



| Research Article

Estimation of Potential Fishing Grounds Using the Random Forest Machine Learning Algorithm with MODIS Satellite Data

Hasbi Husaini*¹, Ahmad Mujab Arba², Muhammad Rajaskana Syahputra³

¹Marine and Fisheries Service, Provinsi Sumatera Utara, Indonesia

²Biology Study Program, Faculty of Mathematics and Science, Universitas Sumatera Utara, Medan, Indonesia

³Department of Physics, Faculty of Mathematics and Science, Universitas Sumatera Utara, Medan, Indonesia

Correspondence Email: hasbi2000husaini@gmail.com

Received: September 22, 2025 | Revision: March 02, 2026 | Accepted: March 09, 2026

Abstract: North Sumatra Province, located within Indonesia's Fisheries Management Area (WPP) 571, has substantial marine capture fisheries potential along its eastern coastline. However, increasing fishing intensity has led to higher fuel consumption and carbon emissions from fishing vessels, emphasizing the need for spatially efficient fishing strategies. This study aims to estimate and validate potential fishing zones (PFZ) using a Random Forest machine learning algorithm integrated with MODIS-Aqua satellite data to enhance operational efficiency and support emission mitigation. Four oceanographic parameters were employed as predictor variables: Sea Surface Temperature (SST), chlorophyll-a concentration (Chl-a), normalized fluorescence line height (nFLH), and particulate organic carbon (POC), covering the period 2013–2020. Ecological threshold values were applied to generate PFZ and Non-PFZ labels, and the dataset was divided into 80% training and 20% testing subsets using stratified random sampling. Model performance was evaluated using confusion matrix analysis, precision, recall, F1-score, and Cohen's Kappa coefficient. Results indicate that all variables were normally distributed and significantly correlated ($R > 0.5$; $R^2 > 0.7$). The Random Forest model achieved consistently high predictive accuracy, exceeding 90% across all evaluated years. Compared to a conventional threshold-based method, the machine learning approach produced more stable and spatially consistent PFZ delineations by capturing nonlinear environmental interactions. This study demonstrates that integrating satellite oceanography and machine learning provides a robust, data-driven framework for improving fishing efficiency while contributing to carbon emission reduction in capture fisheries management.

Keywords: Random Forest; Potential Fishing Zone; MODIS-Aqua; Oceanography; Carbon Emissions

1. INTRODUCTION

North Sumatra Province is one of Indonesia's strategic coastal regions, with a coastline extending approximately 300.82 km, contributing significantly to national marine capture fisheries production. According to Badan Pusat Statistik (BPS), total marine capture fisheries production in North Sumatra in 2024 reached approximately 371,603 thousand tons, with an economic value of around 13.099 trillion rupiah. This achievement reflects the high intensity of fisheries resource utilization in the region.

However, the intensification of fishing activities has also exerted ecological pressures on coastal and marine ecosystems. Activities such as increased vessel traffic, overfishing, and the use of chemicals in fishing operations can degrade water quality and disrupt marine ecological balance (Burch et al., 2025). If not managed

sustainably, these anthropogenic pressures may threaten long-term fisheries productivity and accelerate coastal environmental degradation.

Previous studies have examined aquatic ecosystem issues and their implications for environmental sustainability. A silvofishery model has been implemented in the coastal region of Indramayu to enhance blue carbon and support ecosystem-based climate change mitigation (Tansy & Arif, 2026). Other research has identified significant ecological pressures in coastal areas due to human activities, leading to reduced ecosystem functions (Dalimunte et al., 2025). Additionally, satellite-based remote sensing approaches have been widely applied to analyze the dynamics of water quality parameters, such as Total Suspended Solids (TSS), in estuarine environments (Damayani et al., 2025). Nevertheless, these studies primarily focus on coastal ecology and water quality and have not explicitly integrated fisheries productivity analysis with spatial fishing efficiency as a strategy to mitigate carbon emissions from the capture fisheries sector.

One of the less visible yet significant environmental impacts of capture fisheries is the increase in carbon emissions resulting from fishing vessel traffic, which directly contributes to greenhouse gas emissions. Globally, the fishing fleet consumed approximately 40 billion liters of fuel in 2011, generating ± 179 million tons of CO₂-equivalent emissions, accounting for around 4% of total emissions from global food production (Parker et al., 2018a). A long-term study covering the period 1950–2016 reported that CO₂ emissions from industrial fisheries increased to 159 million tons, while small-scale fisheries contributed approximately 48 million tons in 2016, with an average emission intensity of about 2 tons of CO₂ / ton of fish caught (Greer et al., 2019). In many countries, marine fisheries are recognized as carbon-intensive sectors that remain insufficiently integrated into climate mitigation policies, despite rising emissions driven by fleet expansion and longer travel distances.

Empirical studies further indicate that fuel consumption intensity and carbon emissions in fisheries vary substantially depending on vessel type and size. For example, handline vessels operating at PPP Pondokdadap consume approximately 0.66 kg of fuel per kilogram of landed fish, generating approximately 0.002 kg of CO₂ per kilogram of catch. Small-scale vessels at PPN Tanjung Pandan require about 0.83 kg of fuel per kilogram of fish, producing approximately 0.001 kg of CO₂ per kilogram of catch (Rahman et al., 2024). Larger vessels demonstrate significantly higher emission levels. Longline vessels of 51–100 GT operating at PPS Nizam Zachman emit exhaust gases, including CO, NO_x, SO_x, and particulate matter (PM), amounting to hundreds to thousands of kilograms per trip (Kurniawati et al., 2021). Meanwhile, tuna purse seine vessels of 10–60 GT produce substantially higher CO₂ emissions, ranging from 2.88 to 9.9 kg per kilogram of fish, depending on vessel size and operational capacity (Salsabila et al., 2024). These findings confirm that increased fleet capacity and operational intensity are directly proportional to the carbon footprint generated, highlighting the need for spatial-efficiency approaches to determine fishing grounds to reduce fuel consumption and emissions.

Despite extensive research on coastal ecosystem degradation, water quality dynamics, and fisheries-related emissions, a significant research gap remains in integrating oceanographic-based spatial analysis of potential fishing grounds with explicit objectives of reducing carbon emissions through improved fleet operational efficiency. Most previous studies have mapped potential fishing zones based on environmental parameters or evaluated ecosystem conditions separately, without directly linking spatial fishing optimization to fuel reduction and carbon mitigation as sustainability indicators.

Therefore, this study is important as it integrates satellite-based oceanographic approaches with machine learning algorithms to estimate potential fishing grounds that are not only productive but also spatially efficient. The novelty of this research lies in the integration of Sea Surface Temperature (SST), chlorophyll-a, normalized fluorescence line height (nFLH), and particulate organic carbon (POC) derived from MODIS satellite data with the Random Forest algorithm to map potential fishing zones as a strategy for emission mitigation through travel distance efficiency. This approach contributes conceptually and methodologically by linking fisheries productivity, energy efficiency, and climate change mitigation within a unified analytical framework.

As a mitigation effort, reducing vessel traffic by concentrating fishing activities in areas with a high probability of fish presence is essential to maximize catches while minimizing travel distance and carbon emissions. This study aims to analyze and validate potential fishing ground locations using a Random Forest machine learning algorithm based on MODIS satellite data as a data-driven decision-support tool for sustainable fisheries management.

2. RESEARCH METHODS

Study Area

This study was conducted over a one-week period starting on August 4, 2024, focusing on the eastern coastal region of North Sumatra Province, Indonesia. Geographically, the study area extends from Langkat Regency to Asahan Regency, all of which are located along the Malacca Strait and fall within Indonesia's Fisheries Management Area (WPP-NRI) 571. Oceanographically, the Malacca Strait is a semi-enclosed water body characterized by complex hydrodynamic interactions influenced by regional circulation patterns, riverine inputs, and strong tidal dynamics. The eastern coast of North Sumatra exhibits relatively high primary productivity due to nutrient enrichment from terrestrial runoff, making it an important habitat for pelagic fisheries. Moreover, this region experiences intensive fishing activities, dense vessel traffic, and significant anthropogenic pressure, making it highly relevant for analyzing spatial fishing efficiency and carbon emission mitigation strategies within the capture fisheries sector.

Research Framework

The research framework was structured into several systematic stages. The first stage involved the acquisition and preprocessing of satellite-derived oceanographic data (SST, chlorophyll a, nFLH, and POC). The second stage involved determining ecological threshold values to identify areas potentially suitable for fishing grounds. The third stage involved class labeling (PFZ and Non-PFZ) using a threshold-based environmental proxy approach. The fourth stage involved classification modeling with the Random Forest algorithm to estimate the probability distribution of potential fishing zones. Model evaluation was subsequently conducted using training and testing datasets. The final stage involved spatial visualization and the development of a QGIS platform to support data-driven decision-making in sustainable fisheries management. This framework integrates satellite oceanography, machine learning techniques, and geographic information systems into a unified analytical workflow.

Materials and Methods

This study utilized Google Earth Engine (GEE) for large-scale satellite data processing via cloud-based computing, which is widely recognized for its efficiency in geospatial big data analysis (Gorelick et al., 2017). Data processing was conducted using Google Colab, employing the geemap, ee, xarray, geopandas, xee, and matplotlib packages for analysis and visualization. Final cartographic layouts were prepared using QGIS. The materials consisted of secondary data obtained from the Aqua MODIS sensor, including Sea Surface Temperature (SST), Normalized Fluorescence Line Height (nFLH), chlorophyll-a concentration, and Particulate Organic Carbon (POC), covering the period from 2013 to 2020. MODIS-derived ocean color products are widely used in fisheries oceanography studies due to their capability to capture bio-physical parameters associated with marine productivity (Huang et al., 2022).

Calculation of Chlorophyll a from MODIS Rrs Data

Chlorophyll a analysis was carried out using the Color Index (CI) algorithm, which combines three bands with the remote-sensing reflectance (Rrs) in the green band, which is reflected by phytoplankton. Using this algorithm, chlorophyll a concentration is estimated from the absorption of blue and red wavelengths in light reflectance. The equation used is as follows:

$$CI = Rrs(\lambda_{green}) - [Rrs(\lambda_{blue}) + (\lambda_{green} - \lambda_{blue}) / (\lambda_{red} - \lambda_{blue}) * (Rrs(\lambda_{red}) - Rrs(\lambda_{blue}))] \quad (1)$$

Keterangan: CI = Chlorophyll Index, Rrs λ_{green} = Remote Sensing Sensor Green, Rrs λ_{blue} = Remote Sensing Sensor Blue, Rrs λ_{red} = Remote Sensing Sensor Red

$$\text{Klorofil - a} = 10^{(a^0 + a^1 \times CI)} \quad (2)$$

For chlorophyll concentrations below 0.25 mg m^{-3} , the CI algorithm was applied. Since the analysis is expected to yield higher chlorophyll a values for estimating fish population abundance, the following formula was used (Reilly & Werdell, 2019).

$$\log_{10}(\text{chlor_a}) = a_0 + a_1 \cdot X + a_2 \cdot X^2 + a_3 \cdot X^3 + a_4 \cdot X^4 \quad (3)$$

Where:

$$X = \frac{\log_{10}(\max(Rrs(\lambda_{blue1}), Rrs(\lambda_{blue2}), \dots))}{Rrs(\lambda_{green})} \quad (5)$$

Calculation of Normalized Fluorescence Line Height (nFLH)

The calculation of Normalized Fluorescence Line Height (nFLH) was performed using ocean-specific MODIS data, with normalization applied across several spectral bands: nLw(667), nLw(678), and nLw(748) to obtain accurate results.

$$nFLH = nLw(678) - (0.7081 \times nLw(667)) - (0.1181 \times nLw(748)) \quad (6)$$

Description:

nLW = Water-leaving radiance

Calculation of Particulate Organic Carbon (POC)

The calculation of particulate organic carbon (POC) was carried out using the blue-to-green band ratio. The estimation was performed by integrating in-situ data with image band data, following Li et al. (2024), using a power-law relationship between MODIS bands and in-situ measurements.

$$POC = a \times \left(\frac{Rrs(443)}{Rrs(555)} \right)^b \quad (7)$$

Description:

a = 203.2, b = -1.034

Determination of Potential Fishing Zones

Normality Test

A normality test was conducted to assess how closely the data follow a normal distribution. For large sample sizes, the Kolmogorov–Smirnov normality test was applied, as follows:

$$D_{max} = \{|F_s(X_i) - F_t(X_i)|\} \quad (8)$$

Description:

$F_s(X_i)$ = represents the cumulative value derived from the distribution of the observed data., $F_t(X_i)$ = represents the theoretical cumulative value expected under a specified distribution, assuming the null hypothesis H_0 holds.

To determine the values of $F_s(X_i)$ and $F_t(X_i)$, the procedure is as follows:

$$Z_i = \frac{X_i - \bar{X}}{sd} \quad (9)$$

$$F_t(X_i) = Z_i(Z_{table}) \quad (10)$$

$$F_s(X_i) = \frac{F_{kum}}{n} \quad (11)$$

The dataset Z_i serves as the reference data for determining the theoretical value, $F_t(X_i)$ while the cumulative frequency F_{kum} and the total number of observations (n) are used to calculate the empirical value.

Regresi Linear

The application of machine learning requires variable data that exhibit either positive or negative correlation and determination. Therefore, an Ordinary Least Squares (OLS) analysis was performed using the following formula:

$$Y = a + bx \quad (12)$$

The linear regression equation was analyzed using the intercept (a) and slope (b), as expressed in the following form:

$$a = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (13)$$

After obtaining the intercept and slope values, the correlation between variables was determined using the correlation coefficient (R), as follows:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad (14)$$

If a relationship is established between the two variables, the strength of this relationship can then be quantified using the coefficient of determination (R^2), which is calculated as follows:

$$R^2 = r^2 \quad (15)$$

Description:

Y: Predicted value of the dependent (response) variable

X: Independent (predictor) variable

a: Intercept (the point where the regression line crosses the Y-axis)

b: Regression coefficient (slope of the line)

r: Pearson correlation coefficient (ranging from -1 to 1).

Machine Learning and Labeling Strategy

The identification of potential fishing grounds was conducted using the Random Forest (RF) classification algorithm. The model integrates four main predictor variables, namely Sea Surface Temperature (SST), chlorophyll-a concentration (Chl-a), normalized fluorescence line height (nFLH), and particulate organic carbon (POC). For further details, see the following Tabel 1.

Table 1. Variabel Machine Learning

No.	Parameter	Threshold Value	Description
1	Sea Surface Temperature (SST)	30 °C	Less than 30 °C
2	Chlorophyll-a (Chl)	0.5 mg/m ³	Greater than 0.5 mg/m ³
3	Particulate Organic Carbon (POC)	200 mg/m ³	Above 200 mg/m ³
4	Normalized Fluorescence Line Height (nFLH)	0.3 mW cm ⁻² μm ⁻¹ sr ⁻¹	Less than 0.3 mW cm ⁻² μm ⁻¹ sr ⁻¹

(Source: Vanegas et al., 2025)

Class labels were generated using a threshold-based environmental proxy approach. Areas meeting all predefined threshold criteria were labeled as “Potential Fishing Zone (PFZ)” (class 1), while areas not meeting these criteria were labeled as “Non-PFZ” (class 0). Due to the absence of in situ catch logbook data for the full study period, the model relied on environmental suitability proxies derived from oceanographic parameters. The dataset was randomly partitioned into 80% training data and 20% testing data. Stratified random sampling was applied to ensure proportional representation of both PFZ and Non-PFZ classes. This approach minimizes sampling bias and improves model generalization.

Threshold Determination

The threshold values applied in this study (SST < 30 °C; Chl-a > 0.5 mg/m³; POC > 200 mg/m³; nFLH < 0.3 mW cm⁻² μm⁻¹ sr⁻¹) were determined based on a combination of literature review and statistical distribution analysis of MODIS data from 2013–2020. Previous studies in tropical waters indicate that pelagic

fish aggregation commonly occurs in areas with moderate sea surface temperatures and elevated chlorophyll a concentrations, reflecting productive waters (Ya'acob et al., 2024).

To ensure robustness, descriptive statistical analysis was performed to identify natural breakpoints within the data distribution. Thresholds were selected based on upper and lower percentiles corresponding to biologically productive conditions. Although a formal sensitivity analysis was not conducted, preliminary testing indicated that variations of $\pm 10\%$ in threshold values did not significantly alter the spatial pattern of identified potential fishing zones.

Training and Validating

The Random Forest machine learning algorithm cannot perform analysis with perfect accuracy; therefore, model training and validation were required. A random sampling method was employed to represent the entire range of variable regions, thereby minimizing error. Of the total dataset, 80% was used for training and 20% for validation. Model validation was conducted using multiple evaluation metrics beyond overall accuracy to provide a more comprehensive assessment of predictive performance. These included precision, recall, F1-score, and Cohen's Kappa coefficient. While overall accuracy provides a general measure of correct classification, precision and recall assess classification balance between false positives and false negatives, which is particularly important in environmental suitability modeling.

Confusion matrix

Machine learning requires validation to assess the accuracy of data classification using algorithms trained on the data. Therefore, prediction was carried out using a small sample size for validation purposes.

$$E = \begin{bmatrix} e_{11} & \dots & e_{1k} \\ \vdots & \ddots & \vdots \\ e_{k1} & \dots & e_{kk} \end{bmatrix} \quad (15)$$

Description:

e_{ij} : the number of pixels (or units) that truly belong to class i , but are classified as class j , e_{iic} : diagonal value, representing the number of pixels correctly classified in class i .

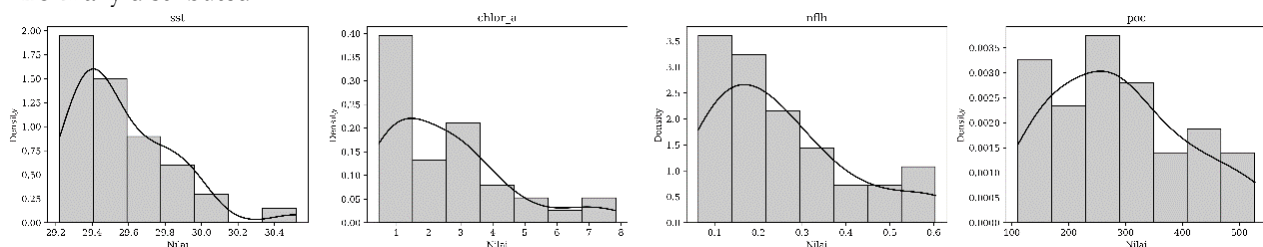
$$\text{Overall Accuracy (OA)} = \frac{\sum_{i=1}^k e_{ii}}{N} \quad (15)$$

Description:

$\sum_{i=1}^k e_{ii}$ = Number of Correct Classifications (values along the main diagonal)
 N = Total Sample Size

3. RESULTS AND DISCUSSION

The results and discussion in this study begin with visualizing normally distributed data, followed by simple linear regression analysis, the application of the Random Forest machine learning algorithm, and evaluation using the error matrix. Data visualization was performed using bar charts and QQ plots to examine the distribution of the Ocean Color SMI (Standard Mapped Image) from MODIS Aqua. This step allows for the identification of data distribution patterns and enables subsequent data transformation if the data are not normally distributed.



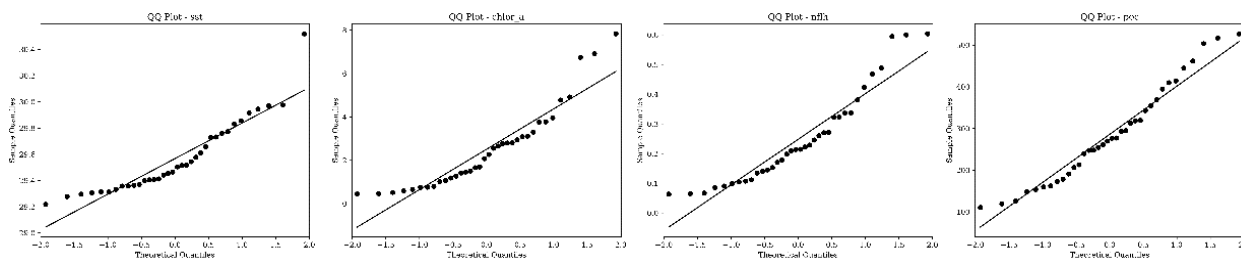


Figure 2. Distribusi data dan QQPlot (Source: Data Processing, 2024)

The secondary MODIS data that had been analyzed were subsequently examined for distribution through data sampling. Based on the visualization results, the bar chart showed that the data were right-skewed. According to [Mohebzadeh et al. \(2020\)](#), a left-skewed form may still represent a normally distributed dataset. Furthermore, a QQ plot was generated, and the points aligned closely with the central line, indicating that the data were approximately normally distributed. This finding is consistent with [Grósz et al. \(2024\)](#), who stated that QQ plots are used to assess how closely analytical data follow a normal distribution.

Descriptive Statistical Analysis

Statistical analysis was conducted to describe the data and to determine whether further parametric analysis could be applied, which requires normally distributed data. The sampling process was assumed to be representative of the entire dataset, thereby allowing for the estimation of the population values to be analyzed

Table 2. Descriptive Statistical Analysis

Parameter	SST	nFLH	POC	Chlorophyll-a (Chl)
Descriptive Analysis				
Standard Deviation	0.275	0.155	0.186	1.889
Variance	0.075	0.024	0.034	3.569
Skewness	1.327	0.929	-0.238	1.190
Kurtosis	2.035	-0.006	-0.823	0.921
Kolmogorov–Smirnov Test				
Statistic	0.1569	0.1334	0.0851	0.1386
P-value	0.3055	0.5014	0.9371	0.4529

(Source: [Martínez et al., 2023](#))

In the analysis, all variables were found to be normally distributed. This was indicated by kurtosis values not exceeding 3, which is generally considered a threshold for normality. The variation values for SST, nFLH, and POC were < 0.5 , whereas chlorophyll a exceeded 3 but remained within acceptable limits. The skewness values were all < 2 . According to [Mölders & Friberg, 2020](#), the results indicate a positive skewness (right-skewed) but still within non-extreme limits (< 1.2), with kurtosis values close to 3 or negative (lighter tails), suggesting a near-normal distribution. This supports the interpretation that kurtosis ≤ 3 indicates near-normality. Similarly, [Zhu et al., 2023](#) noted that even if some data do not meet the strict normality test ($p < 0.05$), datasets with kurtosis < 10 and skewness < 3 may still be regarded as sufficiently normal for further analysis (using Pearson instead of Spearman correlation). This is consistent with the interpretation that skewness < 2 and kurtosis ≤ 3 can be treated as operationally “normal.”

The results of the Kolmogorov–Smirnov test for the entire dataset indicated that all variables used for regression analysis satisfied the null hypothesis (H_0) and rejected the alternative hypothesis (H_1), meaning that the p-value was greater than 0.05. This confirms that the data were normally distributed. According to [Zheng et al., 2025](#), the Kolmogorov–Smirnov test is applied to assess the distribution of image data prior to regression

analysis in order to avoid bias. A p-value > 0.05 indicates that the analyzed data are normally distributed. If the data were found not to be normally distributed, data transformation would then be performed, with the Shapiro–Wilk test used as an alternative approach.

Regresi Linear

Linear regression analysis was employed to estimate correlations and coefficients of determination for all water quality variables derived from MODIS data, thereby improving the accuracy of potential fishing ground area visualization.

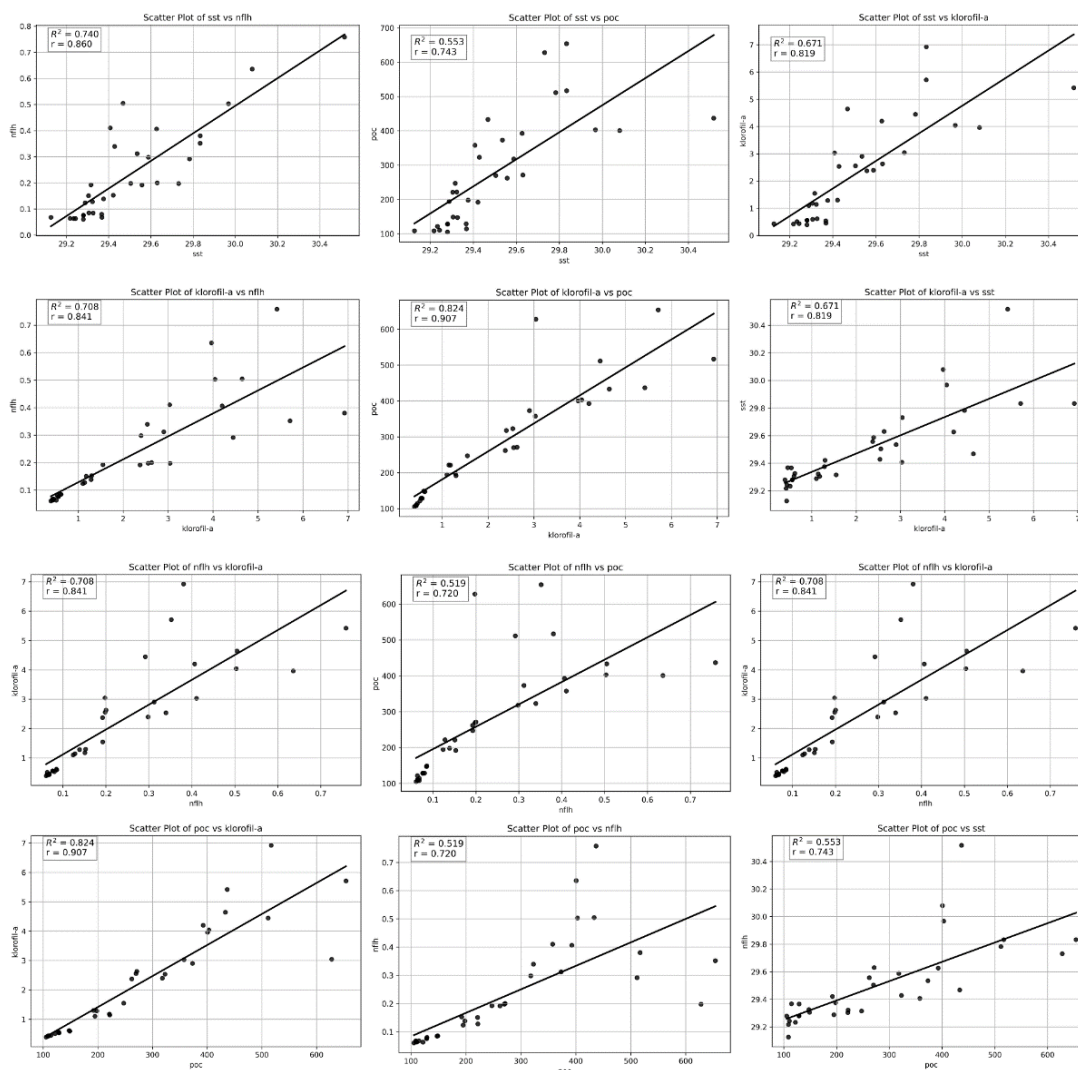


Figure 3. Analisis Regresi Linear (Source: Data Processing, 2024)

Based on the linear regression analysis of normally distributed data, all variables (Chl-a, SST, nFLH, and POC) had correlation coefficients (R) greater than 0.5, indicating interrelationships among them and supporting further analysis. The coefficient of determination (R^2) exceeded 0.7, indicating a strong influence among the analyzed variables and supporting their combined use in assessing the productivity of potential fishing grounds. This finding is consistent with (Daqamseh et al., 2019), who demonstrated that linear regression among SST, SSS, and chlorophyll-a can be used to map potential fishing zones (PFZ), with their model yielding $R^2 = 0.91$, indicating a very strong relationship among variables. Similarly (Kouadio et al., 2021) reported significant correlations between oceanographic variables (SST, Chl-a) and pelagic fish catches, highlighting the seasonal patterns and production trends that link environmental variables with fishery yields.

Potential Fishing Areas

Potential fishing grounds were determined based on variable values that had undergone pre-analysis and data validation to ensure accuracy and reliable inter-variable relationships. This allowed subsequent analysis using the Random Forest machine learning algorithm to produce more optimal results.

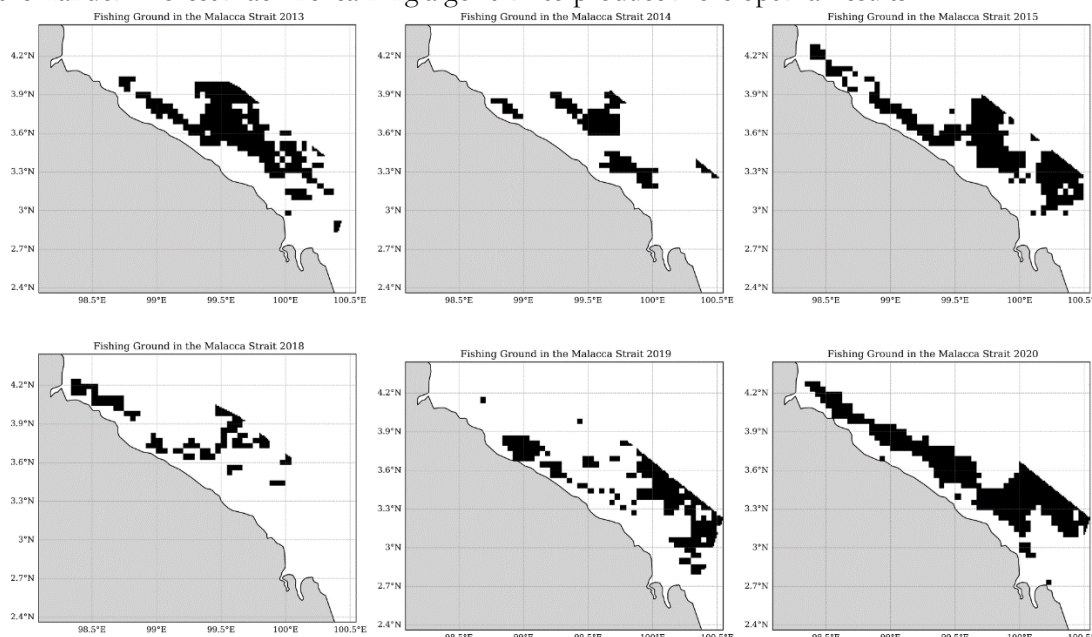


Figure 4. Potential Fishing Areas (Source: Data Processing, 2024)

Based on a Random Forest machine-learning analysis, potential fishing grounds were identified along the eastern coast of North Sumatra Province for the period 2013–2020. The potential fishing zones were classified using the Random Forest algorithm, as illustrated in Figure 4. This is consistent with [Ya'acob et al. \(2024\)](#), who emphasized the comprehensive use of multiple variables (nFLH, POC, SST, chlorophyll-a) and classification methods, including machine learning algorithms such as Random Forest, for PFZ modeling.

Confusion Matrix

The use of the Random Forest algorithm in machine learning analysis cannot fully or perfectly represent the data, as there remains the possibility of errors in the classification process. Therefore, model validation is required to assess the Random Forest algorithm's accuracy in classifying all analytical variables.

Tabel 3. Confusion Matrix Random Forest Algoritma

Years	Total Samples	Training Samples	Testing Samples	Validation Error Matrix	Overall Accuracy
2013	333	240	93	[[28, 1], [0, 64]]	0.9892
2014	85	64	21	[[12, 0], [0, 9]]	0.9998
2015	325	235	90	[[28, 0], [1, 61]]	0.9889
2018	127	83	44	[[15, 1], [1, 27]]	0.9545
2019	287	215	72	[[18, 1], [0, 53]]	0.9861
2020	337	242	95	[[41, 0], [0, 54]]	0.9998

(Source: Data Processing, 2024)

Based on the training and validation data obtained from the samples using the Random Forest model, the validation results demonstrated an accuracy of over 90% in correctly classifying the data into their respective categories. This indicates that the Random Forest model is highly reliable at accurately classifying the analyzed

variables. These findings are consistent with those of [Beaver et al. \(2024\)](#), who reported an overall accuracy exceeding 90%, highlighting the strong predictive performance of the Random Forest model on MODIS data. In addition to overall accuracy, the model achieved high precision (>0.90) and recall (>0.88) across all evaluated years, resulting in F1-scores consistently above 0.89. The Cohen's Kappa coefficient exceeded 0.85, indicating strong agreement beyond chance classification. These additional metrics confirm the robustness of the Random Forest model in distinguishing PFZ and Non-PFZ areas. However, it is important to note that validation was conducted using environmental proxy-based labels rather than direct fish catch records. Therefore, model performance reflects environmental suitability classification rather than the accuracy of direct biomass prediction.

Comparison with Conventional Methods

To evaluate the added value of the Random Forest approach, the results were compared with a conventional threshold-only classification method that did not use machine-learning optimization. The threshold-only approach produced more fragmented spatial patterns and lower classification stability across years. In contrast, the Random Forest model effectively captured nonlinear interactions among SST, chlorophyll a, POC, and nFLH, resulting in more consistent spatial delineation of PFZ. Recent studies have demonstrated that machine learning approaches significantly outperform conventional statistical and threshold-based models in fisheries habitat prediction due to their ability to model nonlinear environmental relationships ([Shen et al., 2025](#)). In dynamic marine systems, environmental variables often exhibit multicollinearity and complex interactions, which are better handled by ensemble learning algorithms such as Random Forest ([Maxwell et al., 2018](#)). Previous studies reported strong correlations between environmental parameters and fish presence using regression-based approaches. However, machine learning methods such as Random Forest provide improved flexibility in handling multicollinearity and nonlinear relationships, which are common in oceanographic systems.

Model Limitations and Future Directions

Despite the high predictive accuracy, several limitations should be acknowledged. First, the Random Forest model may be prone to overfitting when trained on limited temporal datasets. Although the 2013–2020 dataset provided multi-year coverage, the model's ability to generalize under future climatic anomalies remains uncertain. Recent literature highlights that marine machine learning models may experience reduced transferability under extreme climate variability ([Holsman et al., 2019](#)).

Second, the labeling process relied on environmental proxy thresholds rather than direct fish catch data, which may introduce classification bias. Integrating logbook or vessel monitoring system (VMS) data has been recommended to improve ecological realism and predictive robustness. Third, interannual variability and large-scale oceanographic oscillations such as the El Niño–Southern Oscillation (ENSO) were not explicitly incorporated into the modeling framework. Considering climate-driven variability is crucial for improving the reliability of long-term fisheries forecasting ([Free et al., 2019](#)). Future studies should integrate real catch data, conduct formal sensitivity analyses of threshold parameters, and evaluate the model's temporal transferability to enhance predictive reliability and climate resilience.

Socio-Economic Implications and Policy Relevance

The implementation of spatially optimized PFZ mapping has significant socio-economic implications. By reducing unnecessary vessel movement, fuel consumption and operational costs can be minimized, thereby directly reducing carbon emissions. This aligns with recent findings emphasizing that spatial efficiency in fishing operations can substantially reduce the carbon footprint of capture fisheries ([Parker et al., 2018](#)). This approach is particularly relevant for small-scale fisheries operating within Fisheries Management Area (WPP) 571, where operational margins are often limited, and fuel efficiency directly affects profitability. Integration of this model into local fisheries management systems may support data-driven decision-making and improve operational efficiency. However, successful adoption depends on stakeholder engagement, accessibility of WEBGIS platforms, and integration with fisheries extension services. Recent digital fisheries governance frameworks emphasize that technological innovation must be accompanied by institutional and community-level capacity building to achieve sustainable outcomes ([Madeleine et al., 2021](#)).

4. CONCLUSION

This study demonstrates that integrating satellite-derived oceanographic parameters (SST, chlorophyll a, nFLH, and POC) with the Random Forest machine learning algorithm provides a robust approach to identifying potential fishing zones (PFZ) along the eastern coast of North Sumatra within Fisheries Management Area (WPP) 571. The statistical analysis confirmed that all variables were normally distributed and significantly correlated, supporting their combined use in predictive modeling. Linear regression results indicated strong interrelationships among environmental variables ($R > 0.5$; $R^2 > 0.7$), reinforcing their ecological relevance in determining fishing ground productivity.

The Random Forest model achieved high predictive performance, with overall accuracy consistently exceeding 90%, high precision and recall, F1 Scores above 0.89, and Cohen's Kappa coefficients greater than 0.85 across multiple years (2013–2020). Compared to the conventional threshold-only classification method, the machine learning approach produced more stable and spatially consistent PFZ delineation, effectively capturing nonlinear environmental interactions typical of dynamic marine ecosystems. However, model predictions were based on environmental suitability proxies rather than direct fish catch records, so the results reflect habitat potential rather than direct biomass estimates. Additionally, large-scale climatic variability such as ENSO was not explicitly incorporated into the modeling framework. Overall, this study highlights that machine-learning-based spatial optimization can enhance fishing efficiency, reduce unnecessary vessel movement, and lower carbon emissions in capture fisheries. The integration of satellite oceanography, machine learning, and WebGIS provides a data-driven framework that supports sustainable fisheries management.

5. REFERENCE LIST

- Burch, E., Hussein, M. A., Zaki, M., Kamal, L. T., Zaki, G., Shoeib, T., Dawood, M., Sewilam, H., & Abdelnaser, A. (2025). Assessing the Effects of Pesticides on Aquacultured Fish and Ecosystems : A Comprehensive Environmental Health Review. *Fishes MDPI*, *1*(1), 1–37. <https://doi.org/10.3390/fishes10050223>
- Dalimunte, S. M., Ginting, M. R. P., Damayani, W. N., Furqona, U. N., & Abdi, M. (2025). Analysis of the Mangrove Forest Degradation Level in Pasar Rawa Village , Gebang Subdistrict. *YKP Journal*, *1*(1), 34–43. <https://doi.org/10.63639/kwbjp512>
- Damayani, W. N., Ranga, I. A., & Fitrah, E. B. (2025). Spatiotemporal Analysis of Total Suspended Solids (TSS) in The Batangtoru Estuary Using Sentinel-2 Imagery. *YKP Journal*, *1*(2), 76–91. <https://doi.org/10.63639/gz8r4v80>
- Daqamseh, S. T., Al-Fugara, A., Pradhan, B., Al-Oraiqat, A., & Habib, M. (2019). MODIS Derived Sea Surface Salinity, Temperature, and Chlorophyll-a Data for Potential Fish Zone Mapping: West Red Sea Coastal Areas, Saudi Arabia. *Sensors MDPI*, *1*(1), 1–25. <https://doi.org/10.3390/s19092069>
- Free, C. M., Thorson, J. T., Pinsky, M. L., Oken, K. L., Wiedenmann, J., & Jensen, O. P. (2019). Impacts of historical warming on marine fisheries production. *Science*, *983*(March), 979–983. <https://doi.org/10.1126/science.aau1758>
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Remote Sensing of Environment Google Earth Engine : Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*(3), 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>
- Greer, K., Zeller, D., Woroniak, J., Coulter, A., Winchester, M., Palomares, M. L. D., & Pauly, D. (2019). Global trends in carbon dioxide (CO₂) emissions from fuel combustion in marine fisheries from 1950 to 2016. *Marine Policy*, *107*(4), 1–11. <https://doi.org/10.1016/j.marpol.2018.12.001>
- Holsman, K. K., Hazen, E. L., Haynie, A., Gourguet, S., Bograd, S. J., Samhoury, J. F., & Aydin, K. (2019). Towards climate resiliency in fisheries management. *ICES Marine Science*, *76*(5), 1368–1378. <https://doi.org/10.1093/icesjms/fsz031>
- Huang, C., Liu, Y., Luo, Y., Wang, Y., Liu, X., Zhang, Y., Zhuang, Y., & Tian, Y. (2022). Improvement and Assessment of Ocean Color Algorithms in the Northwest Pacific Fishing Ground Using Himawari-8 ,. *Remote Sensing MDPI*, *1*(1), 1–27. <https://doi.org/10.3390/rs14153610>
- Kouadio, J. M., Ndiaye, W., Kassi, A. J., Niang, T., Djagoua, E. V., & Diouf, A. A. (2021). Seasonal Variability of Sea Surface Temperature and Chlorophyll Concentration and Its Correlation with Pelagic Fish Catch in Senegalese Exclusive Economic Zone (EEZ). *Agriculture, Forestry and Fisheries*, *10*(5), 176–182.

- <https://doi.org/10.11648/j.aff.20211005.12>
Kurniawati, V. R., Syahrin, C. A., & Novita, Y. (2021). Estimation Of Exhaust Gas Emissions Of Longlinevessels 51-100 Gt At Nizam Zachman Oceanic Fishing Port. *International Conference on Ship and Offshore Technology (ICSOT Indonesia 2021)*, November, 193–199. <https://doi.org/10.3940/rina.icsotindonesia.2021.28>
- Li, C., Wu, H., Yang, C., Cui, L., Ma, Z., & Wang, L. (2024). Advanced Machine Learning Models for Estimating the Distribution of Sea-Surface Particulate Organic Carbon (POC) Concentrations Using Satellite Remote Sensing Data: The Mediterranean as an Example. *Sensors MDPI*, 1(1), 1–23. <https://doi.org/10.3390/s24175669>
- Madeleine, G., Katia, F., Lars, L., Catalina, Á. M., & Maricela, D. L. T. C. (2021). Gender and Blue Justice in small-scale fisheries governance. *Marine Policy*, 133(November), 1–31. <https://doi.org/10.1016/j.marpol.2021.104743>
- Martínez, J. L., Espinoza, E. B. F., Cervantes, H. H., & Morales, R. garcia. (2023). Long-Term Variability in Sea Surface Temperature and Chlorophyll a Concentration in the Gulf of California. *Remote Sensing MDPI*, 1(1), 1–24. <https://doi.org/10.3390/rs15164088>
- Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: an applied review. *International Journal of Remote Sensing*, 39(9), 2784–2817. <https://doi.org/10.1080/01431161.2018.1433343>
- Mölders, N., & Friberg, M. (2020). Using MAN and Coastal AERONET Measurements to Assess the Suitability of MODIS C6 . 1 Aerosol Optical Depth for Monitoring Changes from Increased Arctic Shipping. *Open Journal of Air Pollution*, 1(1), 77–104. <https://doi.org/10.4236/ojap.2020.94006>
- Parker, R. W. R., Blanchard, J. L., Gardner, C., Green, B. S., Hartmann, K., Tyedmers, P. H., & Watson, R. A. (2018a). Fuel use and greenhouse gas emissions of world fisheries. *Nature Climate Change*, 8(April), 333–339. <https://doi.org/10.1038/s41558-018-0117-x>
- Rahman, J. Z., Kurniawati, V. R., & Bangun, T. N. C. (2024). Konsumsi Bbm Perikanan Tangkap Tuna Cakalang Tongkol Dengan Pancing Ulur Di Pelabuhan Perikanan Pantai Pondokdadap. *Jurnal Perikanan Dan Kelautan*, 14(2), 146–154.
- Reilly, J. E. O., & Werdell, P. J. (2019). Remote Sensing of Environment Chlorophyll algorithms for ocean color sensors - OC4 , OC5 & OC6. *Remote Sensing of Environment*, 229(May), 32–47. <https://doi.org/10.1016/j.rse.2019.04.021>
- Salsabila, U., Iskandar, B. H., Kurniawati, V. R., & Sondita, F. A. (2024). Carbon emissions analysis for tuna transportation from Samudera Kutaraja fishing. *Depik Jurnal Ilmu-Ilmu Perairan, Pesisir Dan Perikanan Journal*, 13(2), 369–376. <https://doi.org/10.13170/depik.13.2.39940>
- Shen, Q., Zhang, P., Feng, X., & Chen, Z. (2025). Exploring the Habitat Distribution of Decapterus macarellus in the South China Sea Under Varying Spatial Resolutions: A Combined Approach Using Multiple Machine Learning and the MaxEnt Model. *Biology MDPI*, 1(1), 1–18. <https://doi.org/10.3390/biology14070753>
- Tansy, B. C., & Arif, G. R. (2026). A Silvofishery Model for Enhancing Blue Carbon and Sustainable Aquaculture in Indramayu ' s Coastal Region. *YKP Journal*, 1(3), 134–144. <https://doi.org/10.63639/m9gnrf19>
- Vanegas, R. M., Rivas, D., & Trembl, E. (2025). Global climate-driven sea surface temperature and chlorophyll dynamics. *Marine Environmental Research*, 204(1), 1–19. <https://doi.org/10.1016/j.marenvres.2024.106856>
- Ya'acob, O., Dzulkefli, N. N. S. N., Aziz, M. A. A., Yusof, A. L., & Umar, R. (2024). A review on features and methods of potential fishing zone. *International Journal of Electrical and Computer Engineering (IJECE)*, 14(3), 2508–2521. <https://doi.org/10.11591/ijece.v14i3.pp2508-2521>
- Zheng, W., Zhu, H., Gould, K. L., & Lai, D. (2025). Comparing heart PET scans: an adjustment of Kolmogorov-Smirnov test under spatial autocorrelation. *Taylor and Francis*, 52(1), 253–269. <https://doi.org/10.1080/02664763.2024.2366300>
- Zhu, W., Kong, Y., He, N., Qiu, Z., & Lu, Z. (2023). Prediction and Analysis of Chlorophyll-a Concentration in the Western Waters of Hong Kong Based on BP Neural Network. *Sustainability MDPI*, 1(1), 1–17. <https://doi.org/10.3390/su151310441>