doi:10.6041/j.issn.1000-1298.2025.02.049

Fishing Effort Estimation of Trawlers Based on Vessel Monitoring System Data

LI Dan[1] LU Feng[1,2] XU Shuo[1,2] WANG Yu[1,2] XUE Muhan[1] NI Hanchen[1] FANG Hui[2,3] ZHANG Man[4] MA Zhenhua[5] CHEN Zuozhi[5] XU Jian[1]

(1. *Institute of Fisheries Engineering*, *Chinese Academy of Fishery Sciences*, *Beijing* 100141, *China*
2. *Laoshan Laboratory*, *Qingdao* 266237, *China*
3. *East China Sea Fisheries Research Institute*, *Chinese Academy of Fishery Sciences*, *Shanghai* 200090, *China*
4. *College of Information and Electrical Engineering*, *China Agricultural University*, *Beijing* 100083, *China*
5. *South China Sea Fisheries Research Institute*, *Chinese Academy of Fishery Sciences*, *Guangzhou* 510300, *China*)

Abstract: Estimating trawler fishing effort plays a critical role in characterizing marine fisheries activities, quantifying the ecological impact of trawling, and refining regulatory frameworks and policies. Understanding trawler fishing inputs offers crucial scientific data to support the sustainable management of offshore fishery resources in China. An XGBoost algorithm was introduced and optimized through Harris Hawks Optimization (HHO), to develop a model for identifying trawler fishing behaviour. The model demonstrated exceptional performance, achieving accuracy, sensitivity, specificity, and the Matthews correlation coefficient of 0.971 3, 0.980 6, 0.963 2, and 0.942 5, respectively. Using this model to detect fishing activities, the fishing effort of trawlers from Shandong Province in the sea area between 119°E to 124°E and 32°N to 40°N in 2021 was quantified. A heatmap depicting fishing effort, generated with a spatial resolution of 1/8°, revealed that fishing activities were predominantly concentrated in two regions: 121.1°E to 124°E, 35.7°N to 38.7°N, and 119.8°E to 122.8°E, 33.6°N to 35.4°N. This research can provide a foundation for quantitative evaluations of fishery resources, which can offer vital data to promote the sustainable development of marine capture fisheries.

Key words: trawler; vessel position data; machine learning; fishing effort
CLC number: TP391 **Document code**: A **Article ID**: 1000-1298(2025)02-0523-10 **OSID**:

0 Introduction

Marine fisheries are a vital component of the marine economy for coastal nations. In pursuit of advancing the marine economy, global fishing intensity has surged, driven by the expansion of fishing fleets and significant advancements in fishing technology and equipment[1]. This increase has imposed considerable pressure on the sustainable development of the marine ecological environment and biodiversity[2-5]. In this context, monitoring fishing efforts is essential for tracking changes in fishery resources, assessing marine ecological vulnerability, and supporting marine spatial planning[6]. Traditional methods for estimating fishing effort have largely depended on fishing logs and catch data, which are prone to various human biases and time delays. Issues such as inconsistent record-keeping, underreporting, misreporting, and the inherent time lag in data collection can significantly undermine the accuracy and timeliness of fishing effort estimates[7]. These challenges highlight the need for more robust and reliable data sources and methodologies to assess fishing pressure on marine resources with great precision. The vessel monitoring system (VMS), which tracks fishing vessels by providing dynamic data on vessel position, speed, and reporting time, offers a promising new data source for fisheries research[8]. VMS not only provides real-time insights into fishing vessel activities but also enables more accurate and timely assessments of fishing

Received date: 2024-09-10 Accepted date: 2024-11-29
Foundation items: Central Public-interest Scientific Institution Basal Research Fund (2024HY-ZC006 and 2023TD91); the National Key Research and Development Program of China (2024YFD2400505 and 2024YFD2400801); Laoshan Laboratory (LSKJ202203003)
First author: LI Dan, Assistant Researcher, Ph. D. E-mail: olivialdl@126.com
Corresponding author: XU Shuo, Researcher. E-mail: xush@cafs.ac.cn

pressure on marine ecosystems. This enhances the precision of fishing effort estimates and supports the development of effective management strategies aimed at ensuring the sustainability of fisheries.

The data obtained from VMS provides critical insights into the operational status of fishing vessels. By analyzing this data, it became possible to discern the fishing activities of vessels[9-15], facilitating the estimation of fishing effort[16]. The accuracy in identifying the operational status of fishing vessels directly impacted the precision of fishing effort estimations. Early fishing behaviour recognition algorithms primarily focused on extracting features such as vessel speed, heading, and operational time, applying thresholds to classify vessel activity[17-18]. However, the threshold-based approach was constrained by the limited features available, making it difficult to generalize across all fishing vessel types[8], and often led to overestimates of fishing effort[19].

Machine learning algorithms offer an advanced approach by exploring the nonlinear relationships between VMS data and fishing behaviour, and have become central to current research on fishing vessel status recognition[8]. For example, ERICO et al. [20] introduced a Hidden Markov model based on speed to identify fishing behaviours in trawlers, achieving an accuracy rate of 85%. This model established a nonlinear relationship between speed and fishing behaviour, enhancing recognition accuracy. However, it relied solely on a single feature input, which limited its overall accuracy. FAUSTINATO et al. [21] extracted geometric features from continuous trajectories and applied a Random Forest model to identify fishing behaviour, reaching an accuracy rate of 88% for trawler fishing behaviour recognition. While this study optimized the feature composition, it neglected spatial position data, which constrained further improvements in accuracy. LI et al. [22] employed XGBoost to identify trawling fishing behaviours in vessels from Liaoning Province and estimated fishing effort. DAVID et al. [23] constructed a comprehensive feature matrix by extracting extensive spatial and vessel dynamic data, using a deep convolutional neural network to recognize fishing behaviours, achieving an accuracy of 96%. However, deep neural networks are encumbered by challenges such as numerous parameters, complex parameter tuning, high computational costs, and long training times.

Leveraging Beidou vessel position data, a Harris Hawks Optimization extreme gradient boosting (HHO-XGBoost) algorithm was presented to develop a fishing behaviour recognition model for trawlers. The proposed model investigated the spatiotemporal distribution patterns of fishing efforts in Shandong Province in 2021, with the aim of offering an innovative approach to estimate trawler fishing effort.

1 Material and Methods

1.1 Data Preparation

The experimental vessel position data were collected from 3 600 trawlers in Shandong Province in 2021, covering the sea area between 119°E to 124°E and 32° N to 40° N. The dataset had a temporal resolution of 3 min and a spatial resolution of approximately 10 m. Each vessel position record included key information such as latitude, longitude, speed, transmission and reception timestamps, and fishing zone details.

In trawling operations, multiple nets were typically towed behind the vessel, which required deceleration and a constant speed to maintain uniform tension on the trawl nets. The duration of trawling sessions, typically ranging from 3 h to 5 h, was influenced by fish density. This study covered the full trawling process, which from net deployment to retrieval, and categorized the trawler status into two primary states: fishing and non-fishing. The fishing state encompassed net deployment, active trawling, and net retrieval, while the non-fishing state was further subdivided into anchored mooring and navigation.

Before annotation, the dataset underwent a thorough cleaning process to remove records with missing or erroneous values for longitude, latitude, speed or timestamps. The distance of each vessel position from the coastline was precisely calculated, and data were segmented based on accurate port entry and exit timestamps. Given the potential for signal fluctuations during satellite data reception, as well as occasional manual obstructions to transmission sources, trajectory segments were carefully segmented if the time elapsed between consecutive position reports exceeded 3 h.

Leveraging the expertise of fisheries professionals, the data from 12 trawling vessels based in Liaoning Province, covering the period from September 2019 to December 2021, were carefully calibrated. During this calibration process, the vessels' fishing status was annotated with "1" for fishing and "0" for non-fishing states. Key operational characteristics of the fishing vessels were then extracted for each record from the calibrated dataset. These parameters included the forward time interval, distance travelled, the shortest distance to China's coastline, the shortest distance to all ports and shelter anchorages in the Yellow Sea and Bohai Sea, theoretical speed, current speed, azimuth difference, rate of change in azimuth, time of day, and month. These critical parameters provided valuable insights into the fishing patterns, navigational behaviour, and operational efficiency of the trawling vessels throughout the study period.

1.2 Algorithm Overview

Beidou vessel position data was utilized to extract feature vectors, which were analyzed to classify the vessels' engagement in fishing activities. Following this classification, fishing effort was quantified. The overall system architecture for this process was presented in Fig. 1.

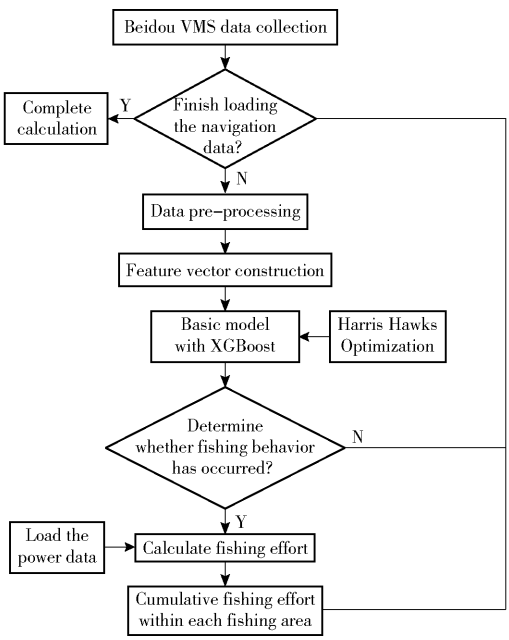

Fig. 1　Overall system architecture

The method outlined for calculating trawler fishing effort, based on Beidou vessel position data, comprised two key components. The first component involved constructing a fishing behaviour discrimination vector from the vessel position data, followed by the application of an XGBoost classifier, optimized via the Harris Hawks Optimization algorithm, to identify the occurrence of fishing activities. The second component calculated the fishing effort by integrating the classifier's output with both static and dynamic information extracted from the vessel position data.

1.3 Principle of HHO - XGBoost Based Fishing Behavior Classifier

Aimed to utilize XGBoost for the identification of trawler fishing behaviour. XGBoost is a gradient-boosted ensemble learning framework[24] that constructs models by iteratively incorporating the n-th weak learner, with the input being the residual of the $(n-1)$-th prediction. This process stacked multiple weak learners to progressively minimize residuals, ultimately approximating the true values. Decision trees were employed as weak learners. The objective function of the fishing vessel status recognition model based on XGBoost was expressed as follows:

$$Obj = L + \sum_{i=1}^{n} \Omega(f_j) \tag{1}$$

where L represented the loss function, $\sum_{i=1}^{n} \Omega(f_j)$ represented the complexity regularization term. In the t-th iteration, the predicted value of the model for the i-th sample was as follows:

$$\hat{y}_i^{(t)} = \sum_{k=1}^{t} f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i) \tag{2}$$

where t represented the iteration number; i denoted the sample index, with x_i representing the training vessel position data; $\hat{y}_i^{(t)}$ was the predicted result for sample i after the t-th iteration; $f_t(x_i)$ was the prediction result from the t-th weak learner; k indicated the index of the weak learner. Based on the above, the objective function can be expressed as:

$$Obj^{(t)} = \sum_{i=1}^{n} L(\hat{y}_i^{(t)}, y_i) + \sum_{i=1}^{n} \Omega(f_i) = \sum_{i=1}^{n} [y_i - (\hat{y}_i^{(t-1)} + f_t(x_i))]^2 + \sum_{i=1}^{n} \Omega(f_i) \tag{3}$$

A second-order Taylor series expansion on the objective function was defined in Equation (4):

$$Obj^{(t)} = \sum_{i=1}^{n} \left[(y_i - \hat{y}_i^{(t-1)})^2 + g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)^2 \right] + \Omega(f_t) + C \tag{4}$$

$$g_i = \partial_{\hat{y}_i^{(t-1)}} (y_i - \hat{y}_i^{(t-1)})^2$$
$$h_i = \partial^2_{\hat{y}_i^{(t-1)}} (y_i - \hat{y}_i^{(t-1)})^2$$

where C represented a constant. Given the above, the objective function can be expressed as:

$$Obj^{(t)} = \sum_{i=1}^{n} \left(g_i \omega_{q(x_i)} + \frac{1}{2} h_i \omega_{q(x_i)}^2 \right) + \gamma N + \frac{1}{2}\lambda \sum_{j=1}^{N} \omega_j^2 = \sum_{j=1}^{N} \left[G_J \omega_J + \frac{1}{2}(H_J + \lambda)\omega_j^2 \right] + \gamma N \tag{5}$$

where $\omega_{q(x_i)}$ represented the prediction made by the current weak learner for sample x_i when it fell into the corresponding node $q(x_i)$. Specifically, $\omega_{q(x_i)} = f_t(x_i) = \omega_j$, where ω_j was the weight associated with leaf node j. Given $G_J = \sum_{i \in I_J} g_i$; $G_J = \sum_{i \in I_J} g_i$, where I_J denoted the set of samples falling into leaf node J, the optimal weight ω_j for leaf node j can be computed to obtain the optimal solution $q(x_i)$. To find the optimal solution $q(x_i)$, the optimal weight ω_j for leaf node j can be calculated as:

$$\omega_j = -\frac{\sum_{i \in I_J} g_i}{\sum_{i \in I_J} h_i + \lambda} \tag{6}$$

The optimal solution for the corresponding objective function was calculated as follows:

$$Obj_{(q)}^{(t)} = -\frac{1}{2} \sum_{j=1}^{N} \frac{\left(\sum_{i \in I_J} g_i \right)^2}{\sum_{i \in I_J} h_i + \lambda} + \gamma N \tag{7}$$

Equation (7) served as a metric for assessing the performance of a weak learner's leaf nodes, with a higher score indicating improved algorithm performance in recognizing fishing vessel behaviour.

However, the application of XGBoost requires the adjustment of multiple parameters, and the selection of parameter values significantly influences its performance. The use of the HHO algorithm was proposed to optimize XGBoost. The HHO algorithm, introduced by Heidari, was inspired by the hunting behaviour of Harris Hawks[25] and involved several phases, including discovery, the transition from exploration to attack, and the attack phase.

Discover: Harris Hawks remained stationary in the desert, continuously observing their surroundings. They attempted to identify the most favourable points by making random exploratory suggestions:

$$X_{mat}(t+1) = \begin{cases} X_{mat_{rand}}(t) - r_1 |X_{mat_{rand}}(t) - 2r_2 X_{mat}(t)| & (q \geqslant 0.5) \\ X_{mat_{rabbit}}(t) - xX_{mat_m}(t) - r_3[AL - r_4(UL - AL)] & (q < 0.5) \end{cases} \tag{8}$$

where $X_{mat}(t+1)$ represented the position vector, $X_{mat_{rabbit}}(t)$ was the hunt position vector, $X_{mat_{rabbit}}(t)$ represented the hawk randomly selected hawk at the current population, $X_{mat_m}(t)$ was the average position of current hawk position and can be denoted as $1/N \sum_{i=1}^{N} \Omega X_{mat_i}(t)$, r_1, r_2, r_3, r_4, q denoted the random value in $[0,1]$. AL, UL denoted the lower and upper values, respectively.

Exploration to attack: Once the hunt was detected, the task of the hawk group was to reduce the energy expenditure of the hunt.

$$E = 2E_0 \left(-\frac{t}{T} + 1 \right) \tag{9}$$

where E was the energy of the escaped hunt, E_0 denoted the initial energy of the hunt, T was the maximum number of iterations. The attack was determined following the strategies of soft, hard, soft with progressive rapid dives, and hard with progressive rapid dives[25-26]. *Soft surround* represented a strategy employed by the Harris Hawks to reduce the energy required for hunting through the use of sudden attacks.

$$X_{mat}(t+1) = \Delta X_{mat}(t) - E|JX_{mat_{rabbit}}(t) - X_{mat}(t)| \quad (r \geqslant 0.5, E \geqslant 0.5) \tag{10}$$

$$\Delta X_{mat}(t) = X_{mat_{rabbit}} - X_{mat}(t)$$

where $\Delta X_{mat}(t)$ denoted the difference between the current position in the t-th iteration and the current position of the hunt, J was the jump strength. *Hard surround* was a situation where the energy of the hunt was considerably reduced.

$$X_{mat}(t+1) = X_{mat_{rabbit}}(t) - E|X_{mat}(t)| \quad (r \geqslant 0.5, |E| \leqslant 0.5) \tag{11}$$

Soft surround with progressive rapid dives indicated that the hunt possessed sufficient energy. The Harris Hawks seeked to deplete the hunt's energy through sudden, aggressive attacks.

$$Y = X_{mat_{rabbit}}(t) - E|\Delta X_{mat}(t)| - X_{mat}(t) \tag{12}$$

where Y denoted the next move. Unlike previous dives, which were assessed for their effectiveness, if the move was deemed inadequate, the hawk continued with sudden dives. The decision-making process employed a Lévy flight distribution structure to guide these actions.

$$Z = Y + S \times LF(D) \tag{13}$$

where D was the problem size, S denoted a random vector of size $1 \times D$, LF was the levy functions expresses as follows:

$$LF(x) = 0.01\,\frac{\mu\sigma}{|v|^{\frac{1}{\beta}}} \tag{14}$$

$$\sigma = \left[\frac{\Gamma(1+\beta)\sin\frac{\pi\beta}{2}}{\Gamma\left(\frac{1+\beta}{2}\right)\beta \times 2^{\frac{\beta-1}{2}}}\right]^{\frac{1}{\beta}}$$

where μ, v were random values inside $(0,1)$, β was default constant set to 1.5. Therefore, the strategy for updating the positions of hawks during the soft besiege phase can be formulated as follows:

$$X_{mat}(t+1) = \begin{cases} Y & (F(Y) < F(X_{mat}(t))) \\ Z & (F(Z) < F(X_{mat}(t))) \end{cases}$$
$$(r < 0.5, |E| \geqslant 0.5) \tag{15}$$

Hard surround with progressive rapid dives indicated that the hunt lacks sufficient energy. In response, the Harris Hawks initiated a more aggressive, forceful surround:

$$Y = X_{mat_{rabbit}}(t) - E|\Delta J X_{mat}(t)| - X_{mat}(t) \tag{16}$$

$$Z = Y + S \times LF(D) \tag{17}$$

$$X_{mat}(t+1) = \begin{cases} Y & (F(Y) < F(X_{mat}(t))) \\ Z & (F(Z) < F(X_{mat}(t))) \end{cases}$$
$$(r < 0.5, |E| < 0.5) \tag{18}$$

Noting that the coloured dots represented the location footprints of LF-based patterns in one trial, and only Y or Z would serve as the next location for the new iteration.

To optimize the process of searching for XGBoost hyperparameters, the HHO – XGBoost procedure was expressed as follows:

Step 1: Define the number and ranges of parameters to be optimized. The parameters to be optimized in XGBoost included the number of weak learners, learning rate, number of iterations, random undersampling ratio, column sampling ratio, and regularization parameters γ and λ. Set the upper and lower bound arrays "ub" and "lb" for each parameter according to their respective value ranges.

Step 2: Use the average accuracy from cross-validation of the XGBoost model as the fitness function for the HHO algorithm.

Step 3: Initialize the population of Harris hawks and compute the fitness values for each hawk.

Step 4: Check whether the maximum iteration count reached. If the current iteration count exceeded the maximum, proceed to Step 5. Otherwise, update the positions of the hawks and continue the optimization process by using the formulas in Equations (4) ~ (7).

Step 5: Output the results of the parameter optimization.

By following these steps, the trawler fishing behaviour recognition model based on HHO – XGBoost was obtained. The pseudo-code for HHO – XGBoost was provided in Algorithm 1.

Algorithm 1　Pseudo-code of HHO – XGBoost algorithm

```
Inputs: The population size N and maximum number of iterations T
 The upper and lower boundaries of the nine parameters of XGBoost
Outputs: The location of the rabbit and its fitness value.
 Initialize the positions of X_i (i = 1,2,…,n)
while iteration t ≤ T do
 Calculate the cross-validation score of XGBoost as the fitness
 Update bF, X_b
 Set X_mat_rabbit as the location of rabbit
 for each hawk X_i do
   Update the initial energy E_0 and jump strength J
   Calculate the E by equation (9)
   if |E| > 1 then
     Update the location vector using equation (8)
   if |E| < 1 then
     if (r ≥ 0.5, |E| ≥ 0.5) then
       Update location vector using equation (10)
     else if (r ≥ 0.5, |E| < 0.5) then
       Update location vector using equation (11)
     else if (r < 0.5, |E| ≥ 0.5) then
       Update location vector using equation (15)
     else if (r < 0.5, |E| < 0.5) then
       Update location vector using equation (18)
Return bF and X_b
```

1.4 Training and Testing Phase of Classifier

The fishing behaviour of trawlers based on HHO – XGBoost was implemented by using the Python programming language. During the training process, a dataset consisted of 165 860 position data points from 10 randomly selected vessels was assembled, ensuring an equal split between fishing and non-fishing behaviours to maintain a balanced training sample. Approximately one-fifth of this dataset was set aside as the test set, resulting in 12 547 points for model training and 44 314 points for testing. To obtain the optimal hyperparameters using HHO, a five-fold cross-validation approach was employed. Following the completion of the training process, external validation was conducted by using the 44 314 position points. Several machine learning algorithms were selected for comparison to evaluate the performance and advantages of the HHO – XGBoost algorithm.

1.5 Performance Evaluation of Classifier

To evaluate the performance of the classifier of trawler fishing behavior, four metrics[27–28] were utilized, i. e., specificity (SP), sensitivity (SN), accuracy (ACC), and the Matthews correlation coefficient (MCC).

1.6 Fishing Effort Calculation Method

According to the calculation method proposed by the Food and Agriculture Organization of the United Nations, fishing effort is typically expressed in terms of engine power and fishing operation days (kW·d)[29–30]. The method for calculating fishing effort was adopted from references [29, 31], with time measured in hours, and fishing effort quantified in kW·h. When a trawler was engaged in fishing activities, assuming the study area could be divided into grids, the formula for calculating fishing effort within each study grid was as follows:

$$E=\sum_{s=0}^{S}\sum_{i=0}^{I}\sum_{n=0}^{N}\left(T_{i,m,n,s}-T_{i,m-1,n,s}\right)W_iP_{i,m,n,s} \tag{19}$$

where m represented a position in a grid for a specific trawler, $T_{i,m}$ and $T_{i,m-1}$ denoted the time at consecutive points along the trajectory of trawler i, W_i denoted the engine power of the trawler i, $P_{i,m}$ represented the operational status of trawler i at position m at $T_{i,m}$, N indicated the number of positions for trawler i in the specific grid, I represented the number of trawler in the s-th grid, S denoted the number of grids, and E signified the total fishing effort within the study area.

2 Results

2.1 Evaluation of HHO – XGBoost Based Fishing Behavior Recognition

During the training of the HHO – XGBoost classifier, the HHO algorithm parameters were configured with a population size of 50 and an epoch of 20. With this configuration, HHO – XGBoost completed the training process and identified the optimal parameter combination, which included a learning rate of 0.096 8, 443 estimators, a subsample rate of 0.834 4, a colsample_bytree of 0.734 2, and a seed of 734. After determining these optimal parameters, model testing was conducted, and the final test results were summarized in Tab. 1.

Tab. 1 Comparison of precision rate of machine-learning classifiers

Methods	ACC	MCC	SN	SP
ELM	0.922 1	0.845 7	0.890 0	0.953 4
RF	0.971 1	0.942 2	0.981 8	0.961 9
LightGBM	0.970 0	0.939 8	0.976 5	0.964 3
XGBoost	0.970 9	0.941 8	0.979 6	0.963 4
HHO – XGBoost	0.971 3	0.942 5	0.980 6	0.963 2

The test set used in the experiment was collected from new trawlers that were not involved in the training phase. This dataset consisted of 20 489 fishing position points and 23 825 non-fishing position points, was specifically designed to evaluate the generalizability of the proposed model. The experimental results demonstrated that the HHO – XGBoost model outperformed other ensemble learning algorithms, including Extreme Learning Machine (ELM), Random Forest (RF), and LightGBM. Compared with ELM, RF, and LightGBM, HHO – XGBoost correctly identified 1,176, 7 and 57 additional samples in the test set, respectively. Furthermore, when compared with XGBoost trained with empirical parameters and XGBoost optimized by using the Genetic Algorithm (GA), HHO – XGBoost correctly identified 15 and 6 more samples, respectively. These results strongly supported the effectiveness of the proposed algorithm, through its innovative optimization process, significantly improved the performance in recognizing trawler fishing behaviour.

To further enhance the precision of fishing behaviour recognition, three strategieswere proposed. Firstly, feature enhancement was recommended. Considering that vessels often operated in fleets, with multiple trawlers working in pairs, integrating spatiotemporal information between vessels could refine the identification of fishing operation characteristics. Secondly, further algorithm optimization was suggested, potentially incorporating additional optimization techniques. However, the overall complexity of the model should be carefully considered. Lastly, expanding the data samples was essential for improving model performance. Continuously updating the dataset with data from various fishing vessels and extending the sample period would increase the model's accuracy and robustness.

2.2 Estimating Spatiotemporal Fishing Effort

The fishing behaviour of trawlers in Shandong Province was recognizedby using the HHO – XGBoost classifier, and the fishing effort for the entire year of 2021 was subsequently calculated, as illustrated in Fig. 2.

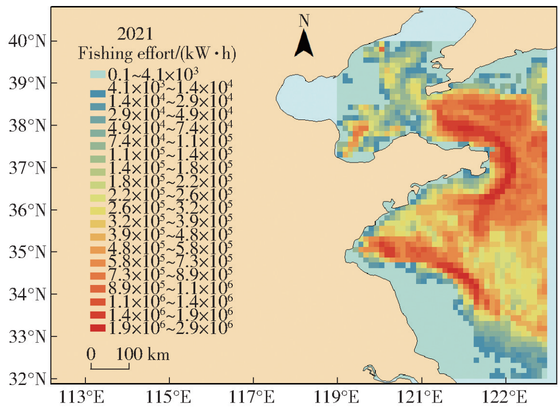

Fig. 2　Distribution of fishing effort of trawlers in Shandong Province in 2021

The analysis of Fig. 2 revealed that the majority of fishing effort was concentrated in two specific regions: ① The first region, spanning from 121.1°E to 124°E and from 35.7° N to 38.7° N, accounting for 387 963 366.6 kW·h of fishing effort, representing approximately 48.98% of the total recorded fishing effort within the surveyed area. This area is primarily located within the Yantai – Weihai and Shidao Fishing Grounds, is a key fishing area for trawlers in Shandong Province. ② The second area of focus, extending from 119.8°E to 122.8°E and from 33.6°N to 35.4°N, contributed roughly 17.46% of the total fishing effort recorded in the region, with concentrations in the Haizhou Bay Fishing Ground and the southwestern sectors of the Lianqingshi Fishing Ground.

To provide a clear understanding of the spatiotemporal distribution of fishing efforts by trawlers in Shandong Province, monthly hotspot maps illustrating the distribution of fishing efforts during the non-fishing closed season was presented, as shown in Fig. 3.

As illustrated in Fig. 3, in January 2021, the fishing effort of trawlers in Shandong Province was predominantly concentrated in two areas: 123° E to 124°E, 32.8°N to 36.7°N, and 121.4°E to 122.8°E, 34.1° N to 35.8° N, both of which were relatively distant from the coastline. By February, overall fishing effort showed a downward trend, with hotspots becoming more concentrated around 123°E to 124°E, 33.1°N to 36.8°N. This shift may be attributed to the Spring Festival, during which some fishermen returned home for the holiday, while others ventured further offshore for trawling.

In March, the distribution of fishing effort expanded, exhibiting a more diffuse pattern and an increase in overall fishing activity. The primary concentration of fishing effort shifted to 122.5° E to 124°E, 33° N to 35.9° N. By April, fishing effort remained stable compared with March but became more widely dispersed, gradually moving closer to the coastline. Key fishing areas were located at 122°E to 124°E, 33.5°N to 37.5°N, and 119.4°E to 120.8°E, 34.8°N to 35.7°N. This redistribution may be linked to the approaching fishing closed season, prompting trawlers to return to their home ports.

After September 2021, a notable increase in fishing activity occurred as a significant number of trawlers ventured offshore, resulting in fishing efforts that were two orders of magnitude higher than those observed during the first half of the year. Primary fishing activities were concentrated in the Yantai – Weihai Fishing Ground and Haizhou Bay Fishing Ground, located closer to the coast at 121.3°E to 124°E, 35.5°N to 38.7°N and 120°E to 122.5°E, 33.5°N to 35.5°N, respectively. Despite minimal fluctuations in fishing effort was between September and October, the spatial distribution became increasingly focused within these fishing grounds, with a significant uptick in

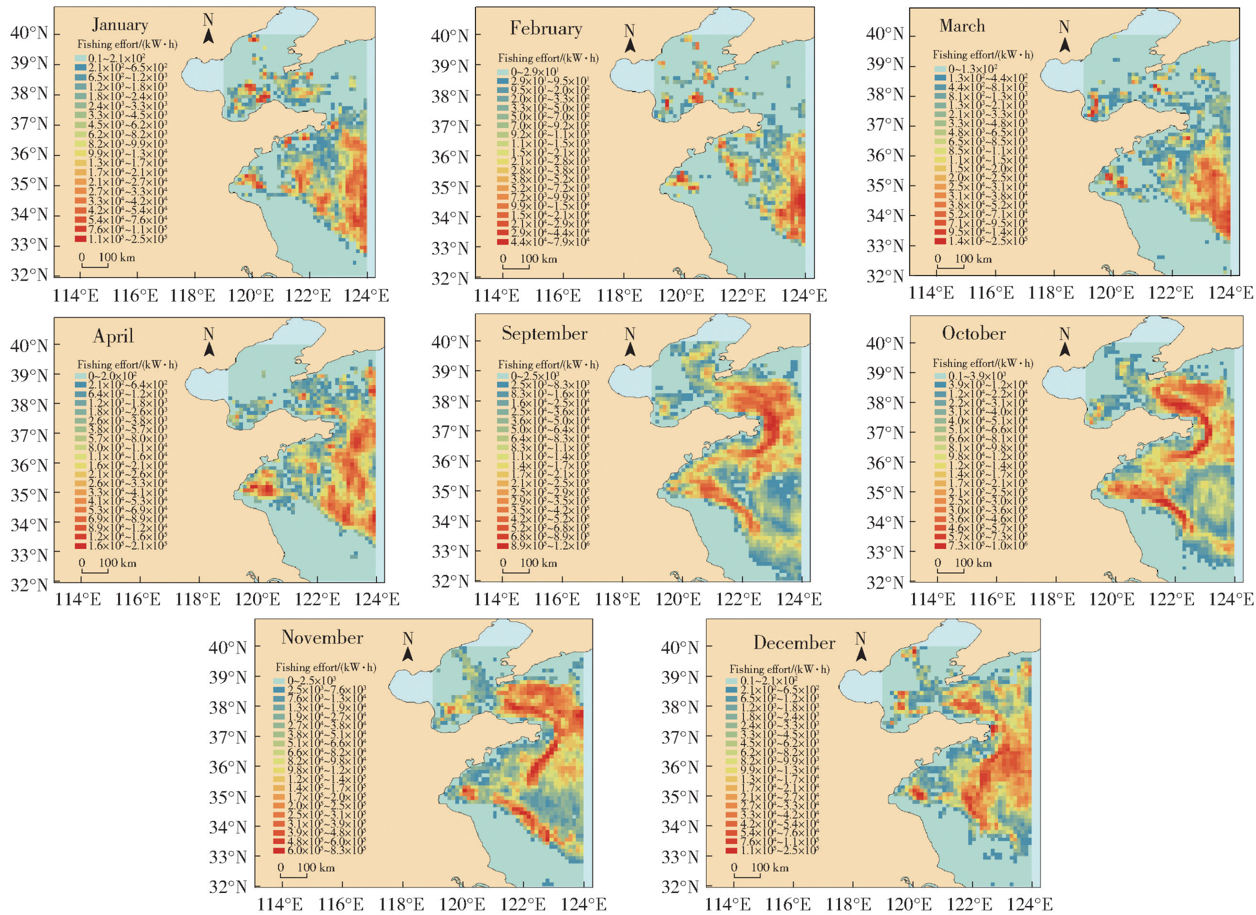

Fig. 3　Monthly distributions of fishing effort of trawlers in Shandong Province

fishing activities. Together, September and October accounted for the majority of the annual fishing effort.

In November, compared with October, the distribution of fishing efforts expanded further into deeper water, with a prominent concentration in the Yanwei Fishing Ground. By December, the fishing effort hotspots shifted even farther offshore, primarily concentrated in the central Yellow Sea at 121.5°E to 124°E, 34°N to 37°N, covering the Shidao Fishing Ground and Lianqingshi Fishing Ground. This shift indicated a growing trend of trawlers targeting deeper waters as the fishing season progressed.

3　Conclusion

(1) A novel algorithm, HHO – XGBoost, was introduced for the recognition of fishing behaviour in trawlers, and leveraging data from over 3 600 trawlers in 2021 was used to calculate fishing effort and estimate its spatiotemporal distribution. The proposed method demonstrated outstanding generalization capability, achieving high performance metrics in terms of accuracy (0.971 3), sensitivity (0.980 6), specificity (0.963 2), and Matthews correlation coefficient (0.942 5) in the test phase.

(2) The spatial distribution of fishing effort was predominantly concentrated in two key areas: 121.1°E to 124°E, 35.7°N to 38.7°N, and 119.8°E to 122.8°E, 33.6°N to 35.4°N. These areas were primarily located within the Yantai – Weihai Fishing Ground, Shidao Fishing Ground, Haizhou Bay Fishing Ground, and Lianqingshi Fishing Ground.

(3) To further enhance the accuracy of fishing behaviour recognition, future efforts should focus on feature enhancement, algorithm optimization, and data expansion. These strategies held substantial potential to further elevate the precision of the model.

参　考　文　献

[1]　FAO. The state of world fisheries and aquaculture[R]. Food and Agriculture Organization of the United Nations, 2022, Rome.

[2]　CHRISTOPHER C, LING C, STEFAN G, et al. The future of food from the sea[J]. Nature, 2020, 588: 95 – 100.

[3]　DANIEL S, JOHN T, VANESSA V, et al. New perspectives on an old fishing practice: scale, context and impacts of bottom

trawling[R/OL]. https://doi.org/10.1016/j.ecolecon.2020.106864.

[4] EISENSTEIN M. How cities are collaborating to help safeguard oceans[J/OL]. Nature Index, https://doi.org/10.1038/d41586-021-02407-8.

[5] LI Y, REN Y, CHEN Y. China fortifies marine protection areas against climate change[J]. Nature, 2019, 573(7774): 346.

[6] NAYLOR L, HARDY W, BUSCHMANN H, et al. A 20-year retrospective review of global aquaculture[J]. Nature, 2021, 591: 551-563.

[7] YANG Shenglong, SHI Huimin, FAN Wei, et al. Spatial distribution of squid fishing vessel operations in the southwest Atlantic Ocean and its relationship with environmental factors[J]. Journal of Fishery Sciences of China, 2022, 27(3): 365-376. (in Chinese)

[8] GUO Ganggang, FAN Wei, ZHANG Shengmao, et al. Advances in mining and application of vessel monitoring system data[J]. Marine Fisheries, 2016, 38(2): 217-224. (in Chinese).

[9] YANG D, LI X Y, ZHANG L Y. A novel vessel trajectory feature engineering for fishing vessel behavior identification[J]. Ocean Engineering, 2024, 310(1): 118677.

[10] GU Y M, HU Z H, ZHAO YC, et al. MFGTN: a multi-modal fast gated transformer for identifying single trawl marine fishing vessel[J]. Ocean Engineering, 2024, 303: 117711.

[11] NGUYEN D, VADAINE R, HAJDUCH G, et al. GeoTrackNet-a maritime anomaly detector using probabilistic neural network representation of AIS tracks and a contrario detection[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(6): 5655-5667.

[12] SÁNCHEZ P D, AMIGO D, GARCÍA J, et al. Architecture for trajectory-based fishing ship classification with AIS data[J]. Sensors, 2020, 20(13): 3782.

[13] SINGH S K, HEYMANN F. Machine learning-assisted anomaly detection in maritime navigation using AIS data[C]//2020 IEEE/ION Position, Location and Navigation Symposium (PLANS). IEEE, 2020: 832-838.

[14] JOO R, BERTRAND S, CHAIGNEAU A, et al. Optimization of an artificial neural network for identifying fishing set positions from VMS data: an example from the Peruvian anchovy purse seine fishery[J]. Ecological modelling, 2011, 222(4): 1048-1059.

[15] RUSSO T, PARISI A, PRORGI M, et al. When behaviour reveals activity: assigning fishing effort to métiers based on VMS data using artificial neural networks[J]. Fisheries Research, 2011, 111(1-2): 53-64.

[16] MILLS C M, TOWNSEND S E, JENNINGS S, et al. Estimating high resolution trawl fishing effort from satellite-based vessel monitoring system data[J]. ICES Journal of Marine Science, 2007, 64(2): 248-255.

[17] ZHANG Shengmao, YANG Shenglong, DAI Yang, et al. Algorithm of fishing effort extraction in trawling based on Beidou vessel monitoring system data[J]. Journal of Fisheries of China, 2014, 38(8): 1190-1199. (in Chinese)

[18] HAN Z, SHENGLONG Y, WEI F, et al. Spatial analysis of the fishing behaviour of tuna purse seiners in the western and central Pacific based on vessel trajectory data[J]. Journal of Marine Science and Engineering, 2021, 9(3): 322.

[19] BERTRAND S, BURGOS J M, GERLOTTO F, et al. Lévy trajectories of Peruvian purse-seiners as an indicator of the spatial distribution of anchovy (*Engraulis ringens*)[J]. Ices Journal of Marine Science, 2005, 62(3): 477-482.

[20] ERICO N S, KRISTINA B, STAN M, et al. Improving fishing pattern detection from satellite AIS using data mining and machine learning[J]. Plos One, 2016, 11(7): e0158248.

[21] FAUSTINATO B, MARIE P E, JÉRÔME G S, et al. Estimating fishing effort in small-scale fisheries using GPS tracking data and random forests[J]. Ecological Indicators, 2021, 123: 107321.

[22] LI D, LU F, XU S, et al. Research on fishing effort algorithm of trawling based on Beidou satellite navigation data[J]. Marine Fisheries, 2023, 45(4): 472-482. (in Chinese)

[23] DAVID A K, JUAN K, TIMOTHY H, et al. Tracking the global footprint of fisheries[J]. Science, 2018, 359: 904-908.

[24] TIANQI C, CARLOS G. XGBoost: a scalable tree boosting system[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016: 785-794.

[25] HEIDARI AA, MIRJALILI S, FARIS H, et al. Harris hawks optimization: algorithm and applications[J]. Future Generation Comp. Sys., 2019, 97: 849-872.

[26] BOZKURT C, ATES A, FIRAT M, et al. A novel strategic water loss management model and its optimization with Harris hawk algorithm[J]. Water Resour Manage, 2024, 38: 1543-1561.

[27] CHAWLA N V, BOWYER K W, HALL L O, et al. SMOTE: synthetic minority over-sampling technique[J]. Journal of Artificial Intelligence Research, 2002, 16: 321-357.

[28] LI D, LU F, XU S, et al. Estimating spatiotemporal fishing effort of trawlers with vessel-monitoring system data: a case study of the sea area of the Bohai Sea and the Yellow Sea, China[J]. Journal of Marine Science and Engineering, 2024, 12(1): 64.

[29] ZHANG S M, CUI X S, WU Y M, et al. Analyzing space-time characteristics of Xiangshan trawling based on Beidou vessel monitoring system data[J]. Transactions of the CASE, 2015, 31(7): 151-156. (in Chinese)

[30] FAO. Definition of some key terms[EB/OL]. Fisheries and Aquaculture Department,1997. http://www.fao.org/docrep/003/w4230e/w4230e09.htm.

[31] YANG S L,ZHANG S M,ZHOU W F,et al. Calculating the fishing effort of longline fishing vessel in the western and central pacific ocean using AIS[J]. Transactions of the CSAE,2020,36(3):198-203. (in Chinese)

基于船舶监测系统数据的拖网渔船捕捞努力量估算

李 丹[1] 鲁 峰[1,2] 徐 硕[1,2] 王 宇[1,2] 薛沐涵[1] 倪翰晨[1] 方 辉[2,3]
张 漫[4] 马振华[5] 陈作志[5] 许 建[1]

(1. 中国水产科学研究院渔业工程研究所, 北京 100141; 2. 崂山实验室, 青岛 266237;
3. 中国水产科学研究院东海水产研究所, 上海 200090; 4. 中国农业大学信息与电气工程学院, 北京 100083;
5. 中国水产科学研究院南海水产研究所, 广州 510300)

摘要: 拖网渔船捕捞努力量的估算对于描述海洋渔业活动、量化拖网作业对海洋造成的生态压力以及修订渔业法规和政策具有重要意义。明确拖网渔船的捕捞投入可为中国近海渔业资源的可持续发展提供科学数据支持。本研究提出了一种基于 Harris Hawks Optimization(HHO)优化的 XGBoost 算法,用于构建拖网渔船捕捞行为识别模型。结果表明,该模型准确率、灵敏度、特异度和马修斯相关系数分别为 0.971 3、0.980 6、0.963 2 和 0.942 5。利用该模型识别拖网渔船的捕捞行为并计算了 2021 年在 119°E~124°E、32°N~40°N 海域内山东省拖网渔船的捕捞努力量。以空间精度 1/8°生成了捕捞努力量热力图,计算结果揭示了捕捞活动的空间分布主要集中在 2 个关键区域:121.1°E~124°E、35.7°N~38.7°N 和 119.8°E~122.8°E、33.6°N~35.4°N。本研究可为渔业资源的定量评估奠定基础,为海洋捕捞渔业的可持续发展提供必要数据。

关键词: 拖网渔船; 船位数据; 机器学习; 捕捞努力量

基金项目: 中国水产科学研究院中央级公益性科研院所基本科研业务费专项(2024HY-ZC006、2023TD91)、国家重点研发计划项目(2024YFD2400505、2024YFD2400801)和崂山实验室科技创新项目(LSKJ202203003)