

Article

Identification of Fishing Vessel Types and Analysis of Seasonal Activities in the Northern South China Sea Based on AIS Data: A Case Study of 2018

Yanan Guan ^{1,2} , Jie Zhang ^{2,3,4}, Xi Zhang ^{2,3,*}, Zhongwei Li ⁴, Junmin Meng ^{2,3}, Genwang Liu ^{2,3}, Meng Bao ^{2,3} and Chenghui Cao ^{2,3}

¹ School of Geosciences, China University of Petroleum, Qingdao 266580, China; b19010076@s.upc.edu.cn

² First Institute of Oceanography, Ministry of Natural Resources, Qingdao 266061, China; zhangjie@fio.org.cn (J.Z.); mengjm@fio.org.cn (J.M.); liu_genwang@126.com (G.L.); baomeng@fio.org.cn (M.B.); chenghui_cao@126.com (C.C.)

³ Technology Innovation Center for Ocean Telemetry, Ministry of Natural Resources, Qingdao 266061, China

⁴ College of Oceanography and Space Informatics, China University of Petroleum, Qingdao 266580, China; lizhongwei@upc.edu.cn

* Correspondence: xi.zhang@fio.org.cn

Abstract: In recent years, concern has increased about the depletion of marine resources caused by the overexploitation of fisheries and the degradation of ecosystems. The Automatic Identification System (AIS) is a powerful tool increasingly used for monitoring marine fishing activity. In this paper, identification of the type of fishing vessel (trawlers, gillnetters and seiners) was carried out using 150 million AIS tracking points in April, June and September 2018 in the northern South China Sea (SCS). The vessels' spatial and temporal distribution, duration of fishing time and other activity patterns were analyzed in different seasons. An identification model for fishing vessel types was developed using a Light Gradient Boosting Machine (LightGBM) approach with three categories with a total of 60 features: speed and heading, location changes, and speed and displacement in multiple states. The accuracy of this model reached 95.68%, which was higher than other advanced algorithms such as XGBoost. It was found that the activity hotspots of Chinese fishing vessels, especially trawlers, showed a tendency to move northward through the year in the northern SCS. Furthermore, Chinese fishing vessels showed low fishing intensity during the fishing moratorium months and traditional Chinese holidays. This research work indicates the value of AIS data in providing decision-making assistance for the development of fishery resources and marine safety management in the northern SCS.

Keywords: Automatic Identification System (AIS) data; the northern South China Sea (SCS); vessel type identification; fishing density; seasonal activities analysis



Citation: Guan, Y.; Zhang, J.; Zhang, X.; Li, Z.; Meng, J.; Liu, G.; Bao, M.; Cao, C. Identification of Fishing Vessel Types and Analysis of Seasonal Activities in the Northern South China Sea Based on AIS Data: A Case Study of 2018. *Remote Sens.* **2021**, *13*, 1952. <https://doi.org/10.3390/rs13101952>

Academic Editor: Jorge Vazquez

Received: 26 March 2021

Accepted: 11 May 2021

Published: 17 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The South China Sea (SCS) has the highest diversity of marine species in China, and it provides a rich fishery resource. Because the SCS is a semiencloded environment, replenishment of the biological resources in the northern SCS from other marine areas is limited and the quantity of resources depends largely on the region's internal primary productivity [1]. Owing to factors such as overfishing and destruction of the ecological environment, offshore fishery resources in the northern SCS are in serious decline. According to a report by the Center for Strategic and International Studies in 2018 [2]: "Since the 1850s, the total number of fishes in the northern SCS has decreased by 70–95%, and the fishing catch has decreased by 66–75% in the past 20 years." To protect the remaining fishery resources and realize the sustainable development of the fishery, it is necessary to understand the dynamics of fisheries through the monitoring and management of fishing vessels.

At present, the operational monitoring of fishing vessels is mainly done by two methods: satellite remote sensing and the Automatic Identification System (AIS). When remote sensing satellites are used to monitor fishing vessels, optical remote sensing and Synthetic Aperture Radar (SAR) images are used. Because of their high spatial resolution, optical satellites can directly monitor and identify fishing vessels during the daytime and obtain considerable information about the appearance of the vessels [3–5]. At night, information on the distribution of fishing vessels is obtained by detecting their lights (such as fish-collecting lights) [6–8]. Compared with optical satellites, radar-based satellite images are not affected by light or clouds and can monitor target ships 24 h a day under all weather conditions to obtain information on the position, geometry and direction of navigation [9–12]. For optical satellite images, Yang et al. [13] proposed a detection method based on sea surface feature analysis, using a linear function combining pixel and regional features to select candidate vessels on various sea surfaces. Shi et al. [14] presented an approach to detect ships in a “coarse-to-fine” manner, which enhances the separability between the vessel and the background mainly by converting the panchromatic image into a pseudohyperspectral format; a hyperspectral algorithm was used to quickly extract the target vessels. In research on ship monitoring with SAR images, based on the subaperture cross-correlation magnitude algorithm, Brekke et al. [15] added a sub-band extraction method to enhance the contrast between small target ships and the background sea, which was conducive to ship detection. A novel algorithm for nearshore ship detection based on multidirectional information from high-resolution SAR images was proposed by Hou et al. [16]. On the basis of the internal and external characteristics of nearshore ships and port areas, the method integrated multiple sources of information such as details of the shoreline and background, scattering mechanism, shape contour and deep feature information for nearshore target ship detection. Jin et al. [17] proposed a new lightweight patch-to-pixel convolutional neural network (CNN) for ship detection in polarimetric SAR images, which used dilated convolutions to expand the receptive field exponentially without adding model parameters.

Current research techniques can accurately detect target vessels at sea using satellite remote sensing, but it is difficult to identify the vessel type from the many targets. Additionally, because satellite remote sensing images can only capture the transient status of fishing vessels and have a long revisit period, it is difficult to continually monitor the navigational and fishing activities of fishing vessels in the long term.

The AIS is a shipboard communication device that sends real-time, uninterrupted information about a ship’s position and speed, for example [18,19]. The AIS is often used to prevent ship collisions. Conventional land-based AIS receives ship information up to 60 km offshore. Satellite-based AIS uses satellite platforms to carry AIS receivers, which can realize ship information monitoring anywhere in the world, especially in areas that lack land-based stations such as open oceans and polar regions. The use of land-based and satellite-based AIS receivers generates a large amount of AIS data, allowing for global dynamic monitoring of ships [20]. With improvement in AIS data processing methods, it is possible to mine and analyze the navigation, fishing and other activities of vessels in a certain region or even globally. This can provide evidence to determine the operational activities of fishing vessels as a whole.

Currently, research based on AIS data mining focuses on two aspects. The first concerns recognition of fishing vessel types. Kroodsma et al. [21] used a CNN algorithm to identify six types of fishing vessels, including trawlers, gillnetters, and longliners. Using speed as an observation variable, Souza et al. [22] developed three classification algorithms for trawlers, longliners and seiners. Huang et al. [23] constructed a fishing vessel trajectory recognition scheme using the feature engineering and machine learning schemes of XGBoost as its two key modules to identify nine types of fishing vessels. The second aspect extracted information such as fishing density and fishing catch from AIS data. Bertrand et al. [24] combined AIS data and fishing data to estimate the position of fishing operations for all *Engraulis ringens* purse seiners and used neural network algorithms

to accurately identify 83% of the real fishing sets. They then combined the fishing set distribution with the stock distribution to analyze the association between them. By adding sensitivity analysis to the neural network classification results, Joo et al. [25] further improved estimates of the number and location of fishing events. Wang [26] extracted the vessel position points during active fishing from a threshold value determined by the speed frequency distribution. After reshaping the trajectories of fishing vessels in Zhoushan fishing grounds, these features were used to determine whether the fishing vessels were actively fishing.

Although existing studies have demonstrated some progress, most of these methods rely on speed and its related information to identify the type of fishing. They use relatively few identification features and cannot detect the differences among various types of fishing vessel in more detail. To identify the types of fishing vessels accurately, it is necessary to explore the characteristic differences among fishing vessel types more comprehensively.

Comprehensive and systematic monitoring of the current status of fishery resources exploitation in the northern SCS is required. The current research on fisheries in the northern SCS focuses on fishery statistics, including fishery development and industry data [27] such as the total economic output value of fisheries, the area of aquaculture, the production of major species, the number of fishing vessels and the number of employees. We could not find any analysis of the current behavior and activities of fishing vessels in the northern SCS using AIS data.

In light of the above problems, we first mined the multidimensional statistical features such as speed, heading and position in the AIS trajectory data. Using machine learning methods, we built an identification model for fishing vessel types to accurately classify the three main fishing vessel types in the northern SCS: trawlers, gillnetters and seiners. We applied the trained model to the AIS data in April, June and September 2018. We used a total of approximately 150 million data points to classify the operational type of fishing vessels in the northern SCS. On the basis of the classification results, we analyzed the spatial and temporal distribution characteristics of the three fishing vessel types and explored the seasonal activity pattern of Chinese fishing vessels. We aimed to provide a reference for the development and scientific management of fishery resources in the northern SCS.

The remainder of this paper is structured as follows: Section 2 introduces the characteristics of AIS data and details the preprocessing steps and contents of the dataset. Section 3 covers the construction of the sample set and the feature extraction process for the different types of fishing vessels. Section 4 explains the LightGBM classification algorithm and gives the classification results of fishing vessels in the SCS. In Section 5, the seasonal activity pattern of Chinese fishing vessels in the northern SCS is analyzed based on the classification results. Finally, Section 6 presents the conclusions.

2. Data and Preprocessing

In this section, we first introduce the AIS data used in this paper. On this basis, a detailed description of the data preprocessing procedure is given.

2.1. Data

AIS data provide ship information in near real-time, including static information, such as the Maritime Mobile Service Identity (MMSI, which is unique for each ship), ship's length and width, IMO number, name, type, and dynamic information such as the time stamp, position (latitude and longitude), speed, heading, destination and draught [28,29]. The AIS data used in this paper were received from vessels in the northern SCS (Figure 1) in 2018. The data were purchased from a commercial company (<http://e.boloomo.com/> (accessed on 5 May 2021)), which can provide all AIS information of all ships in the world received by satellite, shore-based AIS receivers.

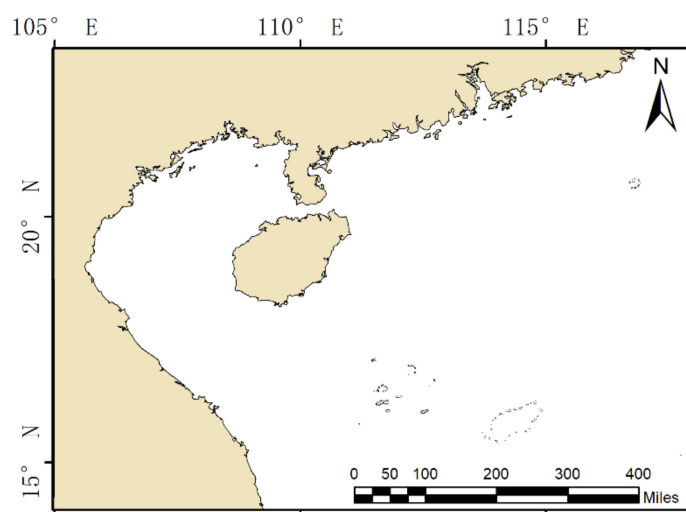


Figure 1. Map of the northern South China Sea (14°N–23.3°N, 105°E–118°E).

To study the seasonal activity pattern of fishing vessels in the northern SCS, April (spring), June (summer) and September (autumn) were selected as typical months for analysis. In April, the gradual increase in the water temperature leads to vigorous fish growth and fishery resources are abundant in the SCS. In June, the Chinese fishing season is closed in the northern SCS. According to Chinese law, from 12:00 a.m. on 1 May to 12:00 p.m. on 16 August, all types of fishing vessels except vessels with fishing tackle, as well as auxiliary vessels that provide support services for fishing vessels, are required to remain ashore. September is the peak season for fishery production, following the growth of juvenile fish in the closed season.

2.2. Data Preprocessing

Before the AIS data could be used, they were filtered. First, AIS data outside the study area were removed based on the vessel's positional information. The AIS defines multiple vessel types based on a unique type code [30]. In this paper, only fishing vessels were of interest (type code = 30); the AIS data for fishing vessels was filtered according to the type code.

Errors and incomplete data remained in the data set after the above processing. These were removed using the following process [31]:

Step 1: duplicate data were removed (identical vessel, time, and location) [32].

Step 2: AIS data with a speed greater than 12 knots were removed. In the work of [22], the AIS data with speed greater than 20 knots were regarded as outliers. However, data analysis revealed that there were few data with speeds greater than 12 knots in our AIS data, and the speed of all types of fishing vessels was usually 0–12 knots [23]. Therefore, the threshold of outliers was set to 12 knots.

Step 3: AIS data in the anchorage area, within three nautical miles from ports, and in the Qiongzhou Strait, were removed [33]. The fishing vessels in these areas are densely distributed and mostly at anchor or getting underway, which may adversely affect the classification of fishing vessel types.

Step 4: fishing vessels with less than 300 AIS track data points in total were removed. When there are not enough AIS data points for a vessel, the activity behavior of that fishing vessel cannot be accurately judged. According to the work of Ferrá et al. [34], the threshold value was set to 300.

Table 1 shows the amount of AIS data before and after each step of the preprocessing operation. The ratio represents the comparison between the number of AIS data points before and after each step of the operation.

Table 1. Comparison of Automatic Identification System data volume before and after pre-processing steps.

	Pre-Processing/Points	After-Processing/Points	The Ratio
Step 1	219,460,183	201,064,402	91.62%
Step 2	201,064,402	201,011,510	99.97%
Step 3	201,011,510	183,509,291	91.29%
Step 4	183,509,291	149,589,489	81.52%

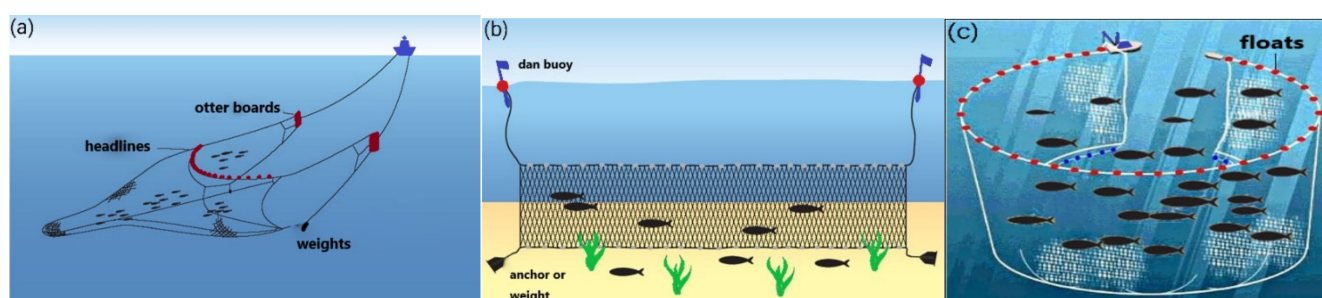
3. Fishing Vessel AIS Sample Set and Feature Extraction

This section first briefly introduces the characteristics of trawlers, gillnetters and seiners and describes the process of constructing the AIS sample set. Then the proposed process of feature analysis and extraction for fishing vessels are described in detail.

3.1. Sample Set

Trawlers, gillnetters and seiners are the main types of fishing vessels in the northern SCS [27]. In 2017, the fishing production of trawlers accounted for 41% of the total fishing production in the South China Sea, gillnetters accounted for 38% and seiners accounted for 15%. These three fishing techniques are mature and stable with a long history and have been adopted by most countries around the world, especially in Southeast Asia [35–39]. So, this work focuses on the identification of these three fishing vessel types.

Trawlers drag or pull trawls behind the vessel either on the sea floor (bottom trawling) or in the water column (pelagic or midwater trawling) [35]. A trawler may also operate two or more trawl nets simultaneously. Trawling is the active pursuit of fish, with a wide fishing range and high efficiency. The fishing targets are mainly valuable aquatic animals, such as fish, cephalopods, shrimps and crabs. Characteristically, the fishing vessel sails to the target fishing ground at a high speed and then slows down to release the net. After the nets are released, the fishing vessel drags the nets to catch the fish. In this process, the speed of the vessel is lower than the sailing speed and the change in heading is relatively large. The schematic diagram of a trawler is shown in Figure 2a.

**Figure 2.** The schematic drawing of a trawler (a), a gillnetter (b) and a seiner (c).

Gillnetters lay long ribbon-shaped nets vertically in the sea [40]. When fish try to cross the net, they are trapped in the mesh or entangled in the netting. After the nets are laid, gillnetters either moor or move slowly with the gear, generally at a speed of less than 2 knots. Their course may deviate to the left or right because of wind and waves. When the float at the top of the net begins to jiggle vigorously, the net is slowly pulled in. The entangled fish are pulled up, shaken out of the net, and thrown into the hold. The net is then reset, and the process begins again. The schematic diagram of a gillnetter is shown in Figure 2b.

The purse seines of seiners are horizontal nets that are deployed to hang vertically from floats around schooling fish on or near the surface by the vessel or by a separate skiff [41]. During the operation, the net surrounds a dense school of fish, forcing the fish to

enter the fishing capsule of the net, and the net is pulled in to catch fish. The schematic diagram of a seiner is shown in Figure 2c.

To obtain an effective and labeled sample data set, first the names of the three fishing vessel types were searched from the fishing vessel registration information officially published by the marine fishery management agencies and the official websites of fishing companies. Then, the MSSI of each ship was obtained from the China Ports Net (<http://ship.chinaports.com/> (accessed on 5 May 2021)) according to the ship name, and the AIS data of the fishing vessels were derived from the AIS database according to the MSSI. The final dataset had 3160 trawlers, 2635 gillnetters and 3240 seiners in the AIS sample data. The complete construction process of the sample data set is shown in Figure 3.

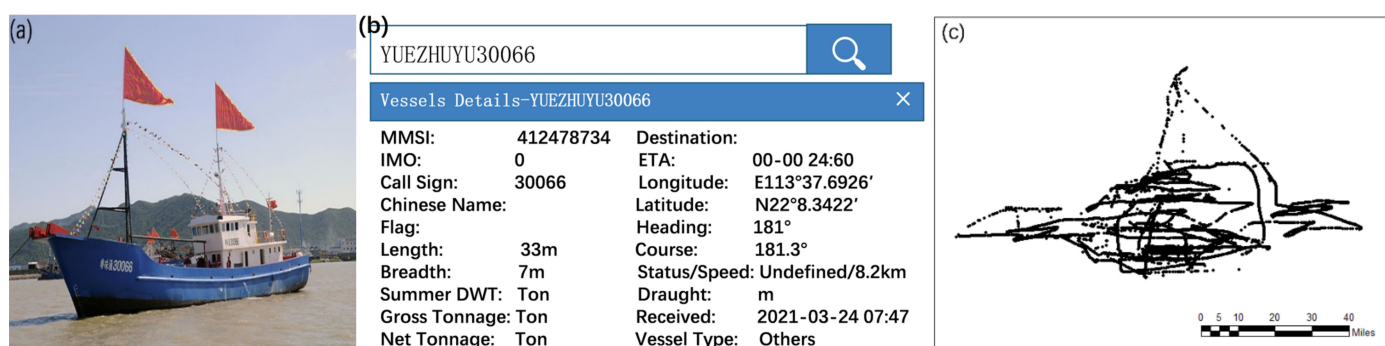


Figure 3. Collection process for the Automatic Identification System sample dataset. (a) Details of a trawler named YUEZHUYU30066 were obtained from the website of Ningbo Dajiang Shipbuilding Co. Ltd (<http://www.nbdjcy.com/content/?828.html> (accessed on 5 May 2021)). (b) The MMSI number of the trawler was identified through the fishing vessel query website of China Ports Net. (c) The ship's activity track in April 2018 was derived from the AIS database based on the MSSI. It should be noticed that the trajectories of all three types are shown in the Figure 6.

3.2. Analysis and Extraction of Fishing Vessels

To explore the differences in behavioral trajectories of different fishing vessel types more accurately, this paper extracted three major types of features from the AIS data: holistic features for speed and heading, features on position change and features for speed and displacement in multiple states. These were used for more precise recognition of fishing vessel types.

3.2.1. Holistic Characteristics of Speed and Heading

The continuous speed and heading changes of different fishing vessel types showed different characteristics, but were similar for the same fishing vessel type [32]. We analyzed the AIS trajectory data in the sample dataset separately, according to their categories, and plotted the frequency histograms and box diagrams of speed and heading for the different types of fishing vessels.

From Figure 4a, it can be seen that the frequency histograms of speed show an obvious bimodal distribution pattern at 0 knots and 4 knots: the difference between the values of the two peaks is relatively large, and the left peak is higher than the right peak. After discussing with experienced fishermen, we learned that the three types of fishing vessels without fishing float on the sea for a rest to save fuel. In addition, after lowering drift gillnets, the unpowered gillnetters move at a relatively low speed under the action of wind and wave. This situation may last from a few minutes to the greater part of the day, depending on current, weather and the number of fish being caught. The trawlers mostly drag nets to catch fish at speeds of about 4 knots. The seiner encircle the school at a speed of approximately 3 to 4 knots after a school of fish is located. According to the distribution of the peaks, three split lines are marked in Figure 4a: 0.6 knots, 2.5 knots and 6.0 knots, respectively. The three split lines divide the ship's speed into four partitions. Compared with speed, the frequency histograms of heading for the three types of fishing vessels in

Figure 4b were more similar and approximated a single-peak distribution. From Figure 5, it can be seen that the statistics of the mean, standard deviation, the lower, median and upper quartile were able to capture the differences between the three fishing vessels in terms of speed and heading. The corresponding values of the statistics are shown in Table 2.

Statistics, such as the mean, lower quartile, median and upper quartile, can be used to characterize data trends. Among them, the mean, also known as the “arithmetic mean”, refers to the average of the sample data. It is the most commonly used basic statistic. The standard deviation (STD) and dispersion coefficient can be used to characterize the discrete trend of the data. The STD is the most commonly used. The larger the sample STD, the greater the probability that the sample value deviates from the sample mean.

From Table 2, it can be seen that the mean, STD, lower quartile, median and upper quartile of speed and heading had large differences for trawlers, gillnetters and seiners, which were caused by the different operational methods. When fishing, gillnetters are mainly affected by wind and waves with small amplitude movements, which makes the five statistical values for their speed and heading much smaller than those of others. Therefore, we calculated these five statistics for the speed and heading in the AIS trajectory data of fishing vessels, which constituted a 12-dimensional feature vector, labeled F_1 – F_{12} , which could be used to represent the difference in the holistic characteristics of speed and heading.

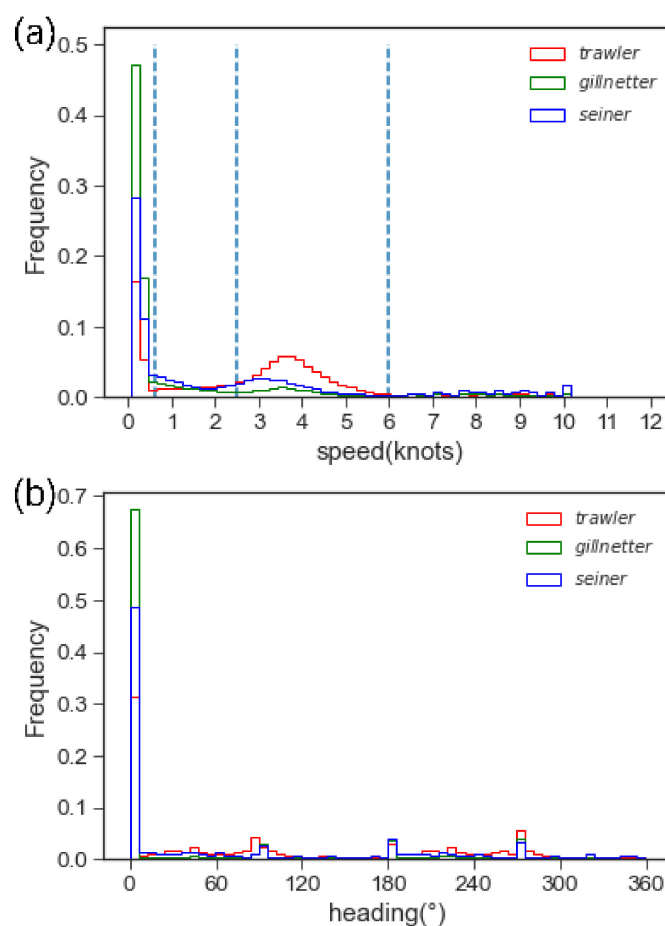


Figure 4. Frequency histograms of speed (a) and heading (b) for trawlers, gillnetters and seiners.

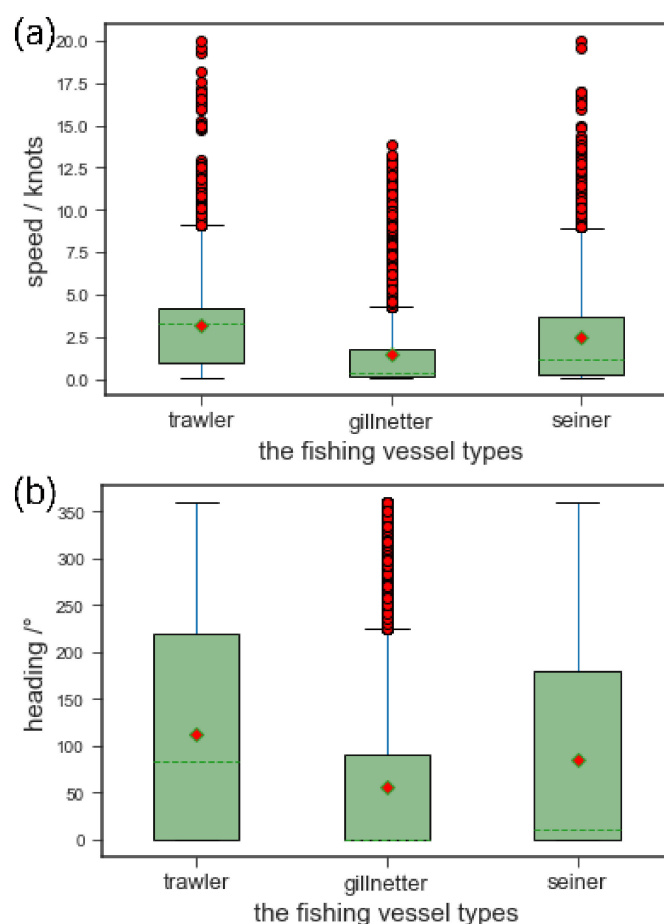


Figure 5. Box diagrams of speed (a) and heading (b) for trawlers, gillnetters and seiners.

Table 2. Statistical values corresponding to the box diagrams of the speed and heading for trawlers, gillnetters and seiners.

Statistics	Speed/Knots			Heading/°		
	Trawler	Gillnetter	Seiner	Trawler	Gillnetter	Seiner
mean	3.14	1.47	2.49	113.39	57.12	84.88
STD	2.41	2.34	2.89	110.72	97.46	109.09
lower quartile	0.92	0.11	0.22	0.00	0.00	0.00
median	3.29	0.32	1.19	84.00	0.00	11.00
upper quartile	4.21	1.78	3.72	219.00	90.00	180.00

3.2.2. Characteristics of Location Changes

Significant differences in the trajectories of different types of fishing vessels were observed (as shown in Figure 6). Trawlers mainly relied on sailing to drag the fishing gear. During a single operation, the trawl was generally towed in one direction, so the trajectory showed continuous changes of direction over a short time. Figure 6a clearly demonstrates that the trajectory of this trawler was crisscrossed east-west, and the fishing site was relatively large but concentrated. The working locations of gillnetters were generally far away from port and not continuous in space. During fishing, the operation track of gillnetters was relatively straight and tight (Figure 6b). For seiners, Figure 6c shows that their fishing areas were usually far from port, and the operation was carried out by a single vessel releasing the net and then returning to the origin, so a single operation trajectory of a seiner resembles a closed loop.

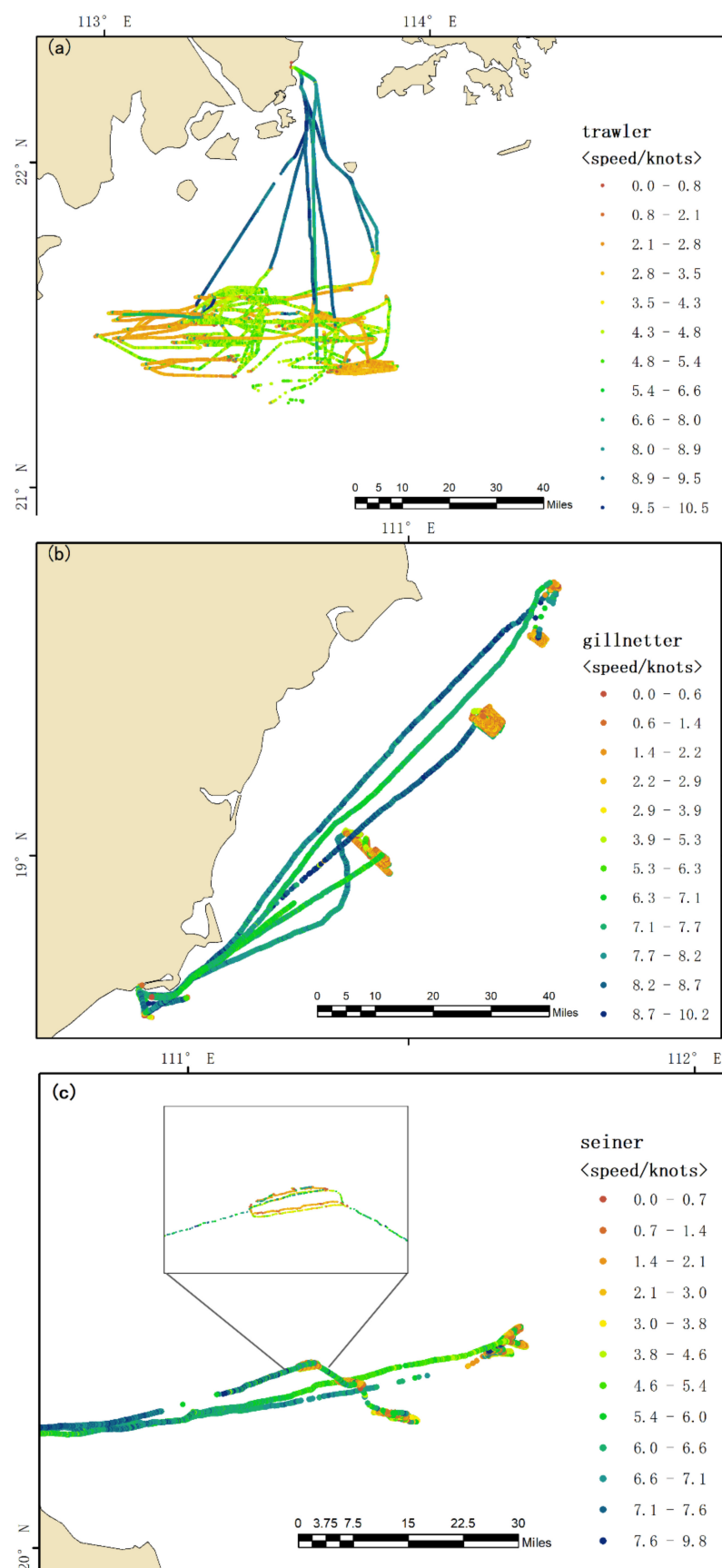


Figure 6. Fishing trajectory of a trawler (a), gillnetter (b) and seiner (c), where different colors are used to indicate speed intervals.

To further explore the characteristics of the position changes of the three fishing vessels during fishing operations, the change curves of the latitude and longitude values in the AIS data were plotted (Figure 7). Because the trajectories generally showed continuous east-west or north-south variations, the longitude of the trawler was more variable, random and irregular compared with the latitude in (Figure 7a). Figure 7b,c, show that the changes in longitude and latitude were more synchronized for gillnetters and seiners, especially for gill net vessels. However, compared with the other vessels, the fishing sites of the gillnetters were more scattered.

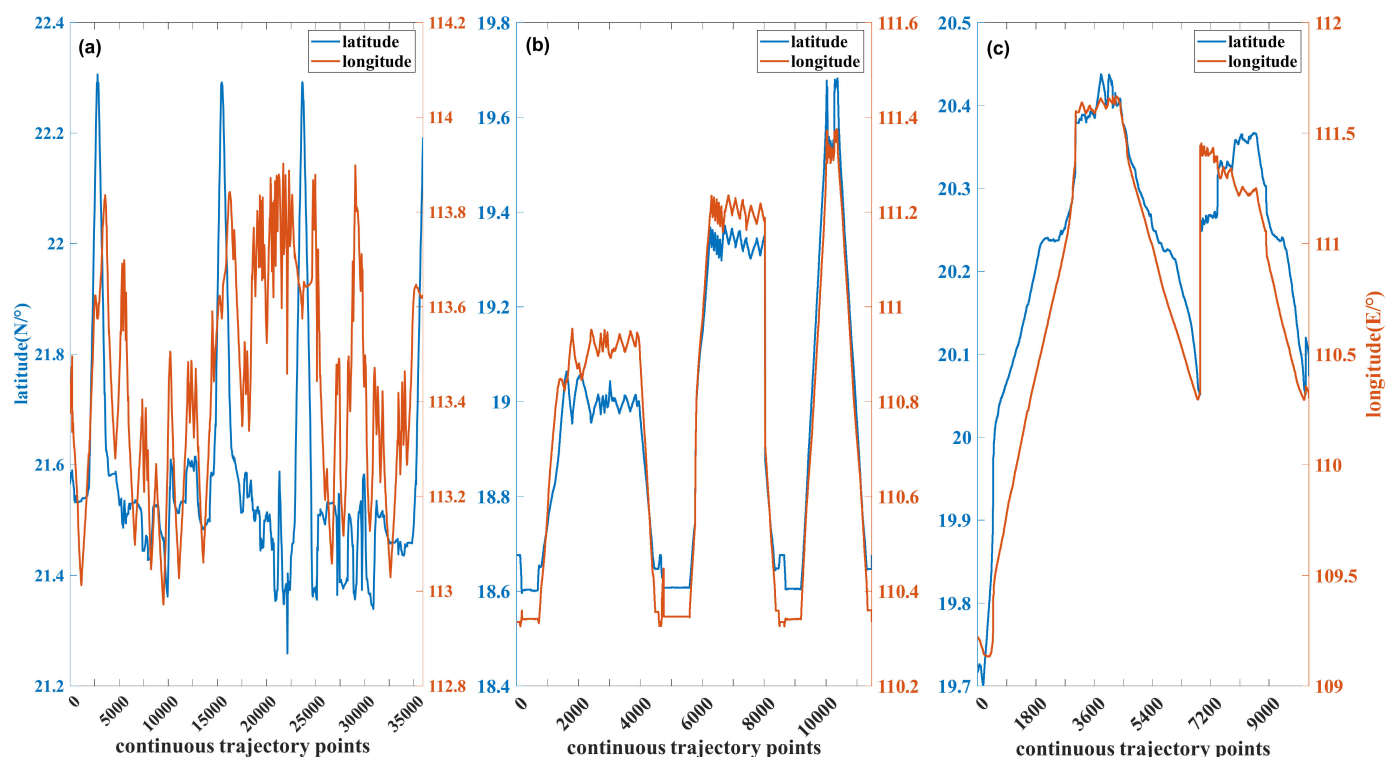


Figure 7. Latitude and longitude change curves of a trawler (a), gillnetter (b), and seiner (c).

On the basis of the above analysis, to characterize the positional change of different types of fishing vessels accurately, we calculated for each vessel the positional change at each moment (x, y) relative to the initial moment (x_0, y_0) ; that is, we calculated the longitude and latitude change values Δlon_{x-x_0} , Δlat_{y-y_0} . The mean, lower quartile, median, upper quartile, STD and dispersion coefficient of the change values were calculated, which constituted a total of 12 features showing the positional change of fishing vessels labeled as F_{13} – F_{24} .

3.2.3. Classification Features in Multiple States

To explore the characteristic differences in dynamic information, such as speed and displacement, of different fishing vessel types in different states, such as fishing and sailing, we divided the fishing vessel tracks into multiple states by segmenting the speed and displacement. The specific operations were as follows.

According to the three split lines labeled in Figure 4a, the sailing speed of fishing vessels was divided into four segments, (0.0 knots, 0.6 knots), (0.6 knots, 2.5 knots), (2.5 knots, 6.0 knots) and (6.0 knots, 12 knots). We calculated the statistical values, including the mean, lower quartile, median, upper quartile, STD and dispersion coefficient of each vessel in the four speed segments. These were used to represent the differential characteristics of the speed in different states. These characteristics constituted 24 features, labeled as F_{25} – F_{48} .

To extract features that could characterize the displacement differences in multiple states, we first calculated the displacement per second based on the latitude and longitude in the AIS data, which was the average displacement per second between trajectory data points at adjacent times. The calculation formula is as follows:

$$s = 2 * R * \arcsin(\sqrt{\sin^2(|x_{i+1} - x_i|/2) + \cos(x_i) * \cos(x_{i+1}) * \sin^2(|y_{i+1} - y_i|/2)}) / (t_{i+1} - t_i) \quad (1)$$

$$i = 0, 1, 2, \dots, N - 1$$

$$R = 6371.004 \text{ km}$$

where x_i , x_{i+1} denotes the longitude of the adjacent moment, y_i , y_{i+1} denotes the latitude of the adjacent moment, t_i , t_{i+1} denotes the time stamp and R is the Earth's radius.

According to the segmental speeds, the segmental displacements per second were 0.3, 1.3, 3.1 and 6.2 m (1 knot = 0.514 m/s). The displacements per second of fishing vessels were divided into four segments (0.0 m, 0.3 m), (0.3 m, 1.3 m), (1.3 m, 3.1 m) and (3.1 m, 6.2 m). The mean, lower quartile, median, upper quartile, STD and dispersion coefficient of the displacement in each displacement section formed 24 features, labeled as F_{49} – F_{72} , which represented the displacement differences of different fishing types under different states.

Having completed the extraction of mean, lower quartile, median, upper quartile, STD and dispersion coefficient features for speed, heading, latitude and longitude variations and displacement, we constructed a 72-dimensional feature vector. Compared with the feature extraction based on speed alone in traditional methods, this feature vector not only included speed but also features such as heading and position changes, which characterize the feature differences of different types of fishing vessels in a multidimensional fashion. The detailed description is shown in Table 3.

Table 3. Description of each feature.

Feature	Describe
$F_1, F_2, F_3, F_4, F_5, F_6$	the mean, lower quartile, median, upper quartile, STD and dispersion coefficient of speed
$F_7, F_8, F_9, F_{10}, F_{11}, F_{12}$	the mean, lower quartile, median, upper quartile, STD and dispersion coefficient of heading
$F_{13}, F_{14}, F_{15}, F_{16}, F_{17}, F_{18}$	the mean, lower quartile, median, upper quartile, STD and dispersion coefficient of $\Delta \text{lot}_{x-x_0}$
$F_{19}, F_{20}, F_{21}, F_{22}, F_{23}, F_{24}$	the mean, lower quartile, median, upper quartile, STD and dispersion coefficient of $\Delta \text{lat}_{y-y_0}$
$F_{25}, F_{26}, F_{27}, F_{28}, F_{29}, F_{30}$	the mean, lower quartile, median, upper quartile, STD and dispersion coefficient of speed segment (0knots, 0.6knots)
$F_{31}, F_{32}, F_{33}, F_{34}, F_{35}, F_{36}$	the mean, lower quartile, median, upper quartile, STD and dispersion coefficient of speed segment (0.6knots, 2.5knots)
$F_{37}, F_{38}, F_{39}, F_{40}, F_{41}, F_{42}$	the mean, lower quartile, median, upper quartile, STD and dispersion coefficient of speed segment (2.5knots, 6knots)
$F_{43}, F_{44}, F_{45}, F_{46}, F_{47}, F_{48}$	the mean, lower quartile, median, upper quartile, STD and dispersion coefficient of speed segment (6knots, 12knots)
$F_{49}, F_{50}, F_{51}, F_{52}, F_{53}, F_{54}$	the mean, lower quartile, median, upper quartile, STD and dispersion coefficient of displacement segment (0m, 0.3m)
$F_{55}, F_{56}, F_{57}, F_{58}, F_{59}, F_{60}$	the mean, lower quartile, median, upper quartile, STD and dispersion coefficient of displacement segment (0.3m, 1.3m)
$F_{61}, F_{62}, F_{63}, F_{64}, F_{65}, F_{66}$	the mean, lower quartile, median, upper quartile, STD and dispersion coefficient of displacement segment (1.3m, 3.1m)
$F_{67}, F_{68}, F_{69}, F_{70}, F_{71}, F_{72}$	the mean, lower quartile, median, upper quartile, STD and dispersion coefficient of displacement segment (3.1m, 6.2m)

4. Identification of Fishing Vessel Type Based on AIS Data

In this paper, we used the Light Gradient Boosting Machine (LightGBM) algorithm to identify fishing vessel types. In this section, we first introduce the basic rationale of LightGBM. Then the process of feature selection and evaluation results of the performance of the classification model are described.

4.1. LightGBM Model

LightGBM is an efficient gradient boosting decision tree (GBDT) algorithm proposed by Microsoft Research Asia [42,43]. Compared with GBDT, LightGBM has the advantages of faster training efficiency, lower memory usage, higher accuracy, support for parallel and GPU computing and can meet the needs of large-scale data processing, which can overcome the traditional boosting algorithm's disadvantages in scalability and running speed. It is now widely used in classification, regression and sorting applications [44–48]. LightGBM adopts two methods, Gradient-based One-Side Sampling (GOSS) and Mutually Exclusive Feature Bundling (EFB), to improve the training and learning speed. Algorithm 1 illustrates the LightGBM method.

Algorithm 1. The LightGBM algorithm

Input: Training set: $T = \{(x_1, y_1), (x_2, y_2) \cdots, (x_n, y_n)\}$; differentiable loss function: $L(y, f(x))$;

Number of iterations: M ; sampling rate of large gradient data: f , sampling rate of small gradient data: z ;

Output: LightGBM model $\hat{f}(x)$;

(1) Combine the mutually exclusive features of XI (not accepting nonzero features at the same time) through the exclusive feature binding (EFB) technology;

(2) Set $f_0(x) = \operatorname{argmin}_c \sum_i^N L(y_i, c)$;

(3) For $m = 1$ to M do

(4) Calculate the absolute value of the gradient:

$$r_i = \left| \frac{\partial L(y_i, f(x))}{\partial f(x)} \right|_{f(x)=f_{m-1}(x)}, i = \{1, \dots, N\}$$

(5) Resample the dataset using gradient-based one-sided sampling process:

$$highN = f \times \text{len}(T); randN = z \times \text{len}(T); sorted = \text{GetSortedIndices}(abs(r)); F = sorted[1 : highN];$$

$$Z = \text{RandomPick}(sorted[highN : \text{len}(T)], randN); T' = F + Z;$$

(6) Calculate IG

$$V_j(d) = \frac{1}{n} \left(\frac{(\sum_{x_i \in F_l} r_i + \frac{1-a}{b} \sum_{x_i \in Z_l} r_i)^2}{n_l^l(d)} + \frac{(\sum_{x_i \in F_r} r_i + \frac{1-a}{b} \sum_{x_i \in Z_r} r_i)^2}{n_r^r(d)} \right)$$

(7) Build a new decision tree $f_m(x)'$ on the set T'

(8) Update $f_m(x) = f_{m-1}(x) + f_m(x)$

(9) END of for

(10) Return $\hat{f}(x) = f_M(x)$

Given a training set, $X = \{(x_i, y_i)\}_{i=1}^n$, the goal of LightGBM is to find the approximate value $\hat{f}(x)$ of a certain function $f(x)$ to minimize the expected value of a certain loss function $L(y, f(x))$, as shown below:

$$\hat{f}(x) = \operatorname{argmin}_{y,x} E_{y,x} L(y, f(x)) \quad (2)$$

LightGBM integrates many classification and regression trees to approximate the final model, which is:

$$f_T(X) = \sum_{t=1}^T f_t(X), f_t \in \Gamma \quad (3)$$

where f_t represents the k th decision tree and Γ is the set of all trees. The regression tree can be represented as $\omega_{q(x)}$, $q = \{1, 2, \dots, J\}$, J is the number of leaves, q represents the decision rules of the tree, and ω represents the sample weights of leaf nodes.

$$G_t \cong \sum_{i=1}^n \left(g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right) = \sum_{j=1}^J \left(\left(\sum_{i \in I_j} g_i \right) \omega_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) \omega_j^2 \right) \quad (4)$$

where g_i and h_i represent the first and second-order gradient statistics of the loss function. Consider I_j represents the sample set of leaf J .

Given the tree structure $q(x)$, the optimal leaf weight fraction ω_j^* and the extreme value G_T^* on each leaf node are as follows:

$$\omega_j^* = - \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i} \quad (5)$$

$$G_T^* = - \frac{1}{2} \sum_{j=1}^J \frac{\left(\sum_{i \in I_j} g_i \right)^2}{\sum_{i \in I_j} h_i + \lambda} \quad (6)$$

The objective function after adding the split is:

$$G = \frac{1}{2} \left[\frac{\left(\sum_{i \in I_L} g_i \right)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{\left(\sum_{i \in I_R} g_i \right)^2}{\sum_{i \in I_R} h_i + \lambda} + \frac{\left(\sum_{i \in I} g_i \right)^2}{\sum_{i \in I} h_i + \lambda} \right] \quad (7)$$

where I_L and I_R are the sample sets of the left and right branches, respectively, and λ is the coefficient of the penalty model.

LightGBM has greater improvement in accuracy compared with the traditional GBDT algorithm due to the adoption of level-wise tree growth, which enables trees to form with a more complex structure. However, this structure is prone to overfitting. It is necessary to control the number of leaf nodes and other hyperparameters of the LightGBM before training the model. In our experiment, a Bayesian optimization method was adopted to find the parameters that gave the LightGBM the best performance. By iterating the results based on different hyperparameter values until convergence to the highest value of the objective function, the Bayesian approach is able to approximate the optimal hyperparameters of the model accurately and quickly. The following parameters of LightGBM were optimized and controlled with the Bayesian hyperparameter method:

- num_leaves: the number of leaf nodes, which can adjust the complexity of the model;
- learning_rate: the impact of the learning level of each tree on the final result, which controls the learning progress of the model during the iteration;
- n_estimators: increasing the number of iterations within a reasonable range can help the model to converge.

Table 4 lists the LightGBM parameters, their value range, and optimal parameter values. These parameters were optimized using the Bayesian method in the proposed model, and other parameters were default values.

Table 4. Parameter settings for the light gradient boosting machine (LightGBM) algorithm.

Parameter	Range of Values	The Best-Optimized Parameters
learning_rate	{0.01, 0.5, 0.01}	0.05
n_estimators	{5, 100, 5}	45
num_leaves	{1000, 25, 000, 1000}	20,000

In Section 3, we explained how through feature extraction, the mean, lower quartile, median, upper quartile, STD and dispersion coefficient of the variation of speed, course, longitude and latitude, and displacement were extracted, and 72-dimensional feature vectors were constructed. However, not all features contributed to the classification, and some features conflicted with each other. Therefore, we used the feature importance function provided by the LightGBM algorithm to evaluate the features. The importance indicates the occurrence of the feature in the construction classification tree and is defined as the following equation:

$$G = \frac{1}{2} \left[\frac{(\sum_{i \in L_L} g_i)^2}{\sum_{i \in L_L} h_i + \lambda} + \frac{(\sum_{i \in L_R} g_i)^2}{\sum_{i \in L_R} h_i + \lambda} + \frac{(\sum_{i \in I} g_i)^2}{\sum_{i \in I} h_i + \lambda} \right] - \gamma \quad (8)$$

where γ is regularization on the additional leaf, and Gain records the number of times each feature causes the classification tree to split, which is an indicator of the feature's contribution to the classification. Then, we sorted the 72 features according to their importance scores, as shown in Figure 8.

We found that features F_{43} – F_{48} , related to the speed (6 knots, 12 knots), and features F_{67} – F_{72} , related to the displacement (3.1 m, 6.2 m) had relatively low importance, and hardly contributed to the classification. This was due to the fact that when the speed and displacement were in these two groupings, the fishing vessel was sailing at high speed and there was no significant difference in the behavior of all fishing vessel types in this state. Therefore, we removed features F_{43} – F_{48} and F_{67} – F_{72} and calculated the other features for each fishing vessel from the AIS trajectory data, generating a 60-dimensional feature vector. This feature vector was fed into the tuned LightGBM model, and the three classifiers were trained for three types of fishing vessels. For the vessels to be classified, each classifier generated the probability that the fishing vessel belonged to the corresponding type, and then identified the fishing vessel as the type with the highest probability.

4.2. Classification Accuracy Assessment

Cross-validation (CV) was used to evaluate the performance of the proposed fishing vessel type classification model [49]. K-Fold-CV is a statistical analysis method used to evaluate the implemented effect of machine learning models. In this paper, the model was assessed using a five-fold CV method. The dataset was randomly divided into five different subsets of the same size. For each validation stage, one subset (20%) was used as the test dataset, while the remaining four subsets (80%) were used for training. This process was repeated five times until each subset was used.

To evaluate the performance of the fishing vessel classification model, accuracy, precision, recall and the F1-score were considered. Accuracy is the most common evaluation index, which is the ratio of the number of samples correctly classified by the classifier compared to the total. Precision is a measure that indicates the proportion of true positive samples among the positive cases determined by the classifier, while recall reflects the proportion of correctly determined positive cases to the total positive cases. Recall with a high result indicates a low percentage of false negatives (FN), while precision with a high score indicates a low percentage of false positives (FP). Precision and recall with high percentages show that the model accurately returns the classification results [50]. The F1-score is the harmonic average of precision and recall, with a maximum of 1 and a minimum of 0. Table 5 shows the performance evaluation results of the classification model.

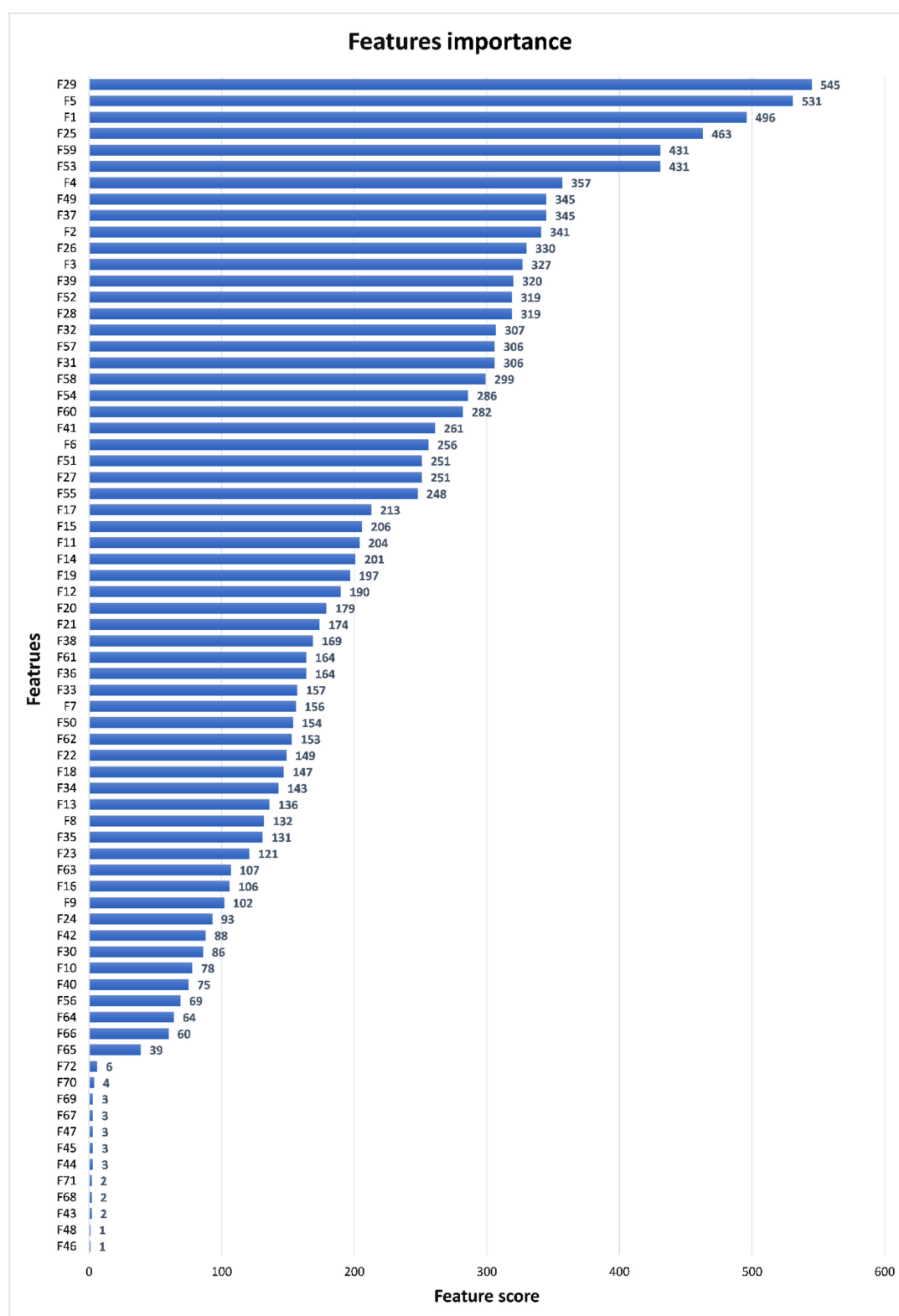


Figure 8. Feature importance ranking.

Table 5. Performance of model evaluation using the five-fold cross-validation procedure.

Fold No.	Accuracy	Precision	Recall	F1_Score
1	0.9568	0.9562	0.9579	0.9569
2	0.9558	0.9551	0.9567	0.9557
3	0.9563	0.9557	0.9571	0.9563
4	0.9573	0.9568	0.9583	0.9574
5	0.9579	0.9573	0.9589	0.9579
AVG	0.9568	0.9562	0.9578	0.9568

As shown in Table 5, the average accuracy score of the model was 95.68%, which was the percentage of correctly classified fishing vessels from the total number of fishing vessels classified. This shows the suitability of the method proposed in this paper to identify the type of fishing vessel. The average recall of the method was 95.78%, which indicates that the method was able to correctly classify more than 95% of the fishing vessel types with low FNs. The average F1-score was 95.68%, suggesting that the results obtained with the five-fold coefficient of variation method were consistent.

For further evaluation, we compare the proposed method with other advanced machine learning methods, such as logistic regression (LR) [51], support vector machine (SVM) [52], XGBoost [53], k-nearest neighbor (KNN) [54] and CatBoost [55]. The performance was evaluated in terms of the accuracy, precision, recall and F1-score. Table 6 shows that all the classifiers achieved high classification accuracy, which confirms that our feature vectors are effective. However, the classification method based on the LightGBM outperformed all the other methods (Table 6). The confusion matrix of the classification effect of each method is shown in Figure 9.

Table 6. Performance comparison of the suggested approach with other machine learning algorithms.

Classifier	Accuracy	Precision	Recall	F1_Score
KNN	0.9236	0.9141	0.8981	0.9053
SVM	0.8325	0.8070	0.7711	0.7835
XGBoost	0.9456	0.9414	0.9236	0.9317
Logistic regression	0.8610	0.8596	0.8626	0.8597
CatBoost	0.9265	0.9197	0.8973	0.9071
LightGBM	0.9568	0.9562	0.9578	0.9568

Although our sample dataset only had AIS data for three types of fishing vessels—trawlers, gillnetters and seiners—the AIS dataset to be classified contained track data for all fishing vessels in the study area, which included categories such as fishing transport vessels, line fishing vessels, and square netters. Before applying the trained classification model to the AIS dataset to be classified, we categorized these fishing vessels as "other" according to the following criteria, which were based on empirical values and the results of many experiments:

- (a) the maximum value among the predicted values of the three types was less than 0.5; and
- (b) the difference between the maximum and the larger value among the predicted values of the three types was less than 0.2.

For example, the probability of a fishing vessel's type being predicted as a trawler, gillnetter, and seiner was 0.42, 0.25 and 0.33, respectively, which meets the above criterion (a), and its type was judged to be "other".

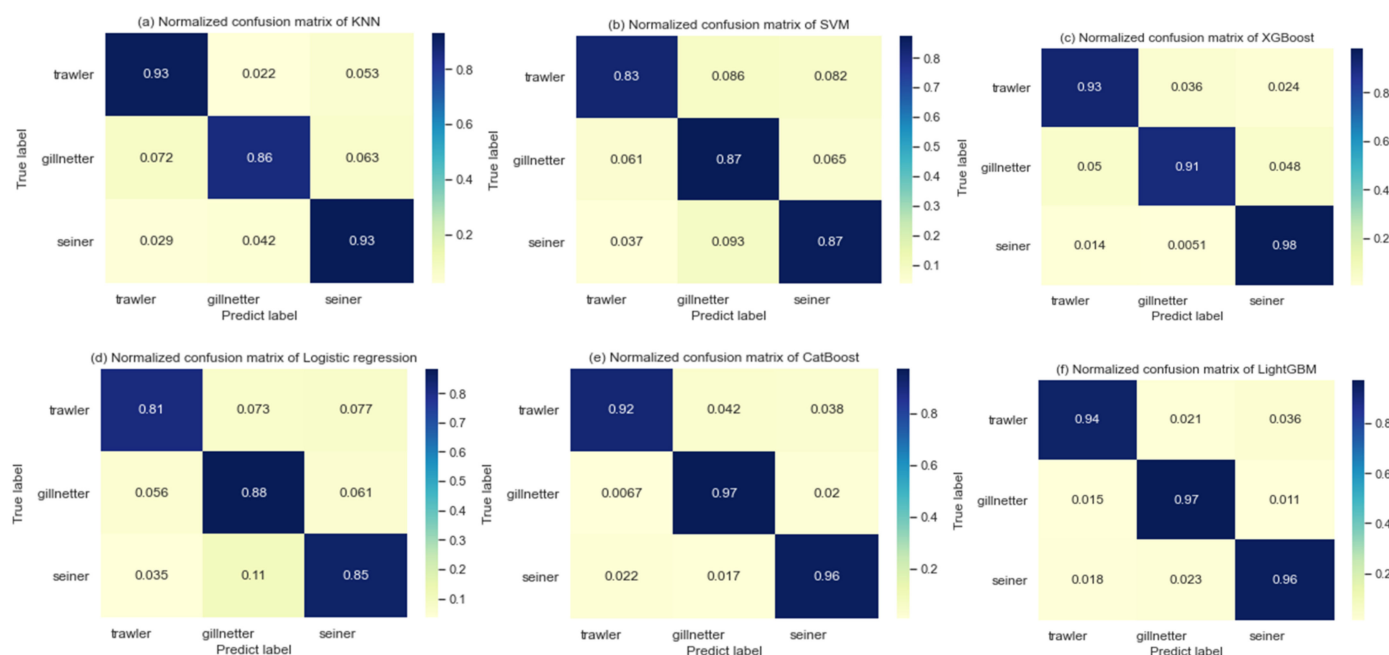


Figure 9. Normalized confusion matrices of classification for trawlers, gillnetters and seiners obtained by the different classifiers. The five classifiers are (a) KNN, (b) SVM, (c) XGBoost, (d) Logistic regression, (e) CatBoost and (f) LightGBM.

We applied the trained classification model to the AIS dataset of all fishing vessels in the northern SCS in April, June and September 2018 and obtained the number of trawlers, gillnetters, seiners and other categories of fishing vessels in each month (Figure 10).

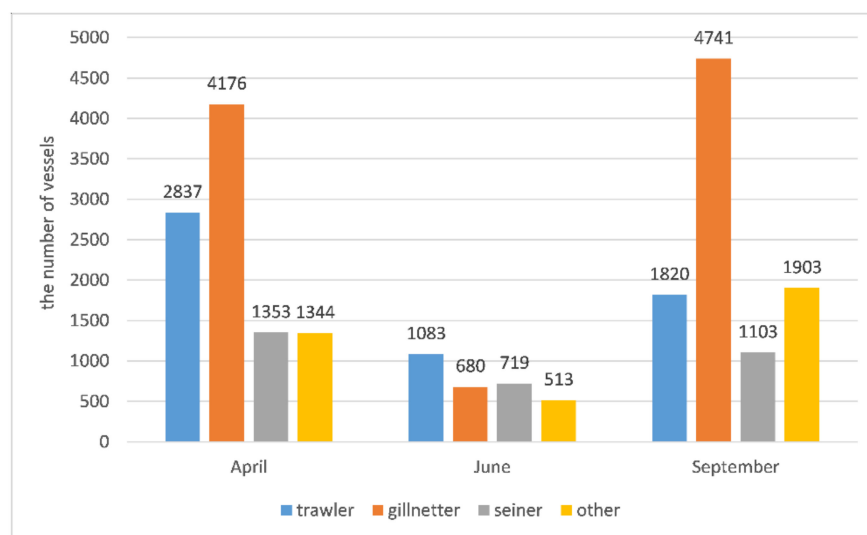


Figure 10. Fishing vessel type classification results for April, June and September 2018.

To verify the validity of the above classification criteria and the accuracy of the classification results of fishing vessels in the northern SCS, we randomly selected 200 fishing vessels from each of the four types of fishing vessels in the April classification results. In total, 800 fishing vessels were selected as the validation set. The official registered operation type of each fishing vessel was queried, and the statistical results are shown in Table 7.

Table 7. Statistical results of the actual operation types of each fishing vessel in the classification results.

Real Type Prediction Type	Trawler	Gillnetter	Seiner	Other	Accuracy
Trawler	190	6	3	1	0.950
Gillnetter	4	189	3	5	0.945
Seiner	2	3	192	3	0.965
Other	19	7	11	163	0.815

To investigate why some vessels were misclassified based on the statistical results in Table 7, we traced the records of these vessels and visualized the trajectory of the 67 fishing vessels that were misclassified. Four of these are shown in Figure 11. Among the misclassified fishing vessels, 35 fishing vessels were misclassified because they showed unclear features in the trajectory data. There were two situations that led to the misclassification.

1. The fishing transport vessels were illegally catching fish. As shown in Figure 11a, the trajectory of a fishery transport vessel matched the characteristics of the trawler's activity trajectory in Figure 6a. It was suspected that the transport vessel deployed some trawling gear and tried to fish illegally.
2. The trajectory data of fishing vessels were not complete enough to extract the trajectory characteristics of fishing. There were more trawlers, gillnetters and seiners among the fishing vessels classified as "other". After observing their trajectory visualization images, it was found the trajectories of 31 vessels were mostly smooth and straight lines (trajectories of the fishing vessels in navigation), as shown in Figure 11b–d. These fishing vessels were classified as "other" because the AIS data only recorded their activity trajectories when sailing but lacked trajectory data during fishing.

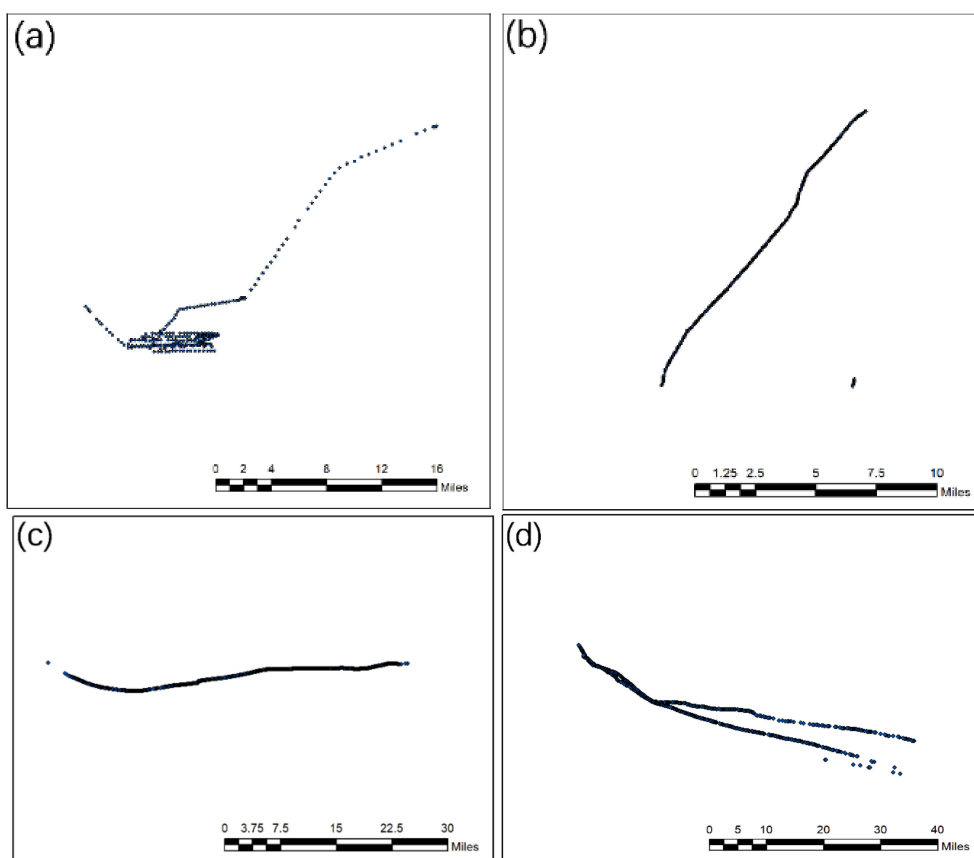


Figure 11. Trajectories of misclassified ships: (a) a fishing transport vessel misclassified as a trawler; (b) a trawler misclassified as "other"; (c) a gillnetter misclassified as "other" and (d) a seiner misclassified as "other".

In the above statistical results, the average accuracy of the three types of fishing vessel classification results was 95.30%. Therefore, the classification results of the fishing vessels fishing in the northern SCS were reliable.

It can be seen from Figure 10 that, among the three types of fishing vessels, the largest number were gillnetters, followed by trawlers and seiners. During the three months, the number of fishing vessels was larger in April and September, and lower in June because of the closed season. Compared with April, the number of gillnetters in September increased by 13.53%, but the number of trawlers and seiners decreased by 38.85% and 18.48%, respectively. This was mainly because the water temperature decreased in September and the fish migrated to the sea areas with suitable temperatures. The fishing method of gillnetters intercepts a greater catch in the channels of the migratory paths, so their numbers increased.

Owing to the high average annual temperature and complex water systems, the fisheries in the SCS have high species diversity, and its rich fishery resources have become the focus of competition among neighboring countries. To explore the multinational fishing situation, we calculated the distribution and number of nations under which vessels were registered. Table 8 shows the total number of countries, the top eight countries with the larger number of fishing vessels and the corresponding number of fishing vessels. It can be seen from Table 8, in addition to Chinese fishing vessels, there were fishing vessels from countries such as Vietnam, Albania and Bhutan fishing in the northern SCS. Vietnamese fishing vessels made up the vast majority of non-Chinese fishing vessels, while the number of fishing vessels from other countries was small and probably unimportant. Especially in June, the number of Vietnamese fishing vessels increased significantly.

Table 8. Statistical results of countries of fishing vessels in the northern South China Sea.

Month	Number of Countries	The Top 8 Countries and the Number of Fishing Vessels
April	65	China (9306), Vietnam (252), Albania (42), Somali Democratic Republic (10), Bhutan (6), Korea (5), Lebanon (5), Fiji (5)
June	26	China (2210), Vietnam (741), Albania (12), United Kingdom (4), Bhutan (2), Malaysia (2), Christmas Island (2), Somali Democratic Republic (2)
September	66	China (9032), Vietnam (334), Albania (44), Somali Democratic Republic (18), Bhutan (10), Lebanon (8), Korea (5), Pitcairn Island (5)

According to the six classification criteria of fishing vessel length in the work of Kroodsma et al. [21], we calculated the length distribution of the Chinese and Vietnamese fishing vessels, as shown in Figure 12. For Chinese fishing vessels, the proportion of fishing vessels in each category was distributed homogeneously, but Vietnamese fishing vessels were mainly larger (24–30 m) vessels. On the basis of the classification results of fishing vessels in Figure 10, we counted the number of trawlers, gillnetters and seiners of China in April, June and September, as shown in Figure 13.

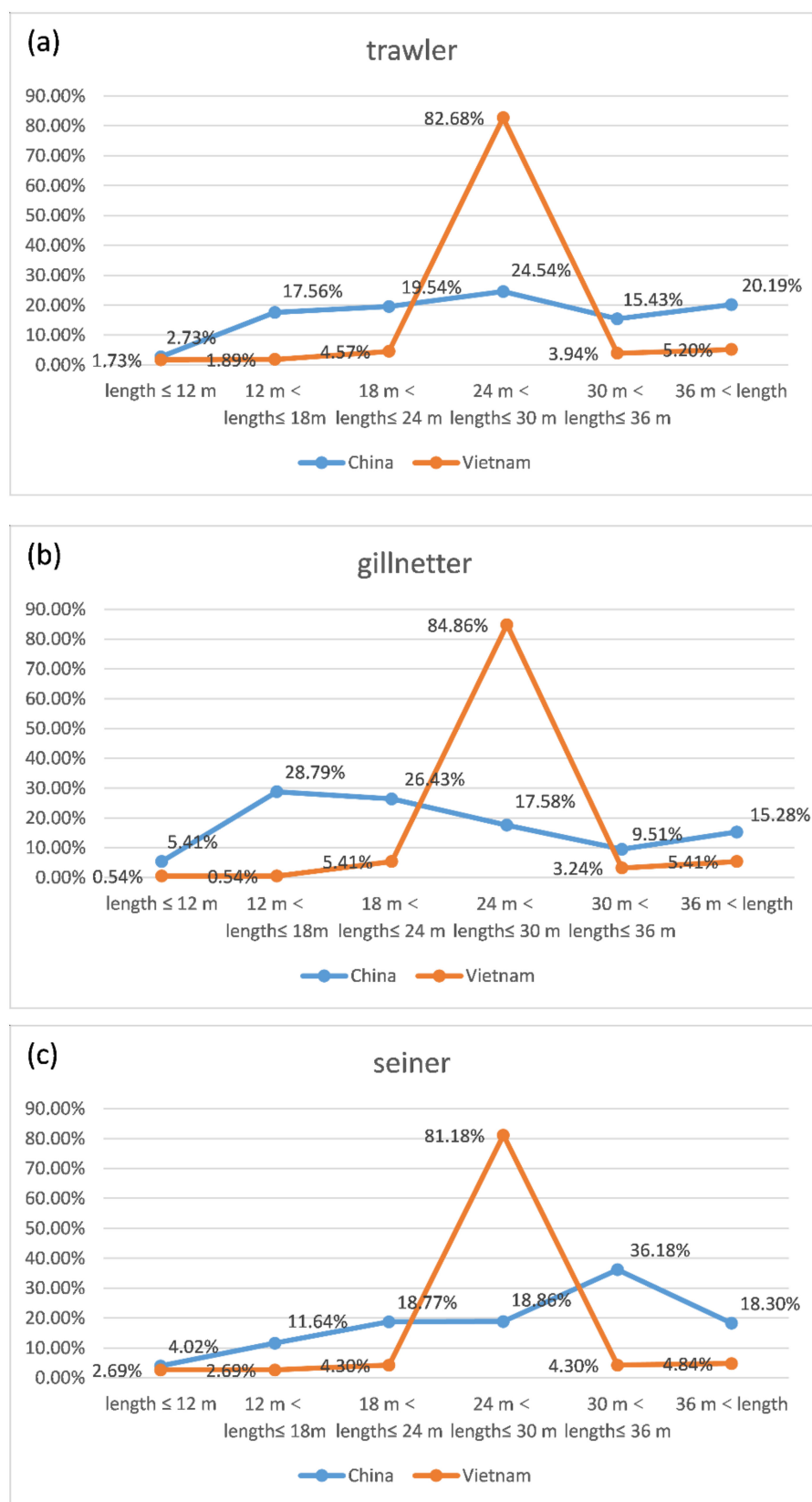


Figure 12. Classification of the length of Chinese and Vietnamese fishing vessels.

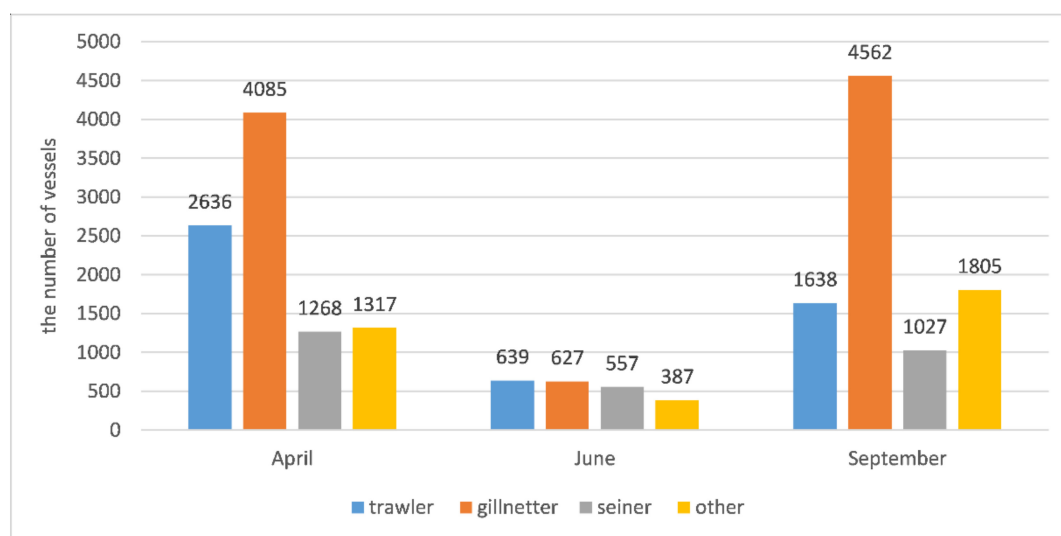


Figure 13. Number of trawlers, gillnetters and seiners for April, June and September for Chinese fishing vessels.

5. Analysis of the Seasonal Activity Pattern of Chinese Fishing Vessels in the Northern SCS

The number of Chinese fishing vessels accounted for more than 92% of the total, and therefore we analyzed the classification results of the Chinese fishing vessels in three aspects: the distribution and spatial variation of the activity density of fishing vessels, the temporal distribution and the fishing duration. The aim of the work was to explore the seasonal activity pattern of fishing vessels in the northern SCS.

5.1. Activity, Density Distribution and Spatial Variation of Fishing Vessels

On the basis of the classification results of Chinese fishing vessels in Figure 13, we first used the kernel density analysis method to map the activity density distribution of Chinese trawlers, gillnetters and seiners in the northern SCS in April, June and September, as shown in Figures 14–16. Overall, trawlers were distributed over a wide range, with hotspots mainly located off the coast of Guangdong Province, Guangxi Zhuang Autonomous Region and Hainan Island, and more evenly distributed elsewhere. Gillnetters were mainly distributed in the coastal areas of Guangdong and Guangxi and the northwestern waters of Hainan Island. Compared with the other types, the hotspots of seiners were mainly located off Hainan Island, but there were also more activity trajectories around the Paracel Islands, which were farther away from land. As the seawater temperature changes with the seasons, the fish migrate northward to waters with suitable temperature. The hotspot areas of the three types showed some evidence of a trend of moving north, especially for trawlers.

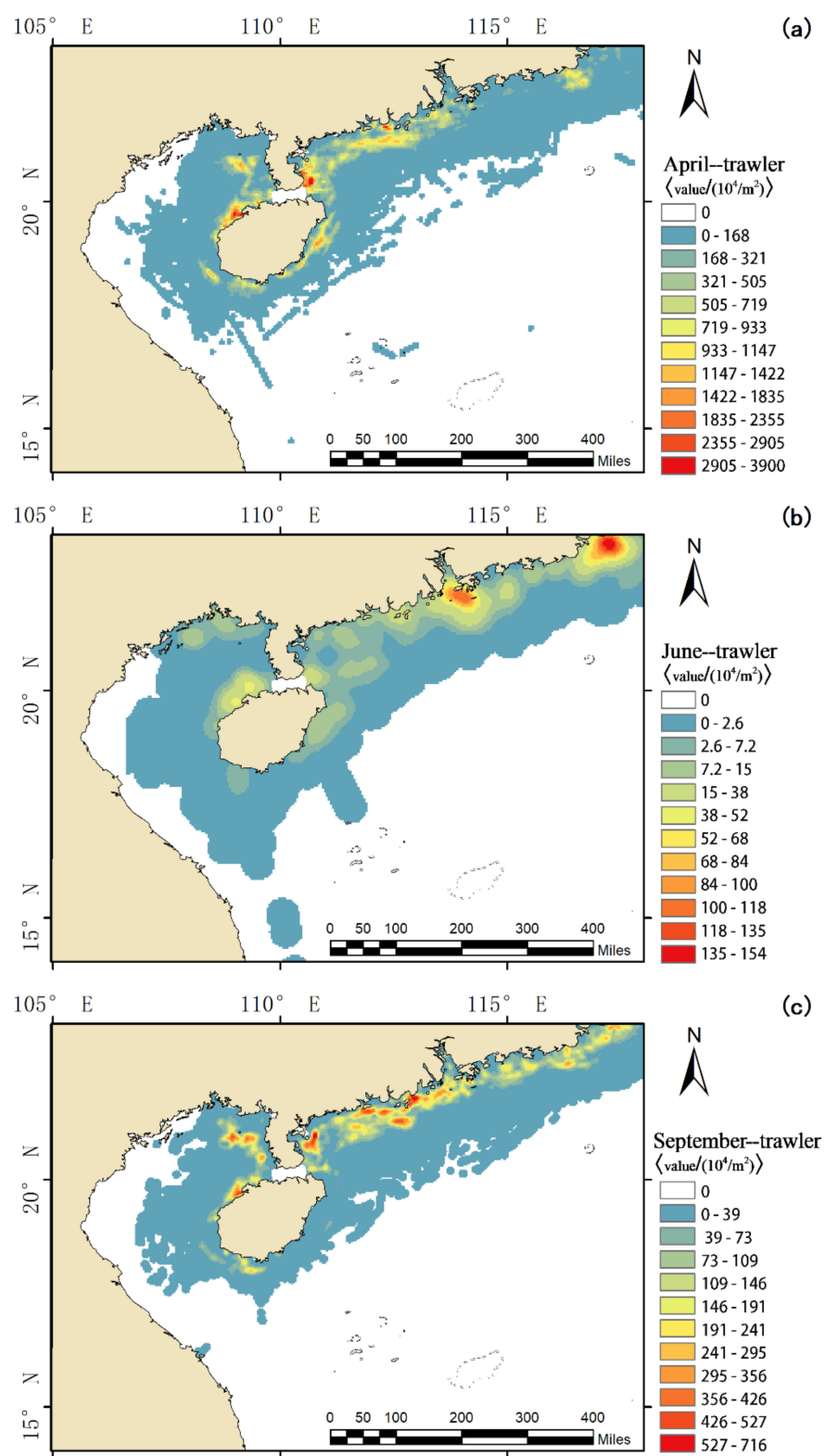


Figure 14. Nuclear density analysis chart of Chinese trawlers in April (a), June (b) and September (c).

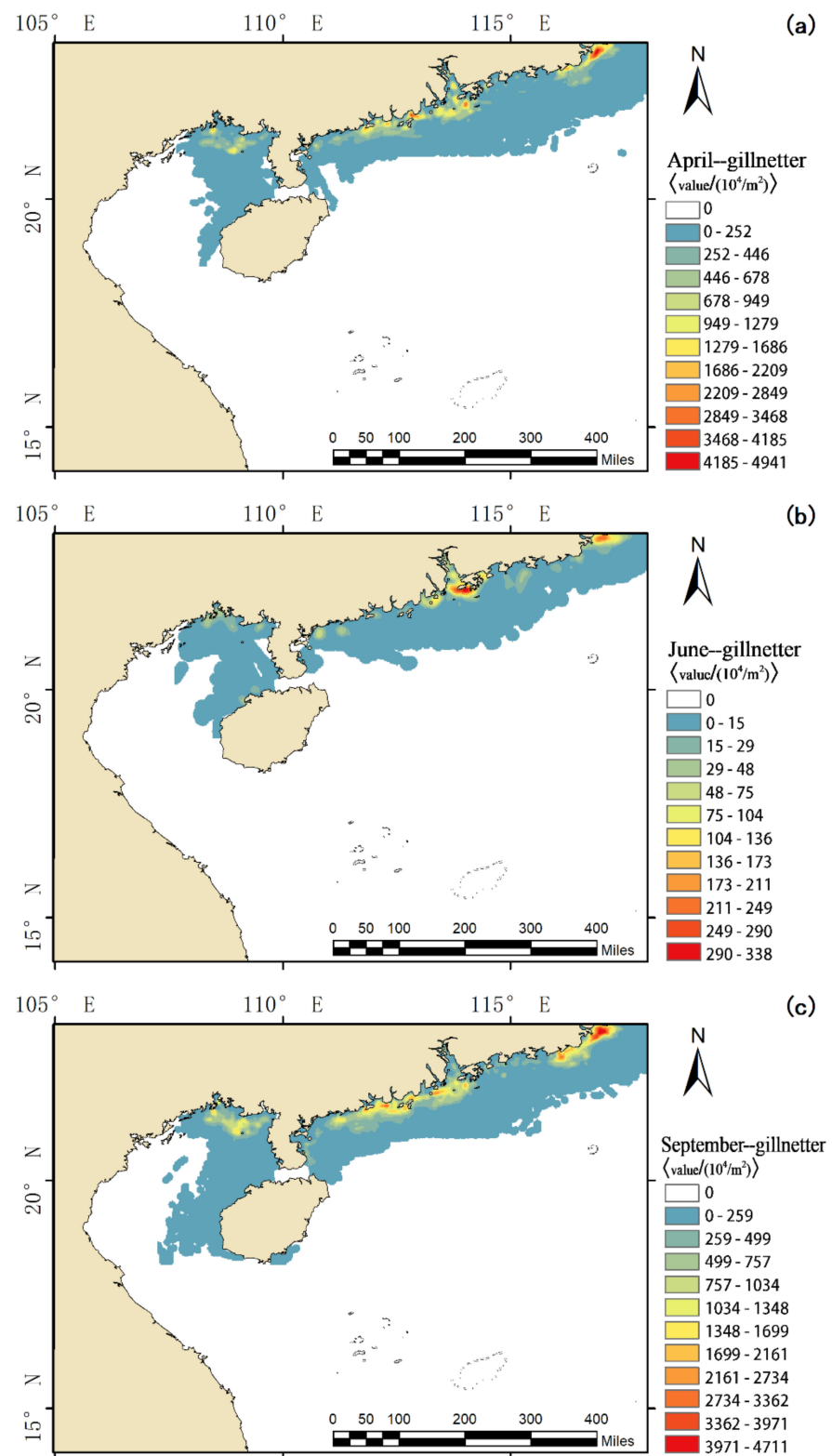


Figure 15. Nuclear density analysis chart of Chinese gillnetters in April (a), June (b) and September (c).

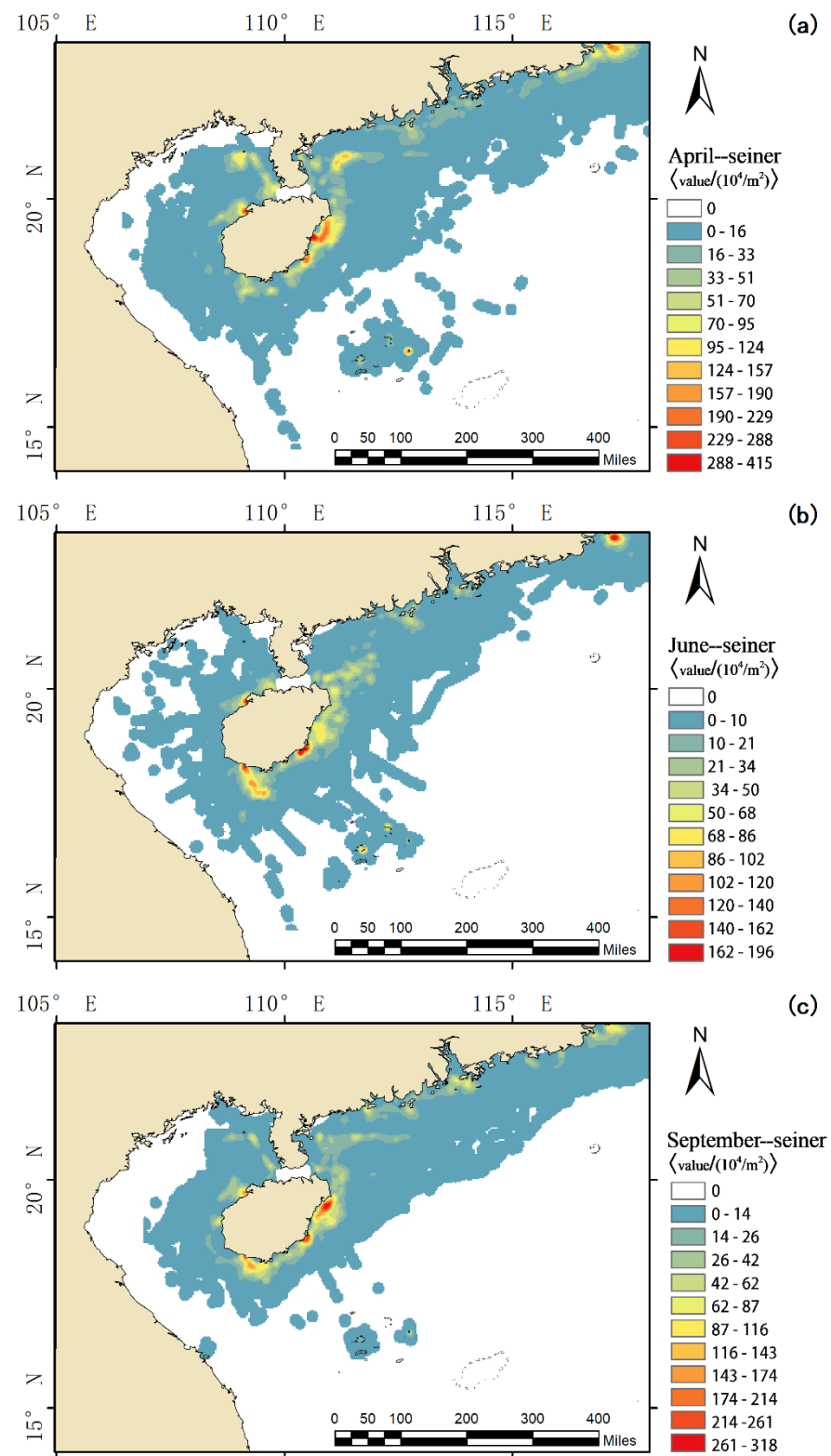


Figure 16. Nuclear density analysis chart of Chinese seiners in April (a), June (b) and September (c).

As can be seen from Figure 14, the active range and hotspot areas of trawlers varied significantly with the seasons. In April, the active hotspots of trawlers were mainly distributed in the coastal waters of Hainan Island and Guangdong Province, especially the northwestern seas of Hainan Island and the eastern seas of Leizhou Peninsula. In June, the area was closed to fishing, and the number of vessels and the scope of fishing decreased

sharply. The hotspots of fishing were mainly located in the inlet of the Pearl River and the inshore area of Shantou. After the closed season, compared with April, the hotspot areas of trawlers increased significantly in September, moving northward along the inshore coastal waters of the mainland.

Gillnetters were active in the inshore waters, and the distribution range was small, mainly located in the northern inshore waters of the Beibu Gulf and the inshore waters of Guangdong Province (Figure 15). The active range of gillnetters varied little with the seasons, except in June when the hotspot areas decreased because of the fishing ban. Compared with April, there was only a small increase in the number and range of gill net hotspot areas in September.

For seiners, hotspots were mainly distributed in the coastal waters of Hainan Island (Figure 16). However, compared with trawlers and gillnetters, their fishing range included a larger range of latitude and longitude, and there were also denser trajectory data in the waters near the Paracel Islands.

5.2. Temporal Distribution Characteristics of Fishing Vessels

In order to explore the temporal distribution of fishing vessels activities in the northern CSC, we separately analyzed the hourly and daily changes in the number of fishing vessels.

5.2.1. Hourly Statistics

Figure 17 shows the number of Chinese trawlers, gillnetters and seiners per hour operating in the northern SCS. There was no significant difference in the trend of the hourly number of the three types. Additionally, since the trawlers and gillnetters operate throughout the day, while the lighted seiners use the light trapping method to catch marine fish at night, the hourly number of the three types were less affected by day and night. However, it is obvious from Figure 17 that there are two peaks: at 6:00–7:00 and 18:00–19:00. This is attributed to the fact that fishermen are mostly concentrated in these two time periods when they go to sea and return to port. Furthermore, there were more fishing vessels in the daytime in April and September, which were the legal fishing months. Fishing vessels operating at night accounted for a large proportion of the total number in June and were more concentrated in the first half of the night.

5.2.2. Daily Statistics

The work of Kroodsma et al. [21] found that, globally, the fishing activities of Chinese fishing vessels were influenced by the Chinese New Year and the annual moratoria during the summer months. The Chinese New Year, Tomb Sweeping Day, the Dragon Boat Festival and the Mid-Autumn Festival are the four major traditional festivals in China. As the festivals approach, most Chinese go home to reunite with families. To further explore the impact of traditional Chinese festivals on Chinese vessels fishing in the northern SCS, we graphed the daily changes in the number of each fishing vessel type in each month, as shown in Figure 18.

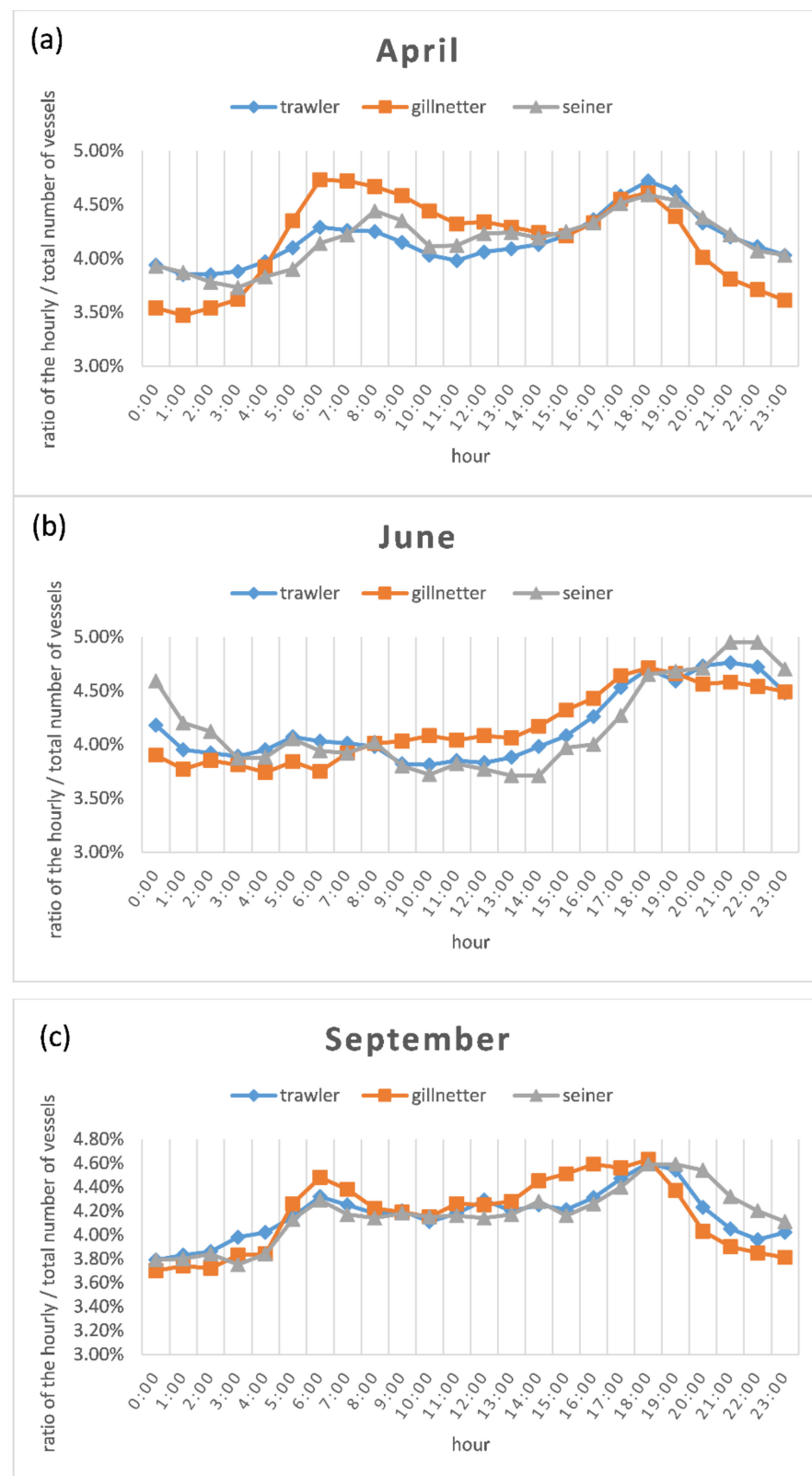


Figure 17. Statistics for Chinese fishing vessels per hour in April (a), June (b) and September (c).

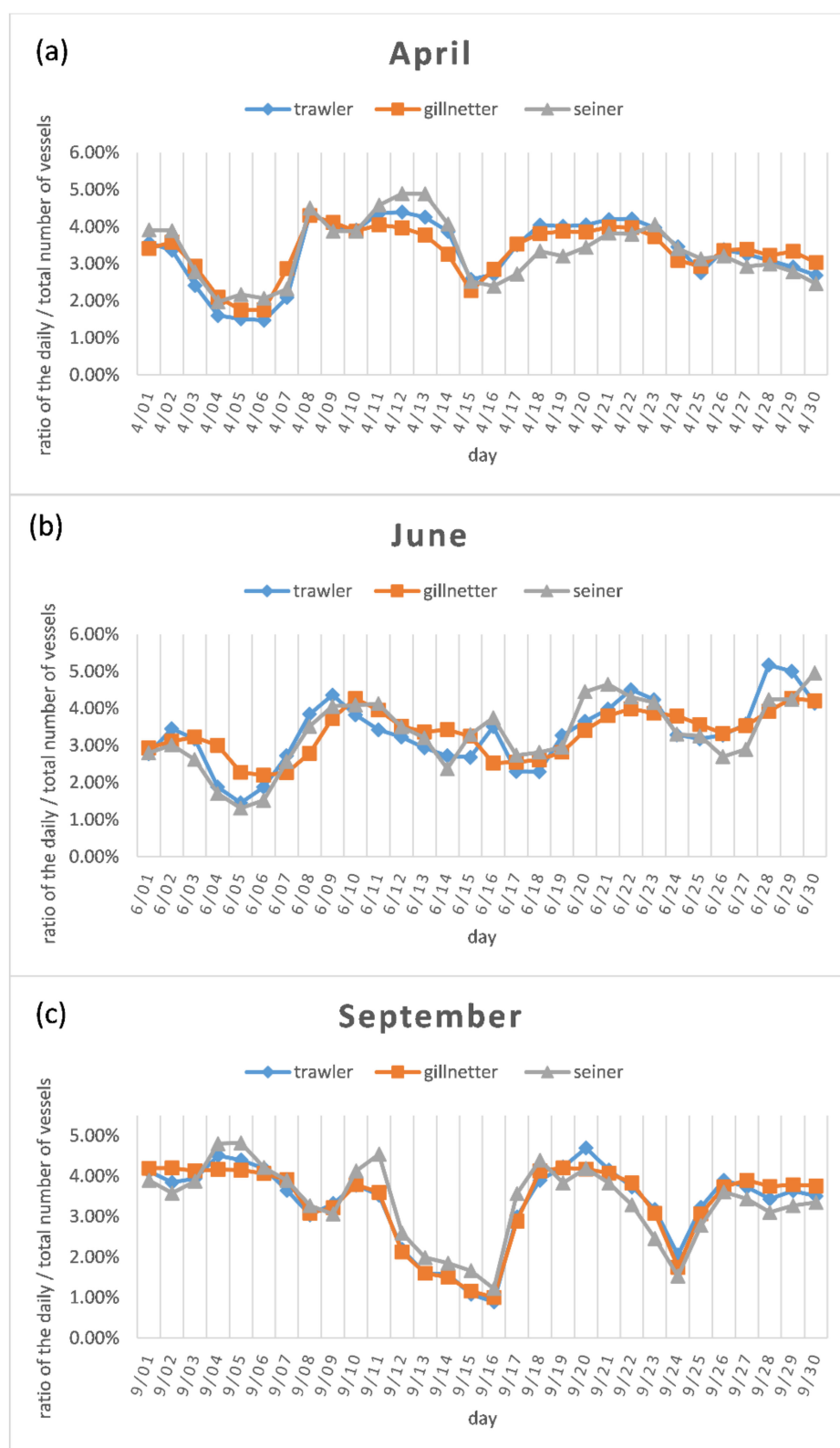


Figure 18. Statistics for Chinese fishing vessels per day in April (a), June (b) and September (c).

For Chinese fishing vessels (Figure 18), we found that in April and September, the change trend in the daily number was quite consistent for the trawlers, gillnetters and seiners, while there was a slight difference in June. The Chinese vessels showed little weekly variation but were influenced by traditional Chinese festivals. In China in 5 April 2018 was Tomb Sweeping Day (the holiday period was from 5 to 7 April), and June 18 was

the Dragon Boat Festival (the holiday time was from 16 to 18 June), and 24 September was the Mid-Autumn Festival (the holiday time was from 22 to 24 September). We counted the proportion of the average fishing vessels per day in the total number of the month during the holiday, and for three days before and after the holiday. It can be seen from the statistics that the number of all three types decreased very significantly during the festivals, as detailed in Table 9. Interestingly, from Figure 18 we found that the number of fishing vessels was lower during the festival and the day before and after the festival, which did not correspond to the legal three-day holiday. For example, during the Tomb Sweeping Day in April, the number of fishing vessels was low from April 4 to April 6, but the number on April 7 showed a significant increase, although it was still within the legal holiday. Meanwhile, the number of fishing vessels rose to a small peak before the holiday, and then began to show a significant downward trend. This showed that, owing to the increased demand for fish during the holiday, fishermen increased their working hours before the holidays to increase their catches and returned to the port for their vacation when the festivals were approaching.

Table 9. Average daily number of Chinese fishing vessels during and around traditional festivals.

Festival	Type	Pre-Holiday (%)	Holiday (%)	After-Holiday (%)
Tomb Sweeping Day	Trawler	3.11	1.53	3.50
	Gillnetter	3.31	1.87	3.76
	Seiner	3.53	2.07	3.56
The Dragon Boat Festival	Trawler	2.97	2.62	4.04
	Gillnetter	3.07	2.63	3.73
	Seiner	3.14	2.83	4.47
The Mid-Autumn Festival	Trawler	4.32	2.12	3.73
	Gillnetter	4.00	2.00	3.91
	Seiner	3.63	1.83	3.44

In addition to being affected by festivals, fishery production was significantly constrained by the weather and sea conditions. As shown in Figure 18a, the number of fishing vessels was low around April 15 and 25. A review of the weather records for that month showed that there were heavy rainstorms and strong winds in the local waters of Guangdong Province and Hainan Province during this period, and the severe weather prevented fishing boats in this area from going out to sea. In June and September, tropical storms “Ewiniar” and “Barijat” and Super Typhoon “Mangkhut” reached parts of the SCS on June 6, September 13 and September 16, respectively. The number of fishing vessels going to sea decreased sharply at these times. However, we found that although the number of fishing vessels operating at sea decreased significantly when the typhoon came, there was a sharp increase after the typhoon. Especially after super typhoon “Mangkhut”, the number of all three vessel types increased an approximately three-fold, far exceeding the growth rate after the holidays. This may be due to the disturbance of the ocean by slow-moving strong typhoons that transport nutrients to the upper layers of seawater [56], which then attract zooplankton and phytoplankton that feed on them. These are palatable bait for fish, thus triggering large numbers of fish to gather [57]. Experienced fishermen generally possess such knowledge, and they will even follow the path of the typhoon to get a larger catch.

5.3. Duration of Fishing Time

We also counted the cumulative and average fishing duration of each type of Chinese fishing vessels in each month (Table 10). According to the statistical results in Table 10, the average fishing duration of gillnetters was the longest, followed by trawlers and seiners. This difference was due to their different fishing methods. Gillnetters lay long nets which the fish pierce and become entangled. During this process, the fishing vessel waits alongside the net. Therefore, the fishing duration of gillnetters mainly depends on the waiting time. The fishing time of trawlers depends on the time that vessels tow the

gear for fishing. For seiners, when a school of fish is found, the vessel quickly surrounds it with nets, and then gradually closes the net to pick up the catch. This process takes less time than the other two methods.

Table 10. Cumulative and average fishing duration of trawlers, gillnetters and seiners in April, June and September.

Country	Type	April		June		September	
		Cumulative Operating Time/h	Average Operating Time/h	Cumulative Operating Time/h	Average Operating Time/h	Cumulative Operating Time/h	Average Operating Time/h
China	Trawler	493,221.96	187.11	60,372.72	94.48	316,134.00	193.00
	Gillnetter	862,098.40	211.04	67,458.93	107.59	1,027,006.56	225.12
	Seiner	90,902.92	71.69	38,466.42	69.06	75,103.20	73.20

For Chinese fishing vessels, the cumulative fishing duration of their three types varied significantly with the month and was greatly affected by the fishing moratorium. The variation in the average fishing duration of seiners was relatively small, while the average fishing duration of trawlers and gillnetters in June was significantly smaller, only approximately 50% of other months. To reduce the fishing time, trawlers may reduce the trawling distance and the number of trawls in a single operation, and gillnetters may reduce the waiting time to launch the nets. Therefore, because of the moratorium on fishing in June, illegal fishing vessels may reduce the possibility of detection by law enforcement officers by adopting the above measures. The fishing characteristics of seiners make it difficult to change the fishing duration significantly. The average fishing duration of three types increased slightly in September compared with April because the fishing intensity was greater just after the end of the moratorium.

6. Conclusions

The SCS is an important geographical and strategic location with rich fishery resources. It is essential to carry out comprehensive and systematic monitoring of the fishery resources in the SCS to improve the sustainability of the fishing industry. In contrast to previous studies that used fishery statistics provided by relevant departments, to our knowledge this work is the first to explore the spatial distribution characteristics and seasonal activities of different fishing vessel types in the northern SCS based on AIS data. The AIS dataset contained trajectory data for approximately 22,000 fishing vessels in the northern SCS in April, June and September 2018, with a total of around 150 million data points.

To achieve more accurate classification of fishing vessel types, which is traditionally based on the classification of changes in vessel speed, we mined three major categories of features in fishing vessel AIS tracking data (holistic characteristics of speed and heading, location changes and classification features in multiple states) to comprehensively explore the different features characterizing the main types of fishing vessels. After feature selection, based on the feature importance function provided by the LightGBM model, the following major findings were obtained. The features F_{31} – F_{36} and F_{55} – F_{60} extracted from the low-speed segment (0.6 knots, 2.5 knots) and the displacement segment (0.3 m, 1.3 m), which are highly relevant to the fishing activity of vessels, had higher importance. In contrast, the features F_{43} – F_{48} associated with the speed segment (6 knots, 12 knots) and F_{67} – F_{72} associated with the displacement segment (3.1 m, 6.2 m) had lower importance, indicating that there was no significant difference between various fishing vessels while sailing at high speed. The optimal 60-dimensional feature vector was finally determined after deleting features that contributed little to the classification.

To obtain the optimal LightGBM model, this work adopted a Bayesian optimization method to determine the optimal set of parameters as $\text{num_leaves} = 45$, $\text{learning_rate} = 0.05$, and $\text{n_estimators} = 20,000$. This combination of hyperparameters was used to build the optimal fishing vessel type identification model. To verify the performance of this model,

we performed five-fold cross-validation based on the sample set and compared the model with machine learning algorithms such as KNN, logistic regression, SVM and XGBoost. The results showed that the classification accuracy of the LightGBM model was high, reaching 95.68%, and the evaluation indexes such as precision, recall and F1-score outperformed other methods.

Finally, we analyzed the seasonal activities of Chinese fishing vessels in the northern SCS and drew the following conclusions.

- (1) Distribution and spatial variation of fishing vessels. As the year progressed, the hotspots of Chinese fishing vessels showed a trend of moving northward, especially for trawlers. The hotspots were mainly located in Guangdong Province, Guangxi Zhuang Autonomous Region and the coastal waters of Hainan Island.
- (2) Hourly variation of vessel flows. The hourly flow of Chinese fishing vessels showed less variation. In general, in April and September, the number of Chinese fishing vessels fishing at sea during daytime was slightly higher than that at night, and the opposite was true in June.
- (3) Daily variation of vessel flows. Fishing vessels in the northern SCS showed little weekly variation but were greatly affected by the Chinese traditional festivals such as Tomb Sweeping Day, the Dragon Boat Festival and the Mid-Autumn Festival. Chinese fishing vessels returned to the port for a vacation during traditional festivals, showing low fishing intensity.
- (4) Fishing duration of fishing vessels. The average fishing duration of trawlers and gillnetters was greatly affected by month, while the seiners were almost unaffected. In June, the cumulative fishing duration of Chinese fishing vessels dropped significantly, and the average fishing duration of trawlers and gillnetters was only 50% of that of other months.

Our work will provide information on fishery production in the northern SCS for marine spatial planners and managers as well as for the general public. It also lays the foundation for future scientific research on the operational behavior of fishing vessels of different types. In the future, our efficient and practical method of classifying fishing vessels can also be used in other waters, especially the Asian waters, to explore the distribution patterns of activities and provide guidance for the development and management of fisheries. In addition, we need to further develop the corresponding methods on how to use visible/radar satellite images to verify our results. At the same time, the results of our research can also provide assistance for the classification of fishing operations based on satellite images.

Author Contributions: X.Z. proposed the idea. Y.G., X.Z. and Z.L. developed the method. Y.G., J.Z., X.Z. and J.M. analyzed the results and wrote the text. G.L., M.B., and C.C. supervised the work. All authors commented the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Key R&D Program of China under Grant 2017YFC1405204, in part by the National Natural Science Foundation of China under Grant 61971455 and U2006207.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhang, K.; Guo, J.; Xu, Y.; Jiang, Y.; Fan, J.; Xu, S.; Chen, Z. Long-term variations in fish community structure under multiple stressors in a semi-closed marine ecosystem in the South China Sea. *Sci. Total Environ.* **2020**, *745*, 140892. [CrossRef] [PubMed]
2. Defusing the South China Sea Disputes: A Regional Blueprint. Available online: <https://www.csis.org/analysis/defusing-south-china-seadisputes> (accessed on 11 October 2018).
3. Li, B.; Xie, X.; Wei, X.; Tang, W. Ship detection and classification from optical remote sensing images: A survey—ScienceDirect. *Chin. J. Aeronaut.* **2020**, *34*, 145–163. [CrossRef]
4. Xu, F.; Liu, J.; Dong, C.; Wang, X. Ship Detection in Optical Remote Sensing Images Based on Wavelet Transform and Multi-Level False Alarm Identification. *Remote Sens.* **2017**, *9*, 985. [CrossRef]

5. Wang, H.-L.; Zhu, M.; Lin, C.-B.; Chen, D.-B. Ship detection in optical remote sensing image based on visual saliency and AdaBoost classifier. *Optoelectron. Lett.* **2017**, *13*, 151–155. [\[CrossRef\]](#)
6. Li, J.; Cai, Y.; Zhang, P.; Zhang, Q.; Jing, Z.; Wu, Q.; Qiu, Y.; Ma, S.; Chen, Z. Satellite observation of a newly developed light-fishing “hotspot” in the open South China Sea. *Remote Sens. Environ.* **2021**, *34*, 112312. [\[CrossRef\]](#)
7. Yamaguchi, T.; Asanuma, I.; Park, J.G.; Mackin, K.J.; Mittleman, J. Estimation of vessel traffic density from Suomi NPP VIIRS day/night band. In Proceedings of the OCEANS 2016 MTS/IEEE Monterey, Monterey, CA, USA, 19–23 September 2016; pp. 1–5. [\[CrossRef\]](#)
8. Ezequiel, C.; Carlos, A. Use of VIIRS DNB satellite images to detect jigger ships involved in the *Illex argentinus* fishery. *Remote Sens. Appl. Soc. Environ.* **2016**, *4*, 167–178.
9. Chaturvedi, S.K. Study of synthetic aperture radar and automatic identification system for ship target detection. *J. Ocean Eng. Sci.* **2019**, *4*, 173–182. [\[CrossRef\]](#)
10. Sandirasegaram, N.; Vachon, P.W. Validating targets detected by SAR ship detection engines. *Can. J. Remote Sens.* **2017**, *43*, 451–454. [\[CrossRef\]](#)
11. Gerard, M.; José, B.M.; Antonio, T. Operational Ship Monitoring System Based on Synthetic Aperture Radar Processing. *Remote Sens.* **2009**, *1*, 375–392.
12. Liu, F.; Zhao, F.; Yu, W.; Shi, L.; Wang, R. Ship detection and speed estimation based on azimuth scanning mode of synthetic aperture radar. *Int. Radar Sonar Navig.* **2012**, *6*, 425–431. [\[CrossRef\]](#)
13. Yang, G.; Li, B.; Ji, S.; Gao, F.; Xu, Q. Ship Detection from Optical Satellite Images Based on Sea Surface Analysis. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 641–645. [\[CrossRef\]](#)
14. Shi, Z.; Yu, X.; Jiang, Z.; Li, B. Ship Detection in High-Resolution Optical Imagery Based on Anomaly Detector and Local Shape Feature. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 4511–4523. [\[CrossRef\]](#)
15. Brekke, C.; Anfinson, S.N.; Larsen, Y. Subband Extraction Strategies in Ship Detection with the Subaperture Cross-Correlation Magnitude. *IEEE Geosci. Remote Sens. Lett.* **2012**, *10*, 786–790. [\[CrossRef\]](#)
16. Hou, X.; Xu, F. Inshore ship detection based on multi-aspect information in high-resolution SAR images. In Proceedings of the 2019 6th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR), Xiamen, China, 26–29 November 2019; pp. 1–4.
17. Jin, K.; Chen, Y.; Xu, B.; Yin, J.; Wang, X.; Yang, J. A Patch-to-Pixel Convolutional Neural Network for Small Ship Detection with PolSAR Images. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 6623–6638. [\[CrossRef\]](#)
18. Tixerant, M.L.; Guyader, D.L.; Gourmelon, F.; Queffelec, B. How can Automatic Identification System (AIS) data be used for maritime spatial planning? *Ocean Coast. Manag.* **2018**, *166*, 18–30. [\[CrossRef\]](#)
19. Tu, E.; Zhang, G.; Rachmawati, L.; Rajabally, E.; Huang, G.-B. Exploiting AIS Data for Intelligent Maritime Navigation: A Comprehensive Survey from Data to Methodology. *IEEE Trans. Intell. Transp. Syst.* **2017**, *19*, 1559–1582. [\[CrossRef\]](#)
20. Zhao, Z.; Ji, K.; Xing, X.; Zou, H.; Zhou, S. Ship Surveillance by Integration of Space-borne SAR and AIS—Further Research. *J. Navig.* **2013**, *67*, 295–309. [\[CrossRef\]](#)
21. Kroodsma, D.A.; Mayorga, J.; Hochberg, T.; Miller, N.A.; Boerder, K.; Ferretti, F.; Wilson, A.; Bergman, B.; White, T.D.; Block, B.A.; et al. Tracking the global footprint of fisheries. *Science* **2018**, *359*, 904–908. [\[CrossRef\]](#)
22. Souza, E.N.D.; Boerder, K.; Matwin, S.; Worm, B. Improving Fishing Pattern Detection from Satellite AIS Using Data Mining and Machine Learning. *PLoS ONE* **2016**, *11*, e0158248.
23. Huang, H.; Hong, F.; Liu, J.; Liu, C.; Feng, Y.; Guo, Z. FVID: Fishing Vessel Type Identification Based on VMS Trajectories. *J. Ocean Univ. China* **2019**, *18*, 403–412. [\[CrossRef\]](#)
24. Bertrand, S.; Diaz, E.; Lengaigne, M. Patterns in the spatial distribution of Peruvian anchovy (*Engraulis ringens*) revealed by spatially explicit fishing data. *Prog. Oceanogr.* **2008**, *79*, 379–389. [\[CrossRef\]](#)
25. Joo, R.; Bertrand, S.; Chaigneau, A.; Niquen, M. Optimization of an artificial neural network for identifying fishing set positions from VMS data: An example from the Peruvian anchovy purse seine fishery. *Ecol. Model.* **2011**, *222*, 1048–1059. [\[CrossRef\]](#)
26. Wang, Y. Estimating catches with automatic identification system (AIS) data: A case study of single otter trawl in Zhoushan fishing ground, China. *Iran. J. Fish. Sci.* **2016**, *15*, 75–90.
27. Wei, P.; Wang, X.; Ma, S.; Zhou, Y.; Huang, Y.; Su, Y.; Wu, Q. Analysis of current status of marine fishing in South China Sea. *J. Shanghai Ocean Univ.* **2019**, *28*, 976–982. [\[CrossRef\]](#)
28. Shelmerdine, R.L. Teasing out the detail: How our understanding of marine AIS data can better inform industries, developments, and planning. *Mar. Policy* **2015**, *54*, 17–25. [\[CrossRef\]](#)
29. Yang, D.; Wu, L.; Wang, S.; Jia, H.; Li, K.X. How big data enriches maritime research—A critical review of Automatic Identification System (AIS) data applications. *Transp. Rev.* **2019**, *39*, 755–773. [\[CrossRef\]](#)
30. Xiao, F.; Lighteringen, H.; Gulijk, C.V.; Ale, B. Comparison study on AIS data of ship traffic behavior. *Ocean Eng.* **2015**, *95*, 84–93. [\[CrossRef\]](#)
31. Russo, T.; D’Andrea, L.; Parisi, A.; Cataudella, S. VMSbase: An R-Package for VMS and Logbook Data Management and Analysis in Fisheries Ecology. *PLoS ONE* **2014**, *9*, e100195. [\[CrossRef\]](#)
32. Lee, J.; South, A.B.; Jennings, S. Developing reliable, repeatable, and accessible methods to provide high-resolution estimates of fishing-effort distributions from vessel monitoring system (VMS) data. *ICES J. Mar. Sci.* **2010**, *67*, 1260–1271. [\[CrossRef\]](#)
33. Marzuki, M.I.; Gaspar, P.; Garello, R.; Kerbaol, V. Fishing Gear Identification from Vessel-Monitoring-System-based Fishing Vessel Trajectories. *IEEE J. Ocean. Eng.* **2018**, *43*, 689–699. [\[CrossRef\]](#)

34. Ferrà, C.; Tasseti, A.N.; Grati, F.; Pellini, G.; Polidori, P.; Scarcella, G.; Fabi, G. Mapping change in bottom trawling activity in the Mediterranean Sea through AIS data. *Mar. Policy* **2018**, *94*, 275–281. [\[CrossRef\]](#)
35. Zeeberg, J.; Corten, A.; De Graaf, E. Bycatch and release of pelagic megafauna in industrial trawler fisheries off Northwest Africa. *Fish. Res.* **2006**, *78*, 186–195. [\[CrossRef\]](#)
36. Gupta, T.; Booth, H.; Arlidge, W.; Rao, C.; Manoharakrishnan, M.; Namboothri, N.; Shanker, K.; Milner-Gulland, E.J. Mitigation of Elasmobranch Bycatch in Trawlers: A Case Study in Indian Fisheries. *Front. Mar. Sci.* **2020**, *7*, 7. [\[CrossRef\]](#)
37. Libini, C.L.; Khan, S.A. Influence of lunar phases on fish landings by gillnetters and trawlers. *Indian J. Fish.* **2012**, *59*, 81–87.
38. Murata, M.; Nakamura, Y.; Kubota, S.; Hashiba, T.; Yamaguchi, H. Some biological aspects of neon flying squid *Ommastrephes bartrami* caught by Japanese squid gillnetters in the North Pacific. *Nippon Suisan Gakkaishi* **1988**, *54*, 1573–1581. [\[CrossRef\]](#)
39. Potier, M.; Elguero, E.; Petit, D. Vessel concentrations in the Javanese purse seine fishery: Structuration through spatial approach. *Aquat. Living Resour.* **2001**, *13*, 57–63. [\[CrossRef\]](#)
40. Zain, J. The Correlation of Fishingtrip Frequencies of Gillnetter on the Amount of Logistic Needed in Different Monsoon Seasons. *J. Ilmu Perair.* **2013**, *8*, 35–40.
41. Gaertner, D.; Pagavino, M.; Marcano, J. Influence of fishers' behaviour on the catchability of surface tuna schools in the Venezuelan purse-seiner fishery in the Caribbean Sea. *J. Can. Sci. Halieut. Aquat.* **1999**, *56*, 394–406.
42. Meng, Q.; Ke, G.; Wang, T.; Chen, W.; Ye, Q.; Ma, Z.; Liu, T. A Communication-Efficient Parallel Algorithm for Decision Tree. In Proceedings of the Neural Information Processing Systems 29 (NIPS 2016), Barcelona, Spain, 5–10 December 2016.
43. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In Proceedings of the Neural Information Processing Systems 30 (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
44. Ma, X.; Sha, J.; Niu, X. An Empirical Study on the Credit Rating of P2P Projects based on LightGBM Algorithm. *J. Quant. Tech. Econ.* **2018**, *35*, 144–160.
45. Tang, M.; Zhao, Q.; Ding, S.X.; Wu, H.; Li, L.; Long, W.; Huang, B. An Improved LightGBM Algorithm for Online Fault Detection of Wind Turbine Gearboxes. *Energies* **2020**, *13*, 807. [\[CrossRef\]](#)
46. Ju, Y.; Sun, G.; Chen, Q.; Zhang, M.; Zhu, H.; Rehman, M.U. A Model Combining Convolutional Neural Network and LightGBM Algorithm for Ultra-Short-Term Wind Power Forecasting. *IEEE Access* **2019**, *7*, 28309–28318. [\[CrossRef\]](#)
47. Shaker, B.; Yu, M.-S.; Song, J.S.; Ahn, S.; Ryu, J.Y.; Oh, K.-S.; Na, D. LightBBB: Computational prediction model of blood–brain-barrier penetration based on LightGBM. *Bioinformatics* **2020**, *2020*, 1–5. [\[CrossRef\]](#)
48. Zhou, K.-B.; Hu, Y.; Pan, H.; Kong, L.; Liu, J.; Huang, Z.; Chen, T. Fast prediction of reservoir permeability based on embedded feature selection and LightGBM using direct logging data. *Meas. Sci. Technol.* **2019**, *31*, 045101. [\[CrossRef\]](#)
49. Kumar, R.D.; Searleman, A.C.; Swamidass, S.J.; Griffith, O.L.; Bose, R. Statistically identifying tumor suppressors and oncogenes from pan-cancer genome-sequencing data. *Bioinformatics* **2015**, *31*, 3561–3568. [\[CrossRef\]](#)
50. Fu, G.; Yi, L.; Pan, J. Tuning model parameters in class-imbalanced learning with precision-recall curve. *Biom. J.* **2019**, *61*, 652–664. [\[CrossRef\]](#)
51. Cheng, Q.; Varshney, P.; Arora, M.K. Logistic Regression for Feature Selection and Soft Classification of Remote Sensing Data. *IEEE Geosci. Remote Sens. Lett.* **2006**, *3*, 491–494. [\[CrossRef\]](#)
52. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [\[CrossRef\]](#)
53. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
54. Cover, T.; Hart, P. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27. [\[CrossRef\]](#)
55. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased boosting with categorical features. In Proceedings of the Neural Information Processing Systems 31 (NIPS 2018), Montreal, QC, Canada, 2–8 December 2018; Volume 31, pp. 6639–6649.
56. Ye, H.; Tang, D.; Pan, G. The contribution of typhoon Megi on phytoplankton and fishery productivity in the South China Sea. *Ecol. Sci.* **2014**, *33*, 657–663.
57. Yu, J.; Tang, D.; Li, Y.; Huang, Z.; Chen, G. Increase in fish abundance during two typhoons in the South China Sea. *Adv. Space Res.* **2013**, *51*, 1734–1749. [\[CrossRef\]](#)