are based on the idea that spatio-temporal sequences extracted from vessel behavior (e.g., positional data, velocity profiles) have specific patterns that depend on the fishing gear, and therefore, they can be classified using traditional supervised learning approaches. As shown in Fig. 1, different fishing gears exhibit trajectories and velocity profiles with peculiar characteristics, which further corroborates the fundamental hypothesis of such studies. The application of AI to detect and prevent IUU fishing can be framed in what is known as AI for Social Good (AI4SG) [11–14], that is, the use of AI technology to address social challenges and provide solutions to improve the well-being of communities. AI4SG involves several areas where traditional approaches have been less efficient or even unfeasible, such as tracking diseases [15–18], monitoring environmental risks and disasters [19–22], or social problems mitigation [23–29]. In this work, we address the problem of classifying fishing gears based on vessel trajectory data, with the purpose of monitoring activities that may suppose IUU fishing practices. As we have seen, preventing such activities is not only a matter of complying with the law, but also of achieving the goals of conserving marine biodiversity and combating climate change. We had processed the records provided by the Management of Agricultural and Fisheries Information Systems (MAFIS) of Tragsatec, a Spanish Government-associated company, on the fishing activities carried out by 828 fishing vessels leaving the ports of Spain. Such records contained information about GPS position, speed, and direction of the vessels over time, along with detailed description of the fishing gears transported. We processed this dataset to obtain a clean corpus for fishing gear classification into 7 different classes. We use the resulting database to extract both local and global features from the data, and explore their use for fishing gear classification using different classification methods under different scenarios. Our feature extraction approach abstracts the concept of vessel trajectory as an ensemble of positional and speed signals over time, and establish a parallelism between these and those obtained from online signatures to leverage from the literature on online signature modeling and verification [30, 31]. Our experiments assess the usefulness of the proposed features to identify fishing gears from vessel trajectory data.

The main contributions of this study are the following:

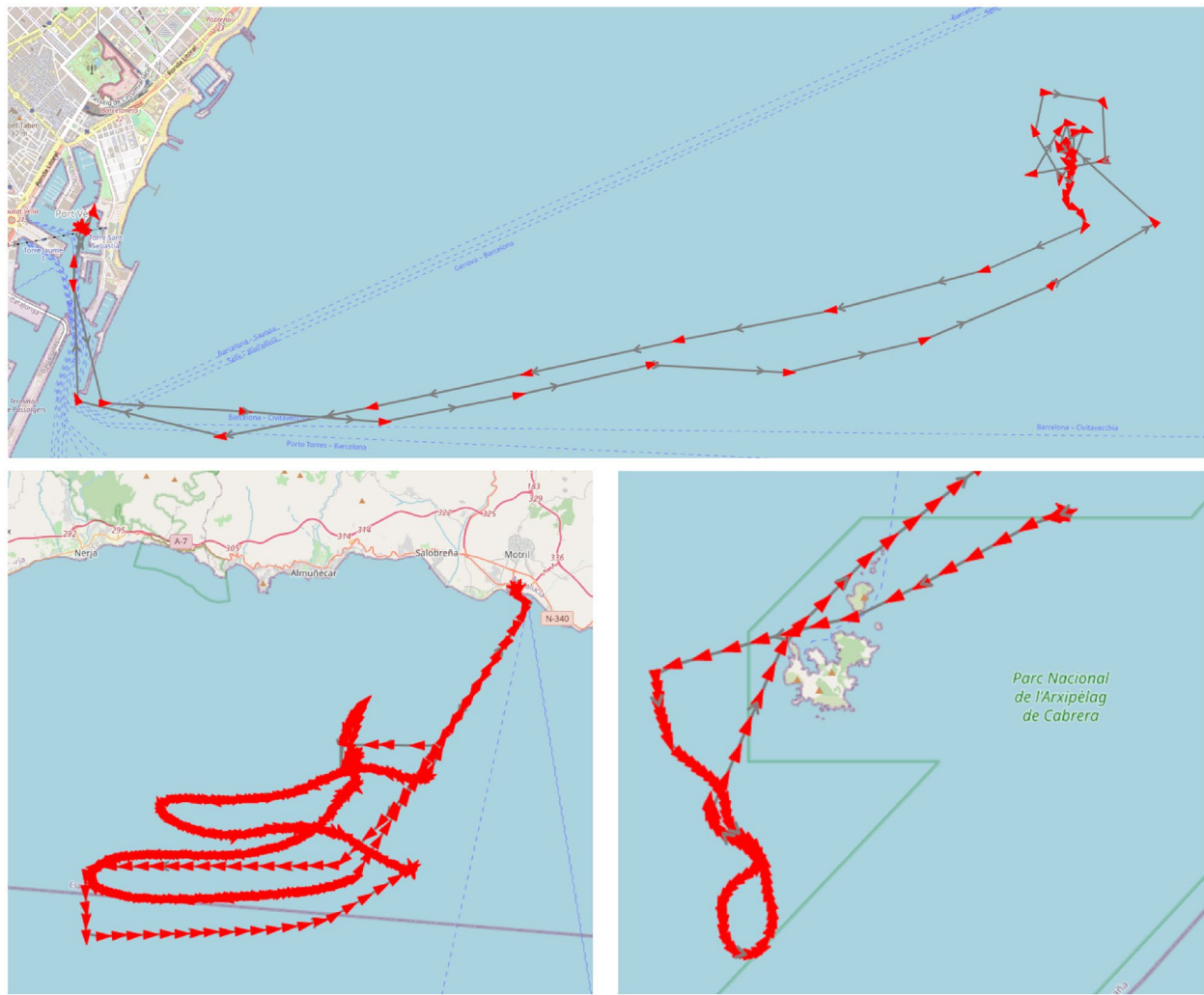
<sup>1</sup> [fao.org/home/en](http://fao.org/home/en).

<sup>2</sup> [fao.org/iuu-fishing/en/](http://fao.org/iuu-fishing/en/)

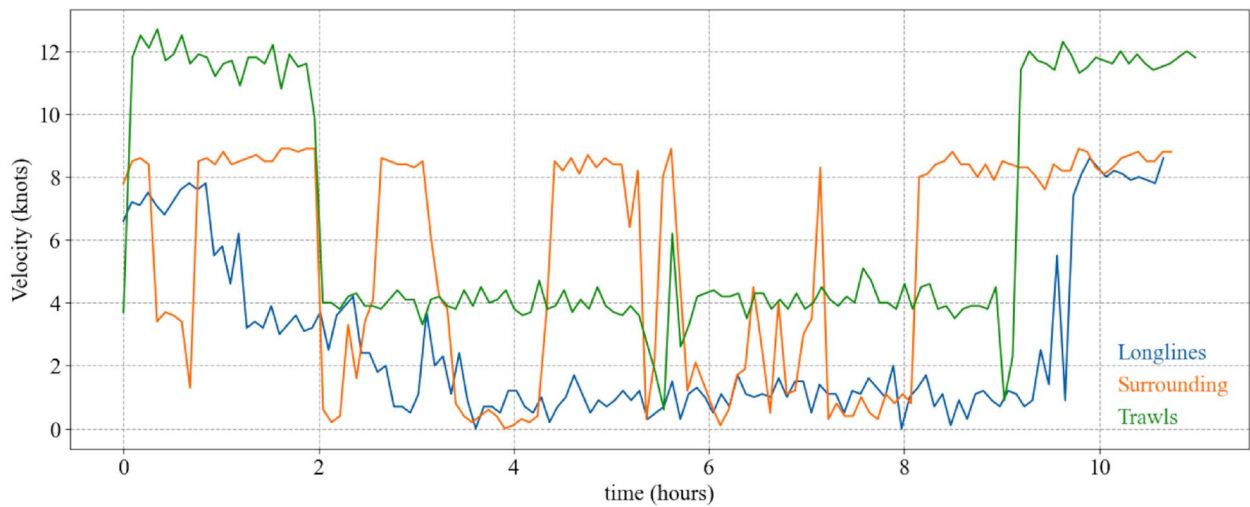
<sup>3</sup> [shorturl.at/emLX5](http://shorturl.at/emLX5).

<sup>4</sup> [eur-lex.europa.eu/eli/reg/2016/2336/oj](http://eur-lex.europa.eu/eli/reg/2016/2336/oj).

<sup>5</sup> [eur-lex.europa.eu/eli/reg\\_impl/2022/1614/oj](http://eur-lex.europa.eu/eli/reg_impl/2022/1614/oj).



(a)



(b)

**Fig. 1** Spatio-temporal kinematics of vessel trajectories: **a** Visual examples of the trajectories for different fishing gears, namely Purse seine fishing (above), longline fishing (bottom-left), and trawling (bottom-right); **b** Velocity profiles for the previous fishing gears

- A new database<sup>6</sup> containing more than nine thousand trajectories recorded from 828 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 band-limit of existing databases.
- We present comprehensive experiments including spatio-temporal features. These features were extracted using two different approaches: one based on local analysis and the other on global analysis of the trajectories.
- 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.
- We present a model specifically trained to detect Trawl Fishing Gear, achieving classification accuracies of over 99%.

A preliminary version of this paper was published in [32]. This article significantly improves [32] in the following aspects:

- We extend the Tragsatec database by increasing the number of fishing vessels. Whilst 357 different fishing vessels were included in [32], here we present information on 828, an increase of almost 2.5 times the previous database.. Furthermore, compared to the 5 fishing gear classes of [32], the database presented here spans 7 different classes. Nevertheless, only 5 classes are included in the multi-classification experimental section due to the limited number of samples for the two new classes.
- We expand our experiments (see Sect. 4.3) by training and evaluating specific trawl detection models, an useful application to prevent IUU fishing. We also provide and ablation study to understand how factors such as data availability or sampling period influence the performance.

The remaining of the paper is organized as follows: Sect. 2 reviews several works related with our work here. Section 3 describes the proposed database, along with the features extracted, and the methodology. Then, Sect. 4 presents the experiments carried out in this work on fishing gear classification, and analyzes the results obtained. Finally, Sect. 5 summarizes the main conclusions of this study.

## 2 Related works

The use of satellite-based data to provide automatic tools for the management, control, and surveillance (MCS) of vessels has increased in recent years. Concretely, two different

systems have proven to be extremely useful to extract rich information about vessel activity, including the detection of illegal fishing. The Vessel Monitoring System (VMS) is a proprietary system integrated with a vessel's GPS. This system transmits detailed, coded information about the vessel to the Regional Fishing Management Organization (i.e., the fishing regulatory authority operating in the seas where the vessel is operating) with high spatial resolution. While it was originally designed to transmit messages with low frequency (e.g., every 2 hours), it has evolved to higher sampling frequencies that can even reach 30 to 15 min. On the other hand, the Automatic Identification System (AIS) is an ITU-standardized system [33] that is also linked to a vessel's GPS, and transmits information such as the identity, the current position, or the course in a broadcast fashion to anyone with a VHF receiver. This means that AIS beacons, as opposed to VMS encoded signals, can be received by other ships. In addition, the AIS system has a significantly higher temporal resolution, with signals transmitted down to a few seconds. While VMS systems were originally designed as a fishing surveillance tool, AIS was intended to prevent collisions and increase safety at sea. For more than 15 years now, the International Maritime Organization (IMO) requires that any vessel with a load higher than 300 tons traveling in international waters, all passenger ships, or cargo ships with a load higher than 500 tons operating in national waters to have an AIS system integrated and turned on [34]. Furthermore, in the EU the AIS system is mandatory for any fishing vessel with an overall length greater than 15 meters from 2014, as noted in the EU Directive 2011/15/EU.<sup>7</sup>

Thus, several studies based on any of these systems have been developed in recent years. In this sense, Dunn et al. argued on the potential that systems based on VMS and AIS have to increase the coverage of vessel management programs, including the visualization of gaps in sea governance, the understanding of fishing activities, or vessel tracking and management [8]. In order to illustrate some of these points, they provide examples using the method proposed in [35]. Concretely, in the latter work Kroodsmat et al. introduced two CNN systems, one of them to detect vessel characteristics (e.g., vessel length or engine power), and the other to identify potential fishing activity positions [35]. The first one was trained with 45 K trajectory data points and achieved an accuracy of 95% in fishing/non-fishing classification. The latter was trained with data from 503 vessels, and obtained an accuracy of 90% in fishing activity detection. They analyzed more than 22 billion AIS messages from more than 70 K industrial fishing vessels, resulting in a spatio-temporal footprint map of fishing activity, from which they concluded that fishing occurs in more than 55% of oceanic areas. In a closely related work [36], the authors

<sup>6</sup> [github.com/BiDALab/TrFGdb](https://github.com/BiDALab/TrFGdb).

<sup>7</sup> [shorturl.at/pvHKV](https://shorturl.at/pvHKV).



proposed to generate high-resolution fishing activity maps from speed profiles obtained from AIS data. They proposed a case study using data on 156 vessels from the Swedish fleet, fitting a bimodal distribution of speed histograms for each vessel with a Gaussian Mixture Model (GMM) with two Gaussians. By fitting these GMMs, they were able to compute the confidence intervals of speed and identify steaming from trawling fishing activity. They validated the proposal on another 112 vessels, and generated fishing effort maps for the case study area. Other works explored as well the use of AIS data to detect fishing activities within a vessel trip [37], or tried to identify behavioral patterns of vessels suspicious of performing dark fishing [10].

A number of works have devised interesting applications for maritime surveillance beyond fishing activity detection. For instance, Nguyen et al. [38] proposed a multi-task framework based on AIS data to simultaneously reconstruct trajectories, detect abnormal behavior, and identify the vessel type (e.g. cargo, passenger vessel). Their framework is built on a Variational Recurrent Neural Network [39], which assumes that the AIS data are noisy, irregular representations of a true, latent data stream. The VRNN allows the model to obtain the regular latent data stream with a sampling period of 10 minutes through an embedding layer, and to detect abnormal behavior by marginalizing the hidden states. Finally, a CNN model is employed to identify the type of vessel. They also introduced a bucketization approach to encode the AIS data as a four-hot coded vector. The approach was tested on data from both the Brittany coast and from the Gulf of Mexico. Huang et al. also focused on vessel type identification, extracting a set of 14-dimensional features reflecting both geometric and trajectory characteristics from nearly 10 K ships operating in the Changhua Wind Farm Channel [40]. They compared the performance of 8 different classifiers, including Random Forests (RF), SVM, or  $k$ -NN, among others, and concluded that good results could be achieved with only 4 of the proposed features. They considered this task to be particularly relevant to hinder illicit practices, since the type of ship can be intentionally manipulated in AIS beacons. Another interesting application is the trajectory prediction, i.e., predicting future trajectories based on past samples in order to prevent potential hazards. Here we can cite the work of [41], where a LSTM-based sequence-to-sequence model with attention was proposed for this task. One key advantage of the model is to extend the input information (i.e., past ship's records), with prior information on the ship's long-term intention (e.g., departure and arrival ports), which was explored in some previous works as a way to improve the performance [42]. They tested the approach on data from the Danish Maritime Authority. Capobianco et al. then

improved their approach to compute a predictive uncertainty confidence [43]. They argued that most of trajectory prediction models do not provide a confidence value to understand how reliable the predicted trajectory actually is. They included the uncertainty prediction in their previous work via Bayesian learning, and tested the approach on the same data from the Danish Maritime Authority.

Regarding the classification of fishing gear based on satellite data (i.e., the objective of this manuscript), this task has been addressed as well using both VMS [6, 44–49] and AIS [7, 9, 50–52] data samples to extract trajectory information. Marzuki et al. [44] proposed to characterize each fishing gear motion by training an independent GMM per gear type, which models both speed and turning angle. They then used the resulting GMMs to extract features from the entire VMS trajectory, and train both RF and SVM models, which combine the GMM-features with position and sinuosity features. They achieved an accuracy of 94.59% in classifying between trawling, purse seine, longline and pole-line in a dataset of more than 3K vessels operating in Indonesian waters in 2012 (i.e., one of the countries with the highest rate of IUU fishing). They extended their previous work in [6], in which the behavioral feature extraction was conducted with a variant of the GMM, namely the Gaussian-Von Mises Mixture Model [53]. They evaluated the same Indonesian VMS dataset using both RF and SVM classifier, increasing the accuracy up to 97.6%. Other proposals is that of Zheng et al. [48], which relies on Neural Network classifiers based on speed profiles to obtain similar accuracies (i.e., 96.6%) on data from China's offshore. On the other hand, among systems using AIS data, we have, for example, the work of [9], which proposes a 3-staged process to classify among 4 different fishing gears. In this framework, trajectories are first reconstructed and divided into segments, which are then fed into a 1D-CNN to perform the final classification. The best performance on Danish data was obtained for the trawl class, with an accuracy of 98.27%. The authors of [7] extracted local and global features from AIS and VMS data of Thai fishing vessels, and classified the fishing gear with a shallow neural network. However, they found that the low sampling period of their data was not enough to obtain sufficient information for certain classes (e.g., purse seine). Xing et al. presented a case study on the East China Sea, combining a grid-based approach with the use of the NLP technique CBOW for feature extraction [52]. The final classification was done with a LightGBM classifier, a variant of the XGBoost classifier. Note that all these proposals agreed on the motivation of their work, that is, the prevention of IUU fishing practices.

**Table 1** Description of the information included in the Tragsatec database

Attributes	Ours	Indonesian VMS [6]	Thai VMS/AIS [7]
Fishing vessels	828	1227	32
GPS positions	1.43 M	5.26 M	184.5 K
Trajectories	9.37 K	–	771
Observed days	66	365	–
Sampling period (min)	$5 \pm 0.83$	$60 \pm 15$	120
Nyquist Band-limit (Hz)	1/600	1/7,200	1/14,400
Classes (Fishing Gears)	Trawls, Purse seines, Trammel, Longline, Gill-nets, Dredges, Pots and traps	Trawls, Longlines, Purse seines, Pole and line	Trawls, Purse Seines, Longlines, Reefer

We included as well information about some similar databases employed in related works for comparison purposes

**Table 2** Description of the raw data provided by Tragsatec’s Management of Agricultural and Fisheries Information Systems used in this study. The data was collected over a capture period between the 15th December 2021 to the 19th February 2022

Info record	Description	Samples	Fields
AIS messages	Messages issued by the vessel’s AIS beacon	5 M	Geom. position, date, hour, speed, course, vessel id
Vessels	Basic data of a vessel	1647	Vessel id, usual fishing gears
Diary statements	Info on when and where the vessels start/end their navigation	31.8 K	Diary id, vessel id, departure date, return date, departure port id, return port id
Fishing gears carried	Fishing gears carried on board	33.1 K	Record id, diary id, fishing gear id
Fishing gears	Information about fishing gears	157	Fishing gear id, name, details, code
Ports	Information about ports	11.6 K	Port id, geometric outline of the port, name

### 3 Database and methodology

#### 3.1 Database

We present in Table 1 the details of the Tragsatec database presented in this work. Note that we included as well the information of similar databases employed in related works for comparison purposes. The Tragsatec database presents a Nyquist band-limit  $B$  of 1/600. The Nyquist theorem establishes 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 significantly outperforms the band-limit of existing databases (i.e., by 12 compared to the database of [6], and by 24 compared to the database employed in [7]). This band-limit is critical when implementing frequency analysis used in time-based feature extraction methods (e.g., RNNs- or HMMs-based techniques), as highlighted in the conclusions of [7]. Furthermore, Tragsatec database comprises 7 different fishing gear classes, whereas the other databases only consider 4 classes.

The raw data collected to create the Tragsatec database were provided by Tragsatec’s Management of Agricultural and Fisheries Information Systems (MAFIS), with the authorization of the General Secretariat of Fisheries of the Spanish MAPA. As a consequence, a detailed data curation procedure was required to obtain the cleaned data corpus presented in Table 1 with highly reliable labels [54], which we will detail in Sect. 3.2. The original raw data from MAFIS contain the information described in Table 2, and were collected over a capture period of about 2 months from December 2021 to February 2022. Raw data was mainly composed of tabular records, including information about the vessels, the fishing gears carried, GPS messages, or the ports of departure and destination. In addition to these records, we considered the expert knowledge provided by MAFIS, about the data format and the properties of the different fishing gears.

Given the high detail of “Fishing gears” available in MAFIS raw data, we decided to aggregate the fishing gears according to the Annex III of Regulation (EU) n° 1379/2013 [55]. The resulting classes of fishing gears that we considered for our study are the following<sup>8</sup>:

<sup>8</sup> seafish.org/

- Trawls: A fishing method that involves dragging a cone-shaped net, usually known as trawl, along the ocean floor to capture the target species.
- Purse seines and surrounding nets: This technique consists in encircling an entire area or school of fish with a surrounding wall of net (i.e. the seine) that hang vertically. Then, the bottom is pulled close to trap the fish inside.
- Gillnets: A fishing method that hangs a wall of net, typically made of nylon, vertically in a water column. Fish swimming into the net are entangled, with a backward structure that prevents their escape.
- Trammel: A variation of the gillnets which employs up to three layers of nets.
- Longline: This technique consists in attaching a long main line with baited hooks behind the boat. The baited hooks are attached at intervals to attract the different species of fish.
- Dredges: This technique involves the use of a rigid structure called dredge to collect shellfish by dragging the dredge along the seafloor.
- Pots and traps: This is a stationary method of capturing sea animals, in which pots and traps are deployed for a period of time (e.g. 24 h) and then hauled aboard to harvest the trapped fish.

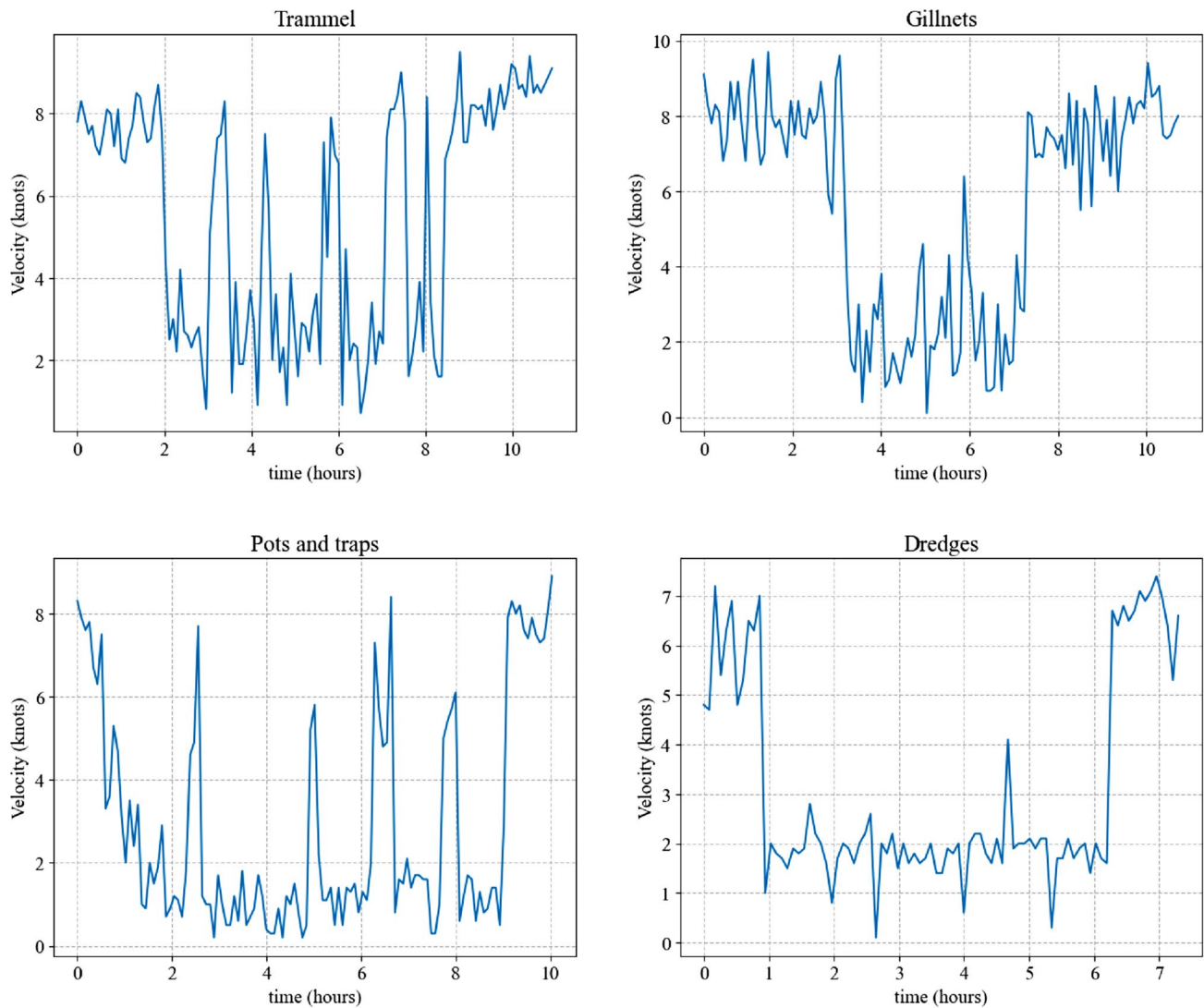
### 3.2 Data curation

In this Section, we describe the data curation process applied to the MAFIS raw data (see Table 2) in order to obtain a clean corpus for fishing gear classification. As already noted by the literature, AIS messages should undergo a data preprocessing process in order to obtain a clean corpus to work, as these data suffers from different quality problems such as gaps in data, duplicated messages, or irregular time sampling [36, 40, 52]. Another commonly applied preprocessing approach is to filter data samples with a speed lower than a certain value, in order to only considered specific parts of the trajectory. Firstly, we filtered diary statements with more than one fishing gear, as we had no method to determine which fishing gear was employed at each time of the navigation. Using the remaining diaries, we identified the vessel's departure and return 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 was no intersection between the vessel's positions and the port outline. Since this may be confused with a loss of coverage, we decided to consider only those trajectories that intersect the outline of a port at both its beginning and its end, given that the correct use of the AIS beacon provides more reliability. We then used the starting/ ending time and the location of a trajectory, determined from the vessel's GPS positions, to obtain the fishing gear reported in the diary statement.

The messages issued by the AIS beacon do not always have a fixed sampling rate of 300 seconds. Consequently, we fixed a threshold of 350 seconds, with which we were able to cover 95.45% ( $2\sigma$ ) of the AIS messages and detect outliers, according to the empirical three-sigma rule of  $68 - 95 - 99.7$  [56]. This threshold represents the maximum time that can elapse between two consecutive messages, which guarantees continuous sampling of GPS positions, hence preventing both loss of coverage and outliers. Conversely, by setting a threshold to obtain a clean, continuous sample, we significantly reduced the number of diary statements, since the trajectories with at least one message exceeding the sampling threshold had to be discarded altogether. This clearly denotes a trade-off between the number of samples and the cleanliness of the data, which in our case (i.e., a threshold of 350 seconds) led to reduction from 31.8 to 19.6 K in the number of diary statements, with nearly a third of the records being filtered. In addition, we discarded trajectories with a low percentage of AIS messages at fishing speed (i.e., a speed lower than 5 knots), or with a total duration of less than 180 minutes. We identified these trajectories with activities other than fishing, such as docking at intermediate ports. The final number of valid diary statements after the whole data curation process is 9376.

### 3.3 Feature extraction

As we previously exposed, both the course and the speed of a vessel present specific behavioral patterns that depend on the fishing gear. We can corroborate this fact on the trajectories illustrated in Fig. 1, or in the velocity profiles depicted in Fig. 2 for the remaining fishing gears analyzed in this work. The evolution of a vessel's trajectory over time  $t$  is described by two time sequences of geographical coordinates, namely the longitude  $long(t)$  and the latitude  $lat(t)$ . These signal are analogous to the positional signal  $x(t)$  and  $y(t)$  describing a trajectory over a 2-dimensional space over time  $t$ . In this sense, the literature on modeling trajectories using machine learning approaches is extensive, and includes a diverse set of applications. Among the different applications of these methods, the work on biometric verification of online signatures [57–59] (i.e., those signatures characterized by chronological sampling of the signature movement) is particularly interesting to model the trajectories of the present study. This is due to the high intra-class variability of signers and the low inter-class variability of forgeries, which requires the extraction of features with significant discriminant power. 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 5 knots, provide an analogy with the contact of digital pens with electronic tablets. Hence, we establish a relationship of



**Fig. 2** Velocity profiles of four fishing gears: trammel (top-left); gillnets (top-right); pots and traps (bottom-left); and dredges (bottom-right)

inverse proportionality between the fishing speed signal  $s(t)$  and the pressure of the digital pen  $p(t)$ .

### 3.3.1 Global features

A trajectory can be described by an  $n$ -dimensional vector, containing features related to its shape and temporal events. The authors of [30] propose to represent a trajectory with a large set of 100 global features. They considered for this representation features that had demonstrated high performance in the literature of online signature verification. Global features are extracted from discrete time signals of digital pen trajectories, namely the positional signals  $x(t)$  and  $y(t)$ , and the pressure signal  $p(t)$ . For the latter, a value of  $p(t_i) > 0$  indicates that the digital pen is down, while  $p(t_i) = 0$  indicates that the digital pen is up

at timestamp  $t_i$ . Each global feature  $f_i$  is normalized using  $\tanh$ -estimators [60] to the interval  $[0, 1]$ . The global features can be grouped into the following four categories:

- **Time:** 25 features related to the duration of the trajectory, events such as raising the digital pen, or local maximums/minimums.
- **Velocity and acceleration:** 25 features obtained from the first and second order temporal derivatives of position-temporal functions, such as the standard deviation of these.
- **Direction:** 18 features extracted from the trajectory, for instance the starting direction, or direction histograms.
- **Geometry:** 32 features associated with the line or aspect of the dynamic trajectory.



In this work, we adapted the extraction of global features proposed by Martinez et al. to our fishing vessel trajectories. We refer the reader to [30] to check the complete list of features. To conduct the extraction, we consider as signals  $x(t)$  and  $y(t)$  the vessel’s GPS position (i.e.  $long(t)$  and  $lat(t)$  respectively) converted to nautical miles. As for the pressure signal  $p(t)$ , we use the fishing speed signal  $s(t)$ , establishing an inverse proportionality analogy between both signals. This means that for a specific timestamp  $t_i$ , a value of  $p(t_i) = 0$  indicates the vessel is at navigation speed (i.e. higher than 5 knots), while  $p(t_i) > 0$  denotes that the vessel is at fishing speed, with high values representing lower speed. Finally, the average sampling period  $T_s$  and the time vector indicating the real-time instant of each data point are considered as well.

### 3.3.2 Local features

Similarly to the case of global features, we adapt here the set of local features proposed in [30] to describe vessel trajectories, using similar correspondences to the ones exposed before. This set of features was an extension of the original set proposed by Fierrez et al [31]. Concretely, based on the signals  $x(t)$ ,  $y(t)$ , and  $p(t)$ , seven discrete functions are defined in [31], for which the first- and second-order derivatives are computed for a total number of 21 signals. From these, all second-order derivatives except those of  $x(t)$  and  $y(t)$  are not considered in [30] due to their low contribution to the verification performance. The resulting set of 16 signals is extended with another 11 functions from the literature, for a final number of 27 local features. A detailed description of these features is provided in [30].

### 3.4 Classification models

In the previous Section, we presented our feature extraction strategy, which results in two different sets of features from the trajectory data, local and global features. To obtain these sets, we draw an analogy between the evolution of trajectories and signatures over time to leverage from the literature on feature extraction for signature verification. Owing to the different nature of local and global features, we use different classification strategies for each of them.

Looking first at the classification with local features, it has been a common practice to process this type of sequential features with recurrent-based classifiers, in order to model their evolution over time. Thus, we decide to use Bidirectional Gated Recurrent Units, i.e., BiGRU-based model, to process local features. Although the GRU unit is less powerful compared to units such as LSTM, its simplicity makes it stronger against overfitting, and can effectively learn long-term dependencies in the data. Due to the limited data of the problem, we believe that the GRU is the perfect

choice to avoid that risk. Furthermore, the use of bidirectional units allows us to process the local features in both time directions. Our model includes a masking layer that prevents trajectory data without information to be considered, followed by a BiGRU layer with 100 units. We take the final state of the BiGRU layer, and use a fully connected layer as the output of the network, with *softmax* activation for the multi-class case or *sigmoid* activation for the binary case to compute the final prediction. This output layer contains the same number of units as the classes considered in the classification problem.

On the other hand, classification on global features can be done using standard Machine Learning classification approaches, since each feature describes an aspect of the data globally, instead of representing its temporal evolution. To this end, we consider three different classification models: (i) Support Vector Machines (SVM) with Gaussian kernel; (ii) Random Forests (RF); and (iii) Multilayer Perceptron (MLP), consisting of a hidden layer with ReLU activation, followed by an output layer with softmax activation (with the same number of output units as classes). Note that these models have been commonly applied by works on the literature (see Sect. 2). Therefore, these classifiers should be powerful enough to obtain competitive results based on a set of discriminant features like the one extracted in this work.

Finally, we will explore a score fusion scheme in our experiments [61], by combining the predictions of both global and local models. The goal of the fusion strategy is to combine in a single prediction the knowledge contained in both approaches, we would capture complementary information of the trajectory data. For this purpose, we compute the fusion score  $s_f$  in the form of:

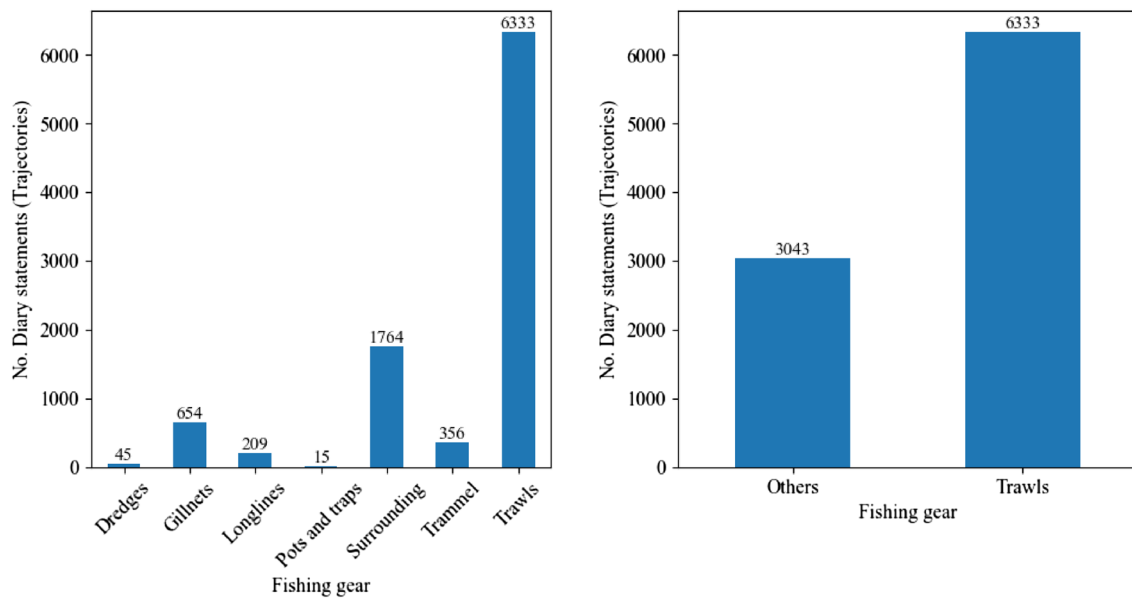
$$s_f = w_g s_g + w_l s_l \tag{1}$$

where  $s_g$  and  $s_l$  are the scores, either Average Precision (AP) or Accuracy, obtained from the global and local models respectively, and  $w_g$  and  $w_l = 1 - w_g$  are the corresponding weights calculated iteratively to provide the best  $s_f$  on a  $K$ -fold cross-validation (CV). Thus, the fusion score is computed as weighted sum of the scores predicted by each classifiers.

## 4 Experiments and results

### 4.1 Experimental protocol

Similar to other existing databases for fishing gear classification, the Tragasec database suffers from class imbalances. As can be seen in Fig. 3 (left), where we depict the distribution of diary statements by fishing gear, the most frequent class in our database is *Trawls* (i.e., 6333 trajectories),



**Fig. 3** Distribution of Diary statements (Trajectories) by fishing gear class: Multi-class (left) and Binary (right)

whose representation is significantly higher than the second one (i.e., *Surrounding*, with 1764 diary statements). Note that for two classes the number of trajectories is less than 50. Actually, we have twice as many samples for the *Trawls* class than for the rest of the classes combined, as illustrated in Fig. 3 (right). This fact seems to imply<sup>9</sup> that trawling is the most widespread fishing practice, despite its potential negative impact on climate change [4].

Given the conditions of our database, two sets of experiments are proposed in this work. In a first part of the experiments, we want to evaluate the proposed methodology for fishing gear classification in a multi-classification configuration using the Tragsatec database (Sect. 4.2). Specifically, we will explore the multi-classification task using a balanced subset of Tragsatec with the 5 most frequent classes. We decided to leave out of this experiment both the *Dredges* and *Pots and traps* classes due to the low number of samples available, which makes unfeasible to obtain sufficient training/test sets to extract significant results. Nevertheless, we have included these classes in the database, as they may be useful for other researchers to evaluate their proposals, or even for future extensions of the database. Thus, in this experiment, we apply an under-sampling procedure to obtain a balanced corpus with 209 samples from the *Trawls*, *Longlines*, *Surrounding*, *Trammel*, and *Gillnets* classes, so that each class is equally represented to prevent potential biases associated with class imbalances.

<sup>9</sup> The data used for the Tragsatec database reflect fishing practices on a short period of time (i.e., 2 months), so further research would be needed to confirm this assertion.

On a second part of the experiments (Sect. 4.3), we consider the proportions illustrated in Fig. 3 (right), and explore the fishing gear classification task as a binary problem, i.e., a One-vs-All configuration in which the aim is to detect *Trawls* from other fishing gears. As we mentioned in Sect. 1, this fishing gear is of particular interest for its impact in biodiversity, therefore international regulation point special emphasis on how to regulate its use. This relevance can be also noted on how this class is one of the most frequently included in works dealing with fishing practices, with some of them even focusing exclusively on trawling [36, 45]. Furthermore, the availability of more data samples here allows us to conduct an ablation study (Sect. 4.3.1) to better understand how characteristics such as the sampling period of the number of data samples available affect the performance.

In both cases, we will follow a similar approach. First, we split the Tragsatec subset for the experiment into a training set with 70% of the samples, and a validation set with the remaining 30%. Using these partitions, we will search for the best hyper-parameter configuration for each classification model. Concretely, for the models using local features, the following hyper-parameters are explored:

- SVM. Two different hyper-parameters are tuned for the SVM, the complexity  $C$ , and the  $\gamma$  value. The complexity controls the trade-off between correctly classifying all training samples (i.e., low values of  $C$ ) and maximizing the margin of the classifier (i.e., high values of  $C$ ). On the other hand,  $\gamma$  controls the curvature of the decision boundary through the RBF function, with high values of

**Table 3** Results of different approaches for fishing gear classification. We report the mean accuracy and its 95% confidence interval (CI) (in %) computed after a  $K$ -fold stratified cross validation with  $K = 10$

Classifier	Features	Hyperparameters	Mean acc. (95% CI) [%]
BiGRU	Local	–	75.60 ± 3.62
MLP	Global	$L = 1,000; \alpha = 0.0001$	82.69 ± 2.00
SVM	Global	$C = 100; \gamma = 0.01$	83.16 ± 1.77
RF	Global	$N = 101$	86.22 ± 2.66
RF + BiGRU	Fusion	$w_{RF} = 0.8; w_{BiGRU} = 0.2$	<b>90.13 ± 3.13</b>

We include as well the hyper-parameter values for each approach

$\gamma$  representing more curvature. We will explore values of  $C \in [1, 10, 100]$ , and  $\gamma \in [0.1, 0.01, 0.001]$ .

- Random Forests. For the RF model, we will only explore the number of estimators  $N$ , which is the number of decision trees included in the forest. In this work, the consider values of  $N \in [101, 1, 10 K]$ .
- Neural Network. Two different hyper-parameters are considered for the NN classifier, namely the number of units  $L$  in the hidden layer, and the learning rate of the network  $\alpha$ . Note that the output layer of the Neural Network contains the same number of units as classes in the multi-class configuration (i.e., 5 output units) and uses softmax, while only 1 output unit with sigmoid activation is used for binary classification. As for the hyper-parameters, we will explore the values of  $L \in [100, 1, 10 K]$ , and  $\alpha \in [1e - 3, 1e - 4, 1e - 5]$ .

Noteworthy, the optimal weights for the local–global fusion scheme are also obtained with this strategy, using the set of optimal hyper-parameters that we have previously found for the local classifiers.

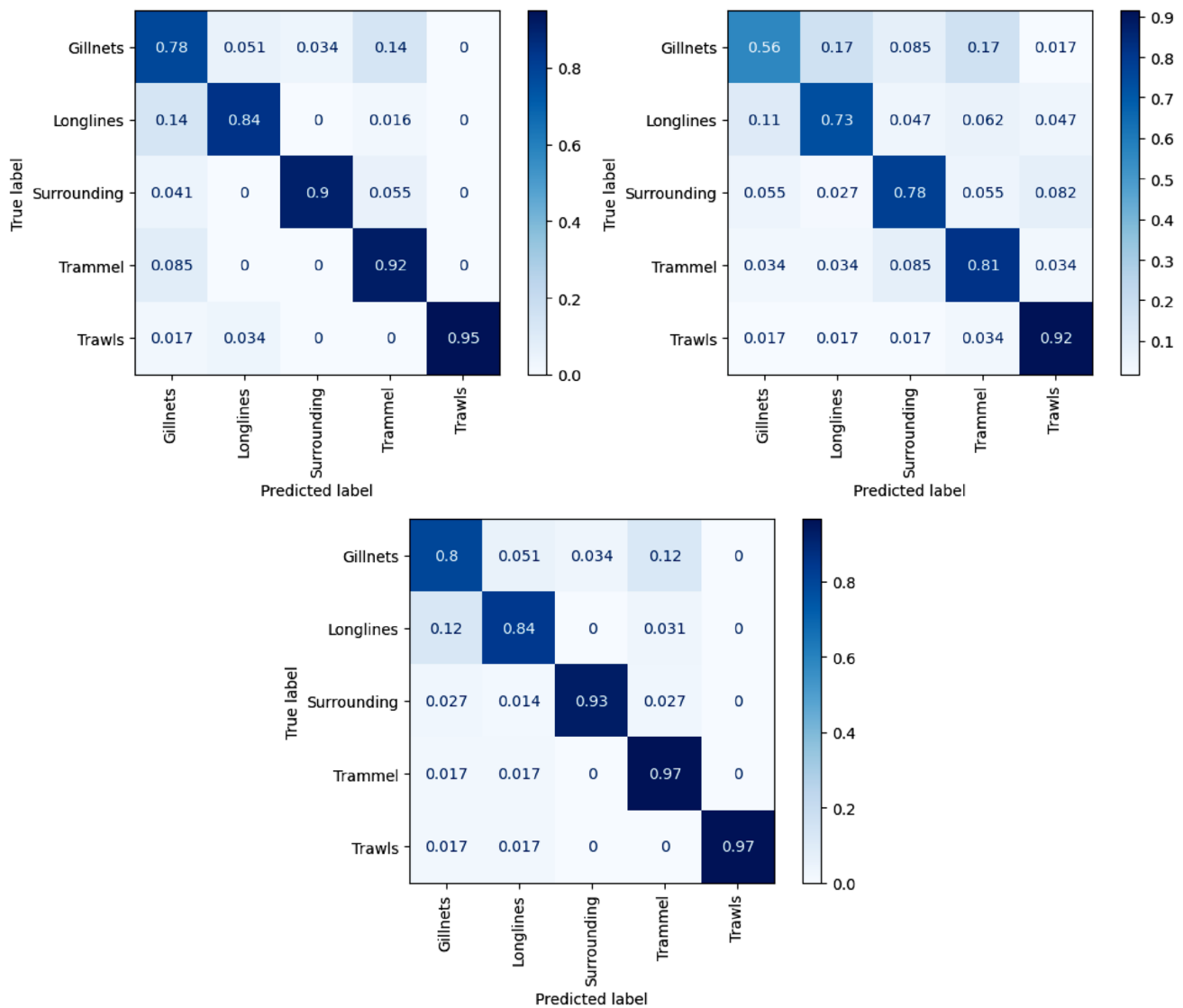
Once we have the optimal hyper-parameters, the final performance is assessed using a  $K$ -fold Stratified Cross Validation (SCV) protocol, in which the data is divided into 10 folds that preserve class proportions. Thus, we train the models 10 times, using in each iteration a different combination of 9 folds for training, and the remaining one for testing. The final performance score is obtained by aggregating the results of all iterations. When assessing the performance in a class-balanced configuration, the accuracy (or the mean accuracy in the 10-fold SCV) is employed. Otherwise, we decide to use the Mean Average Precision (mAP). Note that all our experiments in the multi-class configuration are based on a balanced dataset, so we only use accuracy-based metrics. We will also use some traditional performance tools, such as Confusion Matrix, Receiver Operating Curve (ROC), which illustrates the False Positive Rate (FPR) against the True Positive Rate (TPR) at different classification thresholds, Detection Error Tradeoff (DET) curves, which measures the FPR against False Negative Rate (FNR) at different classification thresholds, and the Equal Error Rate (EER), the operating point at which FPR is equal to FNR.

## 4.2 Multi-class fishing gear classification

As we introduced before, the purpose of this section is to explore the multi-class fishing gear classification task on the newly Tragsatec database. To this aim, we selected the 5 classes with more samples, namely *Trawls*, *Longlines*, *Surrounding*, *Trammel*, and *Gillnets*. An under sampling procedure was applied to our data in order to obtain a balanced corpus including 209 diary statements and trajectories per class.

The results of the 10-fold SCV are reported in Table 3 for the different classifiers considered. Note that we included as well the 95% confidence intervals. The best accuracy provided by a single classifier is 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. By considering in the same prediction information from both global and local features, obtained from the fusion of the scores provided by RF and BiGRU, a raise in performance above 90% is achieved. This is a relative error reduction of 28%. We consider this a promising result, and expect to increase even more the performance with the collection of more data samples, or improving the feature selection process.

In Fig. 4 we illustrate the confusion matrices obtained for the five fishing gears with the following classifiers: (i) RF (top-left), (ii) BiGRU (bottom), and (iii) fusion of RF and BiGRU at score level (top-right). We obtained this confusion matrices from the classifiers trained with the 70/30 training/validation splits used to determine the hyper-parameters. In all three cases, the *Trawls* class is the one exhibiting the best results, while *Gillnets* obtains the worst. Surprisingly, there are almost no errors associated to incorrectly predicting the *Trawls* class, which further highlights the performance when identifying this class. As observed in Fig. 4 (bottom), we obtain here an accuracy of 93% for the *Surrounding* class, a value greater than the total accuracy of the RF-BiGRU classifier (i.e., 90.13%). This is not the case of [7], where this class obtained the worst results. In such work, the data sequences employed a sampling period  $T_s$  of 2 h. This sampling rate not enough to classify the *Surrounding*



**Fig. 4** Confusion matrices obtained in the multi-class configuration with the following classifiers: RF based on global features (top-left); BiGRU based on local features (bottom); and fusion of RF and BiGRU based on both features (top-right)

gear (“Purse seine” in [7]) with an accuracy similar to the other classes. This fact further highlights the benefits of using a higher sampling rate to correctly characterize fishing gears through vessel trajectory data.

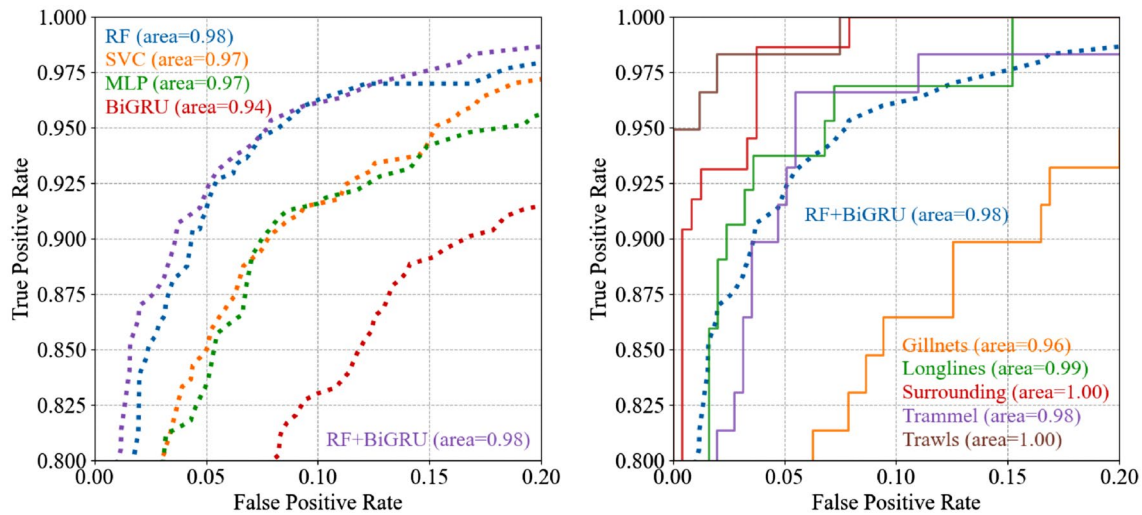
Finally, in Fig. 5 we report the Receiver Operating Characteristic (ROC) curves of different classifiers (left) and different fishing gears when RF + BiGRU is considered (right). As seen, the *Gillnets* obtained the lowest area under the curve, which is consistently with the per-class results of Fig. 4, where an error of 12% between *Gillnets* and *Trammel* was observed. This may be explained by the fact that the trammel fishing gears is basically a variant of the gillnets, as exposed in Sect. 3.

### 4.3 Binary fishing gear classification: trawls detection

Now that we have assessed the performance of the proposed methodology in a fishing gear multi-classification configuration, in this section we explore the binary classification task. Recalling from Sect. 4.1, in this case we use all the dataset to train binary classifier in a One-vs-All configuration, with the purpose of distinguish between the most frequent class (i.e., *Trawls*) and the rest. Given that in this experiment both classes are unbalanced, mAP is used as the performance metric.

The results obtained after the 10-fold stratified CV are reported in Table 4 for the different individual and fusion-based classification approaches. We include as well the





**Fig. 5** ROC curves of different classifiers (left), and different fishing gears (right) using the RF + BiGRU classifier (i.e., the one with best performance).

**Table 4** Mean Average Precision (mAP) results and its 95% Confidence Interval (CI) (in %) in the binary classification setup computed after a  $K$ -fold cross validation ( $K = 10$  folds)

Classifier	Features	Hyperparameters	mAP (95% CI) [%]
BiGRU	Local	–	99.81 ± 0.04
MLP	Global	$L = 1,000; \alpha = 0.0001$	99.82 ± 0.08
SVM	Global	$C = 10; \gamma = 0.01$	99.83 ± 0.03
RF	Global	$N = 1,001$	99.97 ± 0.01
MLP + BiGRU	Fusion	$w_{MLP} = 0.49; w_{BiGRU} = 0.51$	99.91 ± 0.01
SVM + BiGRU	Fusion	$w_{SVM} = 0.52; w_{BiGRU} = 0.48$	99.91 ± 0.01
RF + BiGRU	Fusion	$w_{RF} = 0.75; w_{BiGRU} = 0.25$	<b>99.98 ± 0.01</b>

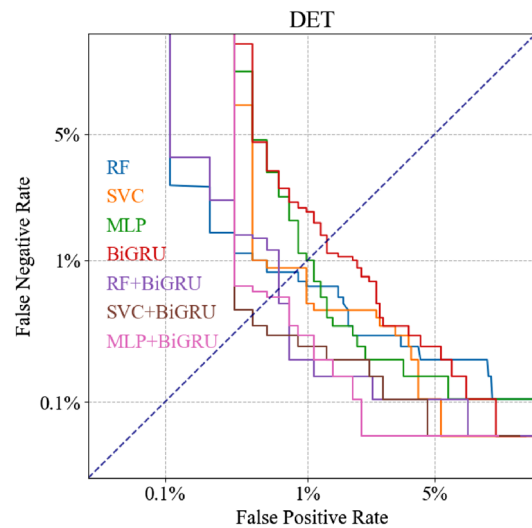
We include as well the optimal hyperparameter values for each approach

95% confidence intervals of the results obtained. The mAP value provided by a single classifier is 99.97%, obtained with RF with the global feature set. The SVM and MLP classifiers showed lower performances with 99.83 and 99.82% respectively. The BiGRU classifier obtained a mAP value of 99.81% using the local feature set. However, the best performance is obtained when combining both global

**Table 5** Equal Error Rate (EER) values (in %) of each approach for the binary fishing gear classification task

Classifier	Features	Threshold (%)	EER (%)
BiGRU	Local	45.11	1.32
MLP	Global	88.93	0.99
SVM	Global	70.32	0.89
RF	Global	53.15	0.86
MLP + BiGRU	Fusion	52.23	0.56
SVM + BiGRU	Fusion	51.06	<b>0.43</b>
RF + BiGRU	Fusion	57.96	0.64

We also include the thresholds from which the EER is obtained in each case



**Fig. 6** DET curves (the closer to the bottom-left the better) obtained with different classifiers for binary fishing gear classification. The intersection of a DET curve with the dotted line represents the EER point

and local information with a score fusion scheme, a fact already noticed in the multi-classification experiment (see Table 3). Note that all the fusion approaches exhibit better performances than the individual classifiers. The best Average Precision is 99.98% provided by the combination of the best global feature classifier (RF) with the best local feature classifier (BiGRU).

Figure 6 reports the Detection Error Tradeoff (DET) curves for all classifiers considered, from which the Equal Error Rate (EER) is obtained for each of them, as shown in Table 5. Given that all the systems analyzed obtained high, similar values of the mAP, the EER metric can help to better understand the possible differences between models. Therefore, the SVM + BiGRU classifier offers the best performance for FPR vs FNR, as its EER is the lowest with a value of 0.43%.

### 4.3.1 Ablation study: effect of number of training samples and sampling period

With the aim of further exploring the binary classification setup, we present here an ablation study to understand how factors such as the size of the training data or the sampling frequency affect the classification performance.

Attending first to the effect of the number of data samples available during the training process, we arbitrarily selected here 400 samples of each class as test set, with which we evaluated the performance of the classifiers trained in diverse scenarios. From the remaining samples, we made available a different number of training samples  $S_t$  to train the classifiers, including in this set an equal number of samples from each class  $S_c \in [100, 200, 500, 1K, 2, 2.64 K]$ . The last value (i.e., 2.64) corresponds to the maximum number of samples available for the *Non-Trawls* class after subtracting the test set samples. With this study we pretend to understand whether having more data samples contributes to performance, and which algorithms are best suitable for scenarios in which few data are available. Since both classes are balanced in this experiment, we use the accuracy as performance metric to assess the classifiers. The Detection Error Tradeoff (DET) curve and the Equal Error Rate (EER) are also used.

Table 6 presents the performance of different individual and fused classifiers when a different number of training samples is available. The performance is measured as the mean accuracy after a  $K$ -fold Cross Validation ( $K = 10$  folds). Note that in this case, instead of evaluating the performance with the corresponding test fold, we use always the fixed test set with 800 data samples (i.e., 400 from each class). To vary the number of training samples, we under-sampled the training sets to obtain balanced training subsets with  $S_t$  samples. We can observe a similar trend in all classifiers, which start from a performance around 94 – 95%

when few data samples are available, and the performance progressively increase until a peak around 98 – 99% when more samples are available. Note that we can observe the same trend in the confidence intervals, which are larger (i.e., more variability in the result across folds) when few data samples are available for training. In general, fusion schemes obtain better results than both local and global classifiers. For small amounts of data, the best single global feature classifier is RF, which improves its performance in combination with the BiGRU local feature classifier, although the performance of the latter is always the lowest when used separately. The SVM, however, exhibits the best performance as an individual classifier for both 2 and 4 K, a point from which it seems to saturate, as observed in the reduction in 5.28 K. Actually, the best performance is obtained with SVM + BiGRU when 4 K data samples are available (i.e., 99.43%). While this performance is slightly reduced in the next configuration (i.e., similar to what occurs with the SVM model alone), we can observe here the best results for several classifiers, including the other two fusion schemes, the RF and the BiGRU. As conclusion, the hypothesis that having more data samples available for training improves the performance seems to be corroborated, but we note that the performance can saturate after certain amount of data is used.

Now that we have observed the effect of the number of training samples on fishing gear classification performance, we want to assess the impact of the sampling period  $T_s$ . To this aim, we progressively reduce here the sampling frequency  $1/T_s$  of the GPS positions, and measure the performance of the classifiers trained in such scenarios. Given that the Tragsatec database presents a sampling period of  $T_s = 5$  minutes, one position is selected every 2, 4 and 10 positions, therefore exploring values of  $T_s \in [5, 10, 20, 35]$  minutes. The maximum sampling period of 35 minutes has been calculated taking into account that the minimum number of points in a trajectory must be 17 for the algorithms to work correctly, and that the trajectories are at most 30 miles offshore, as this is the coverage provided by the AIS beacon, most of them being approximately 10 hours (600 minutes). Therefore, the maximum sampling period is obtained as  $600/17 = 35$  minutes.

The same diary statements are always selected for the 4 sampling periods in order to be able to compare them in a consistent way. The number of samples of the train set is 372 and 731 from the *Non-Trawls* and *Trawls* classes respectively, while the test set comprises 160 and 314 samples from each of these. As the classes are imbalanced in this experiment, we use as performance metric mAP.

The mAP performance values obtained with the 10-fold CV are reported in Table 7 for the different single and combined classifiers. All top scores decrease, and almost all approaches exhibit a decay in performance as the sampling

**Table 6** Performance of different classification approaches depending on the number of training samples  $S_t$  available

$S_t$	Classifier	Features	Hyperparameters	Mean acc. (95% CI) [%]
200	BiGRU	Local	–	84.31 ± 0.87
	MLP	Global	$L = 1,000; \alpha = 0.001$	95.13 ± 0.16
	SVM	Global	$C = 10; \gamma = 0.01$	95.65 ± 0.27
	RF	Global	$N = 1,001$	95.85 ± 0.23
	MLP + BiGRU	Fusion	$w_{MLP} = 0.63; w_{BiGRU} = 0.37$	94.85 ± 0.30
	SVM + BiGRU	Fusion	$w_{SVM} = 0.85; w_{BiGRU} = 0.15$	95.48 ± 0.26
	RF + BiGRU	Fusion	$w_{RF} = 0.98; w_{BiGRU} = 0.02$	<b>95.93 ± 0.24</b>
400	BiGRU	Local	–	88.41 ± 0.79
	MLP	Global	$L = 1,000; \alpha = 0.0001$	96.44 ± 0.23
	SVM	Global	$C = 10; \gamma = 0.001$	96.06 ± 0.25
	RF	Global	$N = 10,001$	97.20 ± 0.16
	MLP + BiGRU	Fusion	$w_{MLP} = 0.72; w_{BiGRU} = 0.28$	96.20 ± 0.29
	SVM + BiGRU	Fusion	$w_{SVM} = 0.57; w_{BiGRU} = 0.43$	95.74 ± 0.18
	RF + BiGRU	Fusion	$w_{RF} = 0.95; w_{BiGRU} = 0.05$	<b>97.24 ± 0.18</b>
1 K	BiGRU	Local	–	94.13 ± 0.68
	MLP	Global	$L = 100; \alpha = 0.00001$	97.44 ± 0.23
	SVM	Global	$C = 10; \gamma = 0.01$	97.94 ± 0.12
	RF	Global	$N = 1,001$	98.26 ± 0.22
	MLP + BiGRU	Fusion	$w_{MLP} = 0.55; w_{BiGRU} = 0.45$	97.69 ± 0.08
	SVM + BiGRU	Fusion	$w_{SVM} = 0.81; w_{BiGRU} = 0.19$	97.90 ± 0.10
	RF + BiGRU	Fusion	$w_{RF} = 0.93; w_{BiGRU} = 0.07$	<b>98.38 ± 0.18</b>
2 K	BiGRU	Local	–	97.35 ± 0.46
	MLP	Global	$L = 1,000; \alpha = 0.0001$	98.85 ± 0.07
	SVM	Global	$C = 100; \gamma = 0.001$	98.71 ± 0.10
	RF	Global	$N = 101$	97.76 ± 0.22
	MLP + BiGRU	Fusion	$w_{MLP} = 0.50; w_{BiGRU} = 0.50$	<b>99.23 ± 0.16</b>
	SVM + BiGRU	Fusion	$w_{SVM} = 0.50; w_{BiGRU} = 0.50$	98.99 ± 0.22
	RF + BiGRU	Fusion	$w_{RF} = 0.75; w_{BiGRU} = 0.25$	98.95 ± 0.17
4 K	BiGRU	Local	–	96.91 ± 1.02
	MLP	Global	$L = 10,000; \alpha = 0.0001$	99.03 ± 0.11
	SVM	Global	$C = 100; \gamma = 0.001$	98.99 ± 0.13
	RF	Global	$N = 1,001$	98.71 ± 0.09
	MLP + BiGRU	Fusion	$w_{MLP} = 0.40; w_{BiGRU} = 0.60$	98.09 ± 0.65
	SVM + BiGRU	Fusion	$w_{SVM} = 0.62; w_{BiGRU} = 0.38$	<b>99.43 ± 0.11</b>
	RF + BiGRU	Fusion	$w_{RF} = 0.67; w_{BiGRU} = 0.33$	98.53 ± 0.35
5.28 K	BiGRU	Local	–	98.10 ± 0.26
	MLP	Global	$L = 10,000; \alpha = 0.001$	98.90 ± 0.12
	SVM	Global	$C = 100; \gamma = 0.01$	98.65 ± 0.13
	RF	Global	$N = 10,001$	98.99 ± 0.10
	MLP + BiGRU	Fusion	$w_{MLP} = 0.50; w_{BiGRU} = 0.50$	<b>99.36 ± 0.09</b>
	SVM + BiGRU	Fusion	$w_{SVM} = 0.51; w_{BiGRU} = 0.49$	99.28 ± 0.12
	RF + BiGRU	Fusion	$w_{RF} = 0.67; w_{BiGRU} = 0.33$	98.99 ± 0.10

We report here the mean accuracy and its 95% confidence interval (CI) (in %) after a 10-fold CV. In all cases, the test set is composed of 400 samples from each class. We include as well the optimal hyperparameter values for each configuration

period increases. We find an exception in the BiGRU classifier, which achieves a peak performance with a sampling period of 10, making the performance of the MLP + BiGRU classifier equal to that of the SVM + BiGRU classifier at that sampling period point. In light of these results, the Data

Curation process applied (see Sect. 3.2) could be modified to avoid discarding trajectories with momentary loss of AIS beacon coverage of up to 10 minutes. For all sampling periods, the best single global feature classifier is SVM, which improves its performance in combination with the BiGRU

**Table 7** Performance of different classification approaches depending on the sampling period  $T_s$  of the data

$T_s$ (min)	Classifier	Features	Hyperparameters	mAP (95% CI) [%]
5	BiGRU	Local	–	98.64 ± 0.26
	MLP	Global	$L = 100; \alpha = 0.001$	99.83 ± 0.02
	SVM	Global	$C = 10; \gamma = 0.01$	<b>99.93 ± 0.01</b>
	RF	Global	$N = 10,001$	99.86 ± 0.01
	MLP + BiGRU	Fusion	$w_{MLP} = 0.87; w_{BiGRU} = 0.13$	99.90 ± 0.02
	SVM + BiGRU	Fusion	$w_{SVM} = 0.93; w_{BiGRU} = 0.07$	<b>99.93 ± 0.01</b>
	RF + BiGRU	Fusion	$w_{RF} = 0.84; w_{BiGRU} = 0.16$	99.90 ± 0.02
10	BiGRU	Local	–	98.93 ± 0.19
	MLP	Global	$L = 100; \alpha = 0.001$	99.83 ± 0.03
	SVM	Global	$C = 10; \gamma = 0.01$	99.90 ± 0.02
	RF	Global	$N = 10,001$	99.77 ± 0.01
	MLP + BiGRU	Fusion	$w_{MLP} = 0.77; w_{BiGRU} = 0.23$	<b>99.91 ± 0.03</b>
	SVM + BiGRU	Fusion	$w_{SVM} = 0.64; w_{BiGRU} = 0.36$	<b>99.91 ± 0.02</b>
	RF + BiGRU	Fusion	$w_{RF} = 0.71; w_{BiGRU} = 0.29$	99.84 ± 0.02
20	BiGRU	Local	–	98.80 ± 0.27
	MLP	Global	$L = 1,000; \alpha = 0.001$	99.77 ± 0.08
	SVM	Global	$C = 100; \gamma = 0.01$	99.87 ± 0.02
	RF	Global	$N = 10,001$	99.66 ± 0.02
	MLP + BiGRU	Fusion	$w_{MLP} = 0.91; w_{BiGRU} = 0.09$	99.87 ± 0.03
	SVM + BiGRU	Fusion	$w_{SVM} = 0.85; w_{BiGRU} = 0.15$	<b>99.90 ± 0.02</b>
	RF + BiGRU	Fusion	$w_{RF} = 0.79; w_{BiGRU} = 0.21$	99.80 ± 0.02
35	BiGRU	Local	–	98.56 ± 0.28
	MLP	Global	$L = 100; \alpha = 0.0001$	99.67 ± 0.05
	SVM	Global	$C = 100; \gamma = 0.01$	99.70 ± 0.03
	RF	Global	$N = 1,001$	99.49 ± 0.04
	MLP + BiGRU	Fusion	$w_{MLP} = 0.71; w_{BiGRU} = 0.29$	<b>99.80 ± 0.04</b>
	SVM + BiGRU	Fusion	$w_{SVM} = 0.67; w_{BiGRU} = 0.33$	99.77 ± 0.03
	RF + BiGRU	Fusion	$w_{RF} = 0.63; w_{BiGRU} = 0.37$	99.71 ± 0.05

We report here the Mean Average Precision (mAP) and its 95% Confidence Interval (CI) (in %) after a 10-fold CV. We include as well the optimal hyperparameter values

classifier, although the performance of the latter is always the lowest as commented before. The performance of the RF classifier is in most cases the lowest of the global-based classifiers. The performance of the MLP classifier is similar to that of the other single global feature classifiers. All single global feature classifiers improve their performance, or remain the same but never get worse, when combined with the local feature classifier. The SVM + BiGRU classifier offers the best performance with a mAP equals to 99.93% with a sampling period of 5 minutes.

## 5 Conclusions

In this work, we have addressed the fishing gear classification task from GPS vessel trajectories data. We processed for this task the data collected by Tragsatec's Management of Agricultural and Fisheries Information Systems, which included information such as AIS beacon positions, date and location of departure and return, or the fishing gears

carried by fishing vessels in Spain waters. After applying a data curation process, we obtain a clean database to train and evaluate fishing gear from GPS trajectories. The proposed Tragsatec database comprises almost 10 K trajectories recorded from 828 fishing vessels, which are classified into one among 7 different fishing gears. This database reduces the Nyquist bandlimit of existing databases by more than 10 times, providing, a new resource to develop AI-based solutions to combat illegal fishing activities

We propose a fishing gear classification framework in which fishing vessels' dynamic trajectories are modeled according to both global and local set of features. We explore the analogy of vessel trajectories with the problem of dynamic handwritten signature verification to this end, adapting feature extraction methods proposed in the state-of-the-art of this biometric trait [30]. Our experiments validated the proposed feature extraction using several supervised learning classifiers, with performances up to 90% for multiclass fishing gear classification, and to 99%



when detecting trawling from other fishing practices. We consider this last results of especial relevance, due to the ecological concerns that bottom trawling has raised among international organizations.

Finally, we presented an ablation study to better understand how factors such as the amount of data available to train the models, or the sampling frequency of the GPS signals impact the performance of the models. We highlighted here how using a sampling period of minutes instead of hours is of significant relevance to obtain better results on fishing gear classification, hence confirming some of the conclusions previously exposed in [7].

**Acknowledgements** 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 work 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 projects INTER-ACTION (PID2021- 126521OB-I00 MICINN/FEDER) and HumanCAIC (TED2021-131787B-I00 MICINN). The work of A. Peña is supported by a FPU Fellowship (FPU21/00535) by the Spanish MIU. A. Morales is supported by the Madrid Government (Comunidad de Madrid-Spain) under the Multiannual Agreement with Universidad Autónoma de Madrid in the line of Excellence for the University Teaching Staff in the context of the V PRICIT (Regional Programme of Research and Technological Innovation).

**Funding** Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This work 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). The work of A. Peña is supported by a FPU Fellowship (FPU21/00535) by the Spanish MIU.

**Data availability** The Tragsatec database collected and used in this work has been made publicly available. Further instructions on how to download the database can be found in the following link <https://www.github.com/BiDALab/TrFGdb>.

## Declarations

**Ethical approval** This article does not contain any studies with human participants performed by any of the authors.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Jarvis RM, Young T (2023) Pressing questions for science, policy, and governance in the high seas. *Environ Sci Policy* 139:177–184
- Commission E (2015) Fighting illegal fishing: commission warns taiwan and comoros with yellow cards and welcomes reforms in Ghana and Papua New Guinea. European Commission Brussels, Belgium
- Pramod G, Nakamura K, Pitcher TJ, Delagran L (2014) Estimates of illegal and unreported fish in seafood imports to the USA. *Mar Policy* 48:102–113
- Sala E, Mayorga J, Bradley DEA (2021) Protecting the global ocean for biodiversity, food and climate. *Nature* 592:397–402
- Scholaert F (2023). Action plan to protect marine ecosystems for sustainable fisheries. European Parliamentary Research Service
- Marzuki MI, Gaspar P, Garelo R, Kerbaol V et al (2017) Fishing gear identification from vessel-monitoring-system-based fishing vessel trajectories. *IEEE J Oceanic Eng* 43(3):689–699
- Chuaysi B, Kiattisin S (2020) Fishing vessels behavior identification for combating IUU fishing: enable traceability at sea. *Wireless Pers Commun* 115(4):2971–2993
- Dunn DC, Jablonicky C, Crespo GO, McCauley DJ et al (2018) Empowering high seas governance with satellite vessel tracking data. *Fish Fish* 19(4):729–739
- Arasteh S, Tayebi M.A, Zohrevand Z, Glässer U, Shahir A.Y, et al. (2020) Fishing vessels activity detection from longitudinal AIS data. In: Proceedings of the international conference on advances in geographic information systems (SIGSPATIAL), pp 347–356
- Shahir A.Y, Tayebi M.A, Glässer U, Charalampos T, Zohrevand Z, et al. (2019) Mining vessel trajectories for illegal fishing detection. In: Proceedings of the IEEE international conference on big data, pp 1917–1927
- Floridi L, Cows J, King TC, Taddeo M (2021) How to design AI for social good: seven essential factors. *Ethics Govern Policies Artif Intell* 2021:125–151
- Floridi L, Cows J, Beltrametti M, Chatila R, Chazerand P et al (2021) An ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Ethics Govern Policies Artif Intell* 2021:19–39
- Tomašev N, Cornebise J, Hutter F, Mohamed S, Picciariello A et al (2020) AI for social good: unlocking the opportunity for positive impact. *Nat Commun* 11(1):2468
- Cows J, Tsamados A, Taddeo M, Floridi L (2021) A definition, benchmark and database of AI for social good initiatives. *Nat Mach Intell* 3:111–115
- Arcadu F, Benmansour F, Maunz A, Willis J, Haskova Z et al (2019) Deep learning algorithm predicts diabetic retinopathy progression in individual patients. *NPJ Digital Med* 2(1):92
- Wang D, Khosla A, Gargeya R, Irshad H, Beck AH (2016) Deep learning for identifying metastatic breast cancer. *arXiv/1606.05718*
- Jia JS, Lu X, Yuan Y, Xu G, Jia J et al (2020) Population flow drives spatio-temporal distribution of COVID-19 in china. *Nature* 582(7812):389–394
- Davenport T, Kalakota R (2019) The potential for artificial intelligence in healthcare. *Future Healthcare J* 6(2):94
- Yang R, Ford B.J, Tambe M, Lemieux A (2014) Adaptive resource allocation for wildlife protection against illegal poachers. In: Proceedings of the international conference on autonomous agents and multiagent systems (AAMAS), pp 453–460
- Rolnick D, Donti PL, Kaack LH, Kochanski AK, Lacoste A et al (2022) Tackling climate change with machine learning. *ACM Comput Surv (CSUR)* 55(2):1–96

21. Witharana C, Lynch HJ (2016) An object-based image analysis approach for detecting penguin guano in very high spatial resolution satellite images. *Remote Sens* 8(5):375
22. Chowdhury JR, Caragea C, Caragea D (2020) On identifying hashtags in disaster twitter data. In: *Proceedings of the AAAI conference on artificial intelligence* 34:498–506
23. Yadav A, Marcolino L, Rice E, Petering R, Winetrobe H et al (2015) Preventing HIV spread in homeless populations using PSINET. In: *Proceedings of the AAAI conference on artificial intelligence* 29:4006–4011
24. Yadav A, Chan H, Jiang AX, Xu H, Rice E et al (2016) Using social networks to aid homeless shelters: dynamic influence maximization under uncertainty. *Proc Int Conf Auton Agents Multiagent Syst (AAMAS)* 16:740–748
25. Wilder B, Onasch-Vera L, Hudson J, Luna J, Wilson N et al (2018) End-to-end influence maximization in the field. *Proc Int Conf Auton Agents Multiagent Syst (AAMAS)* 18:1414–1422
26. Newman N, Bergquist LF, Immorlica N, Leyton-Brown K, Lucier B, et al. (2018) Designing and evolving an electronic agricultural marketplace in Uganda. In: *Proceedings of the ACM conference on computing and sustainable societies (SIGCAS)*, pp 1–11
27. Thorstad R, Wolff P (2019) Predicting future mental illness from social media: a big-data approach. *Behav Res Methods* 51:1586–1600
28. Rahmattalabi A, Adhikari A.B, Vayanos P, Tambe M, Rice E, et al. (2018). Influence maximization for social network based substance abuse prevention. In: *Proceedings of the AAAI conference on artificial intelligence*, 32
29. Jean N, Burke M, Xie M, Davis WM, Lobell DB et al (2016) Combining satellite imagery and machine learning to predict poverty. *Science* 353(6301):790–794
30. Martinez-Diaz M, Fierrez J, Krish RP, Galbally J (2014) Mobile signature verification: feature robustness and performance comparison. *IET Biometrics* 3(4):267–277
31. Fierrez J, Ortega-Garcia J, Ramos D, Gonzalez-Rodriguez J (2007) HMM-based on-line signature verification: feature extraction and signature modeling. *Pattern Recogn Lett* 28(16):2325–2334
32. Melzi P, Rodriguez-Albala J.M, Morales A, Tolosana R, Fierrez J, Vera-Rodriguez R (2023). Fishing gear classification from vessel trajectories and velocity profiles: database and benchmark. In: *Proceedings of the Iberian conference on pattern recognition and image analysis (IbPRIA)*, pp 629–638
33. Series M (2014) Technical characteristics for an automatic identification system using time-division multiple access in the vhf maritime mobile band. Recommendation ITU: Geneva, Switzerland, pp 1371–1375
34. SOLAS, C.V. Safety of navigation'. Regulation 1
35. Kroodsma DA, Mayorga J, Hochberg T, Miller NA et al (2018) Tracking the global footprint of fisheries. *Science* 359(6378):904–908
36. Natale F, Gibin M, Alessandrini A, Vespe M et al (2015) Mapping fishing effort through ais data. *PLoS ONE* 10(6):0130746
37. Wu S, Zimányi E, Sakr M, Torp K (2022). Semantic segmentation of ais trajectories for detecting complete fishing activities. In: *2022 23rd IEEE International conference on mobile data management (MDM)*, pp 419–424
38. Nguyen D, Vadaine R, Hajduch G, Garelo R, et al.(2018). A multi-task deep learning architecture for maritime surveillance using ais data streams. In: *International conference on data science and advanced analytics (DSAA)*, pp 331–340
39. Chung J, Kastner K, Dinh L, Goel K, et al. (2015) A recurrent latent variable model for sequential data. In: *Proceedings of the international conference on neural information processing systems (NIPS)*, 28, pp 2980–2988
40. Huang I, Lee M, Nieh C, Huang J (2023) Ship classification based on ais data and machine learning methods. *Electronics* 13(1):98
41. Capobianco S, Millefiori LM, Forti N, Braca P et al (2021) Deep learning methods for vessel trajectory prediction based on recurrent neural networks. *IEEE Trans Aerosp Electron Syst* 57(6):4329–4346
42. Ristic B, La Scala B, Morelande M, Gordon N (2008). Statistical analysis of motion patterns in ais data: Anomaly detection and motion prediction. In: *International conference on information fusion*, pp 1–7
43. Capobianco S, Forti N, Millefiori LM, Braca P et al (2022) Recurrent encoder-decoder networks for vessel trajectory prediction with uncertainty estimation. *IEEE Trans Aerosp Electron Syst* 59:2554–2565
44. Marzuki MI, Garelo R, Fablet R, Kerbaol V, et al. (2015) Fishing gear recognition from vms data to identify illegal fishing activities in Indonesia. In: *OCEANS 2015 - Genova*, pp 1–5
45. Vermard Y, Rivot E, Mahévas S, Marchal P (2010) other: Identifying fishing trip behaviour and estimating fishing effort from vms data using Bayesian hidden markov models. *Ecol Model* 221(15):1757–1769
46. Ortiz M, Justel-Rubio A, Parrilla A (2013) Preliminary analyses of the iccat vms data 2010–2011 to identify fishing trip behavior and estimate fishing effort. *Collect Vol Sci Pap ICCAT* 69(1):462–481
47. Bez N, Walker E, Gaertner D, Rivoirard J et al (2011) Fishing activity of tuna purse seiners estimated from vessel monitoring system (vms) data. *Can J Fish Aquat Sci* 68(11):1998–2010
48. Zheng Q, Fan W, Zhang S, Zhang H et al (2016) Identification of fishing type from vms data based on artificial neural network. *South China Fisheries Sci* 12(2):81–87
49. Feng Y, Zhao X, Han M, Sun T, et al. (2019). The study of identification of fishing vessel behavior based on vms data. In: *Proceedings of the 3rd international conference on telecommunications and communication engineering*, pp 63–68
50. Kim K, Lee KM (2020) Convolutional neural network-based gear type identification from automatic identification system trajectory data. *Appl Sci* 10(11):4010
51. Souza EN, Boerder K, Matwin S, Worm B (2016) Improving fishing pattern detection from satellite ais using data mining and machine learning. *PLoS ONE* 11(7):0158248
52. Xing B, Zhang L, Liu Z, Sheng H et al (2023) The study of fishing vessel behavior identification based on ais data: a case study of the east china sea. *J Marine Sci Eng* 11(5):1093
53. Mardia KV, Hughes G, Taylor CC, Singh H (2008) A multivariate von mises distribution with applications to bioinformatics. *Canad J Stat* 36(1):99–109
54. Miller R (2014). Big data curation. In: *Proceedings of the International Conference on Management of Data (COMAD)*, 4
55. 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. <https://eur-lex.europa.eu/eli/reg/2013/1379/oj>. Accessed: 2022-11-21(2013)
56. Leonard K (2003) *Schaum's Outline of Business Statistics*, 4th edn. McGraw-Hill, USA
57. Fierrez J, Ortega-Garcia J (2008) On-line signature verification, pp 189–209
58. Tolosana R, Vera-Rodriguez R, Fierrez J, Ortega-Garcia J (2015). Feature-based dynamic signature verification under forensic scenarios. In: *International workshop on biometrics and forensics (IWBF 2015)*, pp 1–6
59. Sae-Bae N, Memon N (2014) Online signature verification on mobile devices. *IEEE Trans Inf Forensics Secur* 9(6):933–947
60. Jain A, Nandakumar K, Ross A (2005) Score normalization in multimodal biometric systems. *Pattern Recogn* 38(12):2270–2285

61. Fierrez J, Morales A, Vera-Rodriguez R, Camacho D (2018) Multiple classifiers in biometrics. Part 1: fundamentals and review. *Inf Fusion* 44:57–64

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.