FORUM GEOGRAFI
ISSN: 0852-0682 | E-ISSN: 2460-3945

HTTPS://JOURNALS.UMS.AC.ID/INDEX.PHP/FG/

Research article

GeoAI-Based Prediction of Yellowfin Tuna (*Thunnus albacares*) Potential Fishing Zones in the Western Sumatra Waters, Indonesia

Amir Yarkhasy Yuliardi[1,*], Muhammad Hanif[2], Septy Heltria[3], Diah Ayu Rahmalia[1], Gandhi Napitupulu[4]

[1] Marine Science, Faculty of Fisheries and Marine Science, Jenderal Soedirman University, 53112, Indonesia; [2] Geoinformatics, College of Computing, Khon Kaen University, 40002, Thailand; [3] Department of Fisheries, Faculty of Animal Science, Universitas Jambi, 36361, Indonesia; [4] Graduate School of Innovation and Practice for Smart Society, Hiroshima University, 739-8529, Japan.

[*)] Correspondence: amiryarkhasy@gmail.com

Citation:
Yuliardi, A. M., Hanif, M., Heltria, S., Rahmalia, D. A., & Napitupulu, G.. (2027). Title of article refer to above title. Forum Geografi. 41(1), 54-71.

Article history:
Received: 31 January 2026
Revised: 23 July 2026
Accepted: 18 August 2026
Published: 21 August 2026

Abstract

The western waters of Sumatra in the eastern Indian Ocean are highly productive fishing grounds for yellowfin tuna (Thunnus albacares). However, the dynamic oceanographic conditions of these waters pose challenges for reliable fisheries prediction. Therefore, this study aimed to identify key environmental drivers and predict Potential Fishing Zones (PFZs) using GeoArtificial Intelligence (GeoAI)-based machine learning (ML) framework integrating Guided Regularized Random Forest (GRRF) and Random Forest (RF). The predictive models were developed using multi-source oceanographic variables as predictors, including sea surface temperature (SST), chlorophyll-a (Chl-a), salinity, current velocity, and net primary productivity (NPP). The results of correlation analysis showed that there was a strong relationship between Chl-a and NPP ($r = 0.87$), indicating the central role of primary productivity in tuna habitat dynamics. Both models achieved high predictive accuracy, with GRRF slightly outperforming RF by reducing variable redundancy and improving interpretability. Tuna occurrence was strongly associated with moderate SST (29.2–29.8°C), high NPP (>350 mg C m^{-2} day^{-1}), and zones influenced by the South Java Coastal Current (SJCC) and the Indonesian Throughflow (ITF). Spatially, PFZs were concentrated along the Mentawai slope and adjacent offshore waters, particularly during monsoon transition periods. These results showed the effectiveness of GeoAI-based methods for predicting PFZs and supported sustainable tuna fisheries management. Moreover, future studies should incorporate interannual climate variability (ENSO, IOD) and advanced deep learning frameworks to enhance spatiotemporal prediction.

Keywords: Potential Fishing Zones (PFZ); Yellowfin tuna (Thunnus albacares); GeoAI and machine learning; Oceanographic variability; Environmental sustainability; Eastern Indian Ocean.

1. Introduction

The prediction of Potential Fishing Zones (PFZs) is essential to improve fishing efficiency, reduce operational costs, and support sustainable fisheries management (Nadeem *et al.,* 2025). This is because accurate identification of fishing grounds is important for highly migratory pelagic species such as yellowfin tuna (Thunnus albacares), whose distribution is strongly influenced by dynamic oceanographic conditions (Nóbrega *et al.,* 2022). Environmental drivers including sea surface temperature (SST), chlorophyll-a (Chl-a) concentration, salinity, current velocity, and primary productivity also vary across space and time in response to monsoonal forcing, regional circulation systems, and air–sea interactions (Singh *et al.,* 2024; Ahmad *et al.,* 2024; Yuliardi *et al.,* 2026). Due to the complexities, predicting tuna fishing grounds remains challenging because habitat suitability is determined by the complex interactions among multiple physical and biological variables. Consequently, understanding these interactions is essential for developing a reliable framework of PFZs and supporting ecosystem-based fisheries management.

To address these complexities, analytical methods capable of integrating and interpreting large volumes of multidimensional geospatial data are essential. Recent advances in machine learning (ML) applied to geospatial data, commonly referred to as GeoArtificial Intelligence (GeoAI), have created transformative opportunities for modeling and predicting complex spatial phenomena (Srivastava & Saxena, 2023; Siddique, 2024; Boutayeb *et al.,* 2025). This interdisciplinary field integrates artificial intelligence (AI), geographic information systems (GIS), and geospatial sciences to intelligently analyze, model, and interpret spatial datasets (Gao *et al.,* 2023). Through several methods such as deep learning and neural networks, ML enables efficient processing of large-scale spatial data, as well as allows extraction of important patterns and produces actionable insights for decision-making (Hosen *et al.,* 2023; Ni & Wang, 2024).

In oceanographic studies, GeoAI applications offer substantial potential for analyzing complex and dynamic marine systems. ML algorithms have shown strong capability in handling multidimensional datasets, capturing nonlinear relationships, and producing reliable predictions in offshore zones where in situ observations are limited (Song *et al.,* 2023a; Hanif *et al.,* 2024; Zhao *et al.,* 2024; Hao *et al.,* 2025). Consequently, ML-based methods have been widely applied to improve the estimation of oceanographic parameters as well as identify PFZs and fish resource distributions for efficient and sustainable fisheries management (Vinston & Ashok, 2022; Sarangi *et al.,* 2024; John *et al.,* 2025; Islam & Moontaha, 2025). The application of ML-based methods is particularly valuable for yellowfin tuna. As a high-value pelagic species distributed across tropical and subtropical waters, including western Sumatra in the eastern Indian Ocean, yellowfin tuna possesses strong sensitivity to environmental variability (Nurholis *et al.,* 2021). The distribution of this species is majorly influenced by various environmental drivers such as SST, Chl-a, salinity, current velocity, and primary productivity, which vary in response to monsoonal circulation and cross-equatorial current systems. Therefore, understanding the spatiotemporal interactions is essential for effective identification of PFZs.

Previous studies in western Sumatra have established the importance of oceanographic drivers in shaping yellowfin tuna distribution. However, most reports relied on descriptive or single-model analyses without critically examining data resolution, model assumptions, or the generalizability of the results. For example, Lisna *et al.* (2024) reported significant relationships between tuna catch and key environmental variables, particularly SST and Chl-a (Lisna *et al.,* 2024). Habitat modeling using MaxEnt further showed high predictive performance (AUC > 0.86), exceeding the predictive accuracy commonly reported for many correlative species-distribution models (AUC ≈ 0.7–0.9), with Chl-a identified as the dominant predictor of tuna abundance (Lisna *et al.,* 2024). Spatial and temporal variability in fishing grounds, effort, and CPUE across the eastern Indian Ocean off Sumatra has also been documented using GIS-based and GAM methods, underscoring the role of spatial dynamics in tuna fisheries (Nurholis *et al.,* 2020).

Additionally, large-scale climate modes, including the Indian Ocean Dipole (IOD) and El Niño–Southern Oscillation (ENSO), were reported to modulate marine productivity and tuna availability along the west coast of Sumatra (Fadhilah *et al.,* 2022). Other studies have addressed fisheries productivity, biological sustainability, and the application of satellite-derived variables to tuna habitat assessment (Brown *et al.,* 2023; Siregar *et al.,* 2025; Al Iqromi *et al.,* 2025; Sari *et al.,* 2024). Despite the significant contributions, these publications have primarily relied on descriptive spatial analyses and relatively simple statistical models, such as regression, generalized additive model (GAM), and MaxEnt, typically using no more than two or three oceanographic predictors, mainly SST and Chl-a. This limited representation of environmental conditions restricts the ability of the models to capture the multivariate physical–biological processes that structure tuna habitat.

Although previous studies have improved understanding of tuna habitat dynamics, important limitations remain because most reports depended on correlation analyses or simple linear regression. More advanced methods such as GAM and MaxEnt have also been used to inherently model nonlinear relationships. However, GeoAI-based ML methods provide the additional advantage of simultaneously incorporating, ranking, and selecting from a larger set of correlated environmental predictors. Most studies considered only a limited set of predictors, typically restricted to SST and Chl-a, which are available satellite-derived variables, but neglect other ecologically relevant data such as currents, salinity, and net primary productivity (NPP) that strongly influence tuna habitat suitability (Heltria *et al.,* 2026; Fadhilah *et al.,* 2022). The widespread use of seasonally or annually aggregated data with coarse spatial resolution has further constrained the representation of local-scale variability derived from fishing point observations. Despite these advances, there is a need for an integrated modeling framework capable of incorporating multiple oceanographic predictors while providing high-resolution spatial predictions of yellowfin tuna fishing potential in western Sumatra waters.

To address the limitations, this study uses a GeoAI framework based on Guided Regularized Random Forest (GRRF) and Random Forest (RF) to handle nonlinear relationships, multicollinearity, and high-dimensional environmental datasets while improving predictor selection and model interpretability. Compared with commonly used models such as regression, GAM, and MaxEnt, RF and GRRF provide greater flexibility for handling nonlinear relationships, high-dimensional datasets, and predictor redundancy simultaneously. Despite the successful application of GRRF and RF to tuna and other pelagic fisheries (Zhang *et al.,* 2023; Sarangi *et al.,* 2024), these models have not been used to generate high-resolution PFZ maps specifically for yellowfin tuna in the western Sumatra waters.

Therefore, this is considered a regional application and methodological extension of GeoAI framework rather than a novel method. Building on the ensemble ML framework of Zhang *et al.* (2023), which achieved high predictive accuracy for albacore tuna in the South Pacific, this study adapts a GeoAI-based method to the eastern Indian Ocean context. The integrated GRRF–RF is used to identify influential environmental drivers and generate spatial predictions of yellowfin tuna fishing potential. Specifically, this study aims to (1) identify the key environmental variables governing yellowfin tuna distribution, (2) develop predictive PFZs using GRRF and RF, and (3) evaluate model performance in mapping spatial fishing opportunities. The proposed framework provides a foundation for developing decision-support tools to enhance efficiency and sustainability in tuna fisheries management in the eastern Indian Ocean.

2. Methods

2.1. Study Area

The study area is located in the western waters of Sumatra within the eastern Indian Ocean (0.5° N–4° S; 97.5° E–101.2° E; Figure 1), including the Mentawai Islands and the adjacent continental margin of West Sumatra, Indonesia. The selection of the area was based on the presence of a properly documented, economically important tuna fishery with strong, well-characterized oceanographic gradients. These characteristics provided suitable conditions for evaluating the relationships between environmental predictors and tuna catch zones within a GeoAI-based PFZ. Quantitative support for this selection was provided by 2,261 georeferenced yellowfin tuna handline fishing events recorded at the Bungus Ocean Fishing Port during 2022 (Table 1), alongside previous catch and catch per unit effort (CPUE) records showing sustained fishing activity throughout the area (Nurholis *et al.,* 2020; Brown *et al.,* 2023). The data provided showed sustained fishing pressure, supporting this designation with quantitative evidence rather than a purely qualitative description.

The western waters of Sumatra represent a dynamic oceanographic transition zone influenced by several major surface circulation systems. The South Equatorial Current (SEC) and South Equatorial Counter Current (SECC) dominate offshore circulation, while the South Java Coastal Current (SJCC) and the Indonesian Throughflow (ITF) transport warm, low-salinity waters into the eastern Indian Ocean along the western coast of Sumatra (Utari *et al.,* 2019; Rahaman *et al.,* 2020). Interactions among these circulation systems generate significant spatial and temporal variability in SST, salinity, Chl-a, and primary productivity, corresponding to the environmental predictors used in this study and representing key factors influencing prey availability and thermal habitat suitability for yellowfin tuna.

Bathymetric conditions further enhance environmental heterogeneity, as the steep continental slope extending from the Sunda Shelf toward the Wharton Basin promotes upwelling and vertical mixing, increasing nutrient availability and phytoplankton productivity along the continental margin (Dalabehara & Sarma, 2021; Xing *et al.,* 2023). Furthermore, persistent frontal zones and enhanced biological productivity from the interaction between circulation and bathymetry create favorable foraging habitats for yellowfin tuna. These environmental characteristics provide a strong ecological basis for investigating the oceanographic controls on tuna distribution and fishing activity within the study area.

2.2. Methodology

This study adopted a GeoAI-based ML framework to predict PFZs of yellowfin tuna in the western waters of Sumatra, with the overall workflow shown in Figure 2. The method comprised sequential stages including data preprocessing, feature correlation analysis, feature importance assessment, and model optimization to ensure robust predictive performance across spatial datasets. The modeling framework integrated GRRF for feature selection and RF for classification and spatial prediction. Specifically, GRRF was applied to reduce redundancy among correlated oceanographic variables and retain the most informative predictors (Izquierdo-Verdiguier & Zurita-Milla, 2020; Iranzad & Liu, 2025).

The selected features were used in RF algorithm to generate probabilistic spatial predictions of tuna fishing potential. Model performance was evaluated using multiple statistical metrics, including accuracy (ACC), precision (P), recall (R), F1-score, mean absolute error (MAE), and root mean square error (RMSE) (Sahib *et al.,* 2025). The optimized RF model was finally applied to the full dataset to produce spatial probability maps of predicted fishing zones, providing an interpretable, data-driven representation of the nonlinear relationships between oceanographic conditions and tuna fishing patterns.

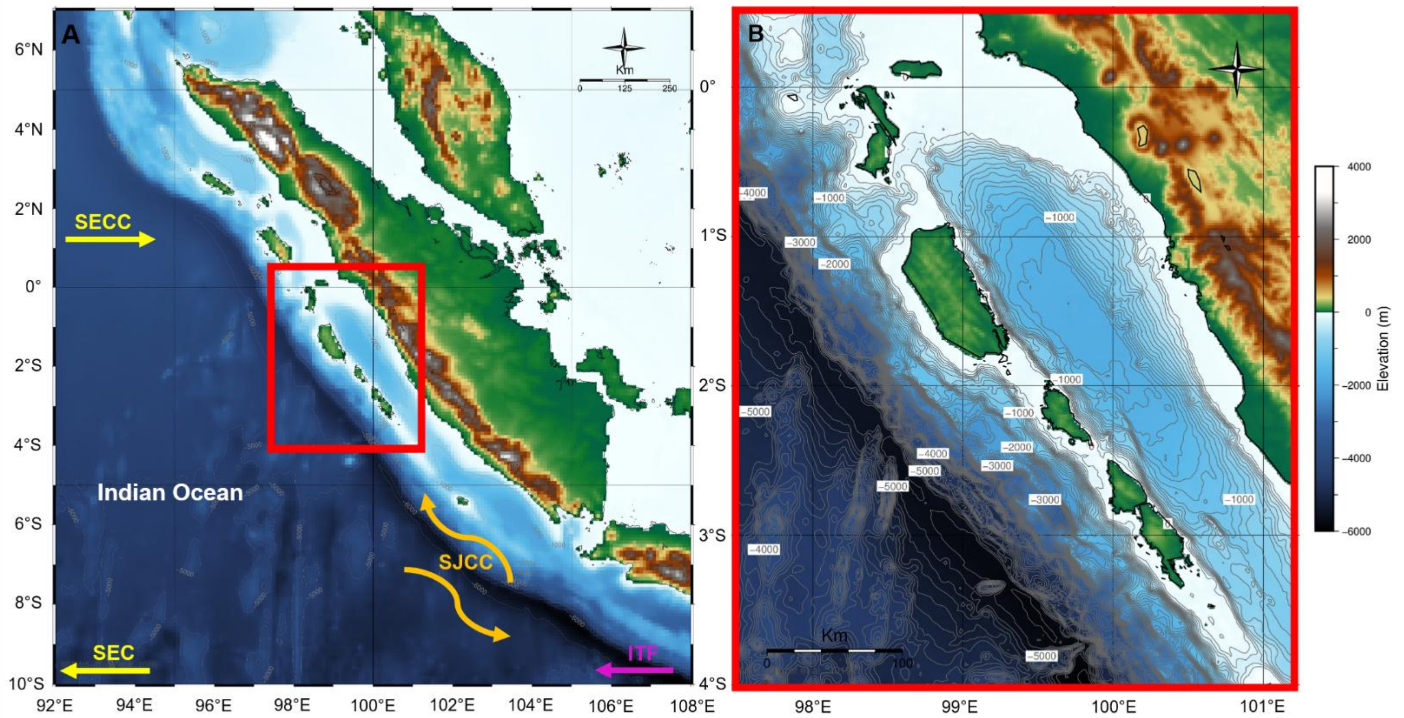

Figure 1. Bathymetric and oceanographic map of the study area in the western waters of Sumatra, Indonesia. (A) Regional setting in the eastern Indian Ocean showing major surface currents, including the South Equatorial Current (SEC), South Equatorial Counter Current (SECC) (Rahaman et al., 2020), South Java Coastal Current (SJCC), and Indonesian Throughflow (ITF) (Utari et al., 2019). (B) Detailed bathymetry of the study area (0.5° N – 4° S; 97.5° E – 101.2° E) illustrating the steep continental slope and the Mentawai Islands chain. Elevation is shown in meters, with color shading representing land topography (green–brown) and ocean bathymetry (light to dark blue). Maps were generated using the Python Generic Mapping Tools (PyGMT) version 0.13.0 (Wessel et al., 2019; PyGMT Development Team, https://www.pygmt.org, 2025).

This study adopted a GeoAI-based ML framework to predict PFZs of yellowfin tuna in the western waters of Sumatra, with the overall workflow shown in Figure 2. The method comprised sequential stages including data preprocessing, feature correlation analysis, feature importance assessment, and model optimization to ensure robust predictive performance across spatial datasets. The modeling framework integrated GRRF for feature selection and RF for classification and spatial prediction. Specifically, GRRF was applied to reduce redundancy among correlated oceanographic variables and retain the most informative predictors (Izquierdo-Verdiguier & Zurita-Milla, 2020; Iranzad & Liu, 2025). The selected features were used in RF algorithm to generate probabilistic spatial predictions of tuna fishing potential. Model performance was evaluated using multiple statistical metrics, including accuracy (ACC), precision (P), recall (R), F1-score, mean absolute error (MAE), and root mean square error (RMSE) (Sahib *et al.,* 2025). The optimized RF model was finally applied to the full dataset to produce spatial probability maps of predicted fishing zones, providing an interpretable, data-driven representation of the nonlinear relationships between oceanographic conditions and tuna fishing patterns.

2.3. Data Reference

This study used 2,261 georeferenced fishing zones of yellowfin tuna obtained from the Bungus Ocean Fishing Port (PPS Bungus), West Sumatra, Indonesia, in 2022. Each point represented validated coordinates of handline fishing operations recorded in official port logbooks. Before analysis, records were screened for duplicate coordinates, positions falling outside the study domain or on land, and entries with missing or inconsistent date/position fields. Flagged records were removed, obtaining a cleaned dataset used as the empirical basis for modeling tuna distribution and PFZs. To account for seasonal variability driven by monsoon-related oceanographic processes in the eastern Indian Ocean, the reference data were classified into four climatological seasons, namely DJF (December–February), MAM (March–May), JJA (June–August), and SON (September–November), as well as an annual composite (Table 1). Each fishing record was temporally matched to the corresponding daily Earth observation (EO) composite for recorded date (Section 2.4). Seasonal subsets were formed by pooling the daily-matched records within each monsoonal period. This grouping enabled the model to capture fluctuations associated with monsoon circulation and productivity cycles. For ML implementation, the dataset was divided into training (70%) and validation (30%) subsets using stratified random sampling to maintain balanced spatial and seasonal representation, thereby reducing bias and enhancing model generalization across both seasonal and annual scales (Abdi, 2020; Chabalala *et al.*, 2023). However, the stratified random split did not explicitly account for spatial autocorrelation among nearby fishing points. The fisheries-dependent dataset also showed the spatial distribution of fishing effort, gear type (handline), and fisher behavior rather than a sample of tuna presence. The issues were treated as limitations of the sampling design and were discussed in Section 4. Therefore, the dataset was restricted to a single year (2022), capturing only the seasonal cycle observed rather than interannual variability in tuna distribution or environmental conditions. This temporal limitation was further discussed in Section 5.

Table 1. Reference Points for Data Prediction.

Seasons	Periods	Number of references
DJF	December – February	531
MAM	March - May	418
JJA	June - August	516
SON	September - November	796
Yearly	January - December	2,261

Sources: Ocean Fishing Port (PPS Bungus).

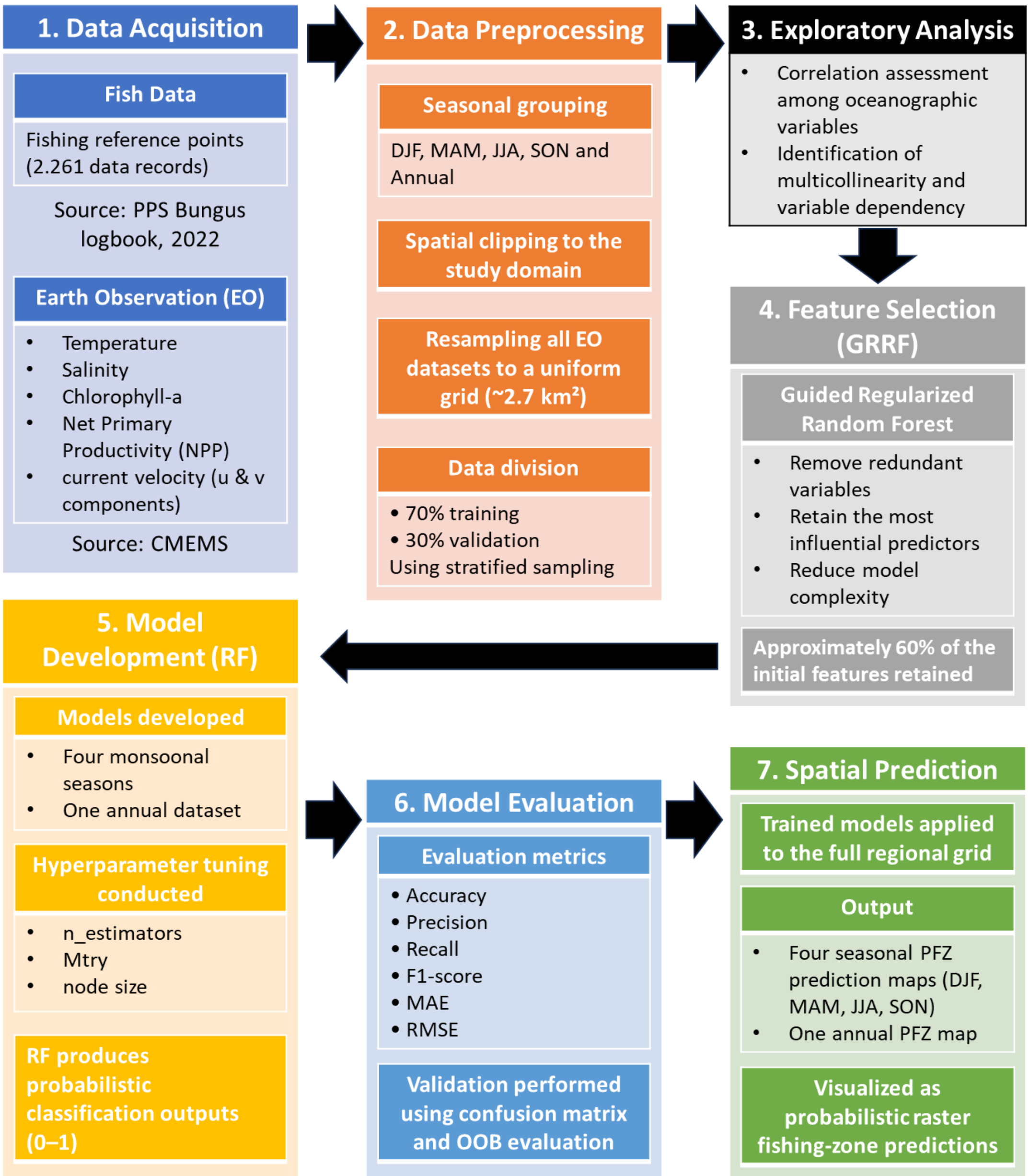

Figure 2. Workflow of the GeoAI-based machine learning framework for predicting yellowfin tuna (*Thunnus albacares*) potential fishing zones in the western Sumatra waters. The process includes data

preprocessing, correlation analysis, feature selection (GRRF), model optimization and training (RF), model evaluation, and final spatial prediction.

2.4. Earth Observation (EO) for Oceanography

To characterize the oceanographic conditions influencing the spatial distribution of yellowfin tuna, this study used multi-parameter Earth Observation (EO) datasets from the Copernicus Marine Environment Monitoring Service (CMEMS) (Lellouche et al., 2018). CMEMS products provide consistent, high-quality, and temporally continuous environmental information suitable for spatially explicit marine habitat modeling. The selected variables included SST, sea surface salinity (SSS), Chl-a concentration, NPP, and surface current velocity components (u and v). These variables serve as key drivers of tuna habitat suitability and fishing ground variability. All EO datasets were extracted for the study domain (0.5° N–4° S; 97.5° E–101.2° E) and matched to each fishing record at daily resolution for the corresponding date in 2022, ensuring temporal consistency between predictors and the reference catch data (Section 2.3). The original spatial resolutions of the datasets ranged from 1/12° (~9 km) to 1/4° (~25 km), as shown in Table 2. To improve spatial compatibility with the densely distributed fishing point observations, all environmental variables were resampled to a uniform grid resolution of approximately 2.7 km² using bilinear interpolation. This resampling represented a three- to nine-fold increase in nominal resolution relative to the native EO products but did not introduce additional spatial information beyond the source grids. However, the procedure established a common analysis grid consistent with the density of the fishing point observations. Consequently, spatial predictions should be interpreted according to the effective resolution of the original EO datasets (~9–25 km) rather than the 2.7 km² analysis grid to avoid an overestimation of spatial detail or predictive capability beyond the limits of the source data.

Table 2. Reference points for data prediction.

No	Name of data	Pixel resolution	Sources
1	Temperature	1/12° (0.083° × 0.083°)	https://doi.org/10.48670/moi-00021
2	Salinity	1/12° (0.083° × 0.083°)	https://doi.org/10.48670/moi-00021
3	Net Primary Productivity	1/12° (0.083° × 0.083°)	https://doi.org/10.48670/moi-00020
4	Chlorophyll-a	1/4° (0.25° × 0.25°)	https://doi.org/10.48670/moi-00015
5	u-component	1/12° (0.083° × 0.083°)	https://doi.org/10.48670/moi-00021
6	v-component	1/12° (0.083° × 0.083°)	https://doi.org/10.48670/moi-00021

2.5. Feature importance and Variable Correlation

GRRF was applied to assess the importance of environmental variables influencing the spatial distribution of yellowfin tuna. This algorithm enhanced model performance by selecting the most relevant predictors while reducing computational complexity and addressing high-dimensional data challenges (Breiman, 2001; Izquierdo-Verdiguier & Zurita-Milla, 2020). Through recursive feature elimination and normalization, GRRF was used to evaluate each variable's contribution to model accuracy and retain predictors with the highest discriminative power for spatial classification (Geerts *et al.,* 2024). Based on the results, variable importance rankings were generated using the default regularization criterion of Izquierdo-Verdiguier and Zurita-Milla (2020). This procedure retained variables with importance scores exceeding a fixed proportion of the maximum observed importance, corresponding to approximately 60% of the extracted data. Since this threshold followed the algorithm's built-in regularization rule rather than an independently tuned or performance-based selection procedure, the resulting subset should be regarded as one plausible configuration rather than a formally optimized solution. Therefore, future investigations should evaluate model sensitivity across a range of selection thresholds. In this study, selected variables were incorporated into RF model to generate high-resolution spatial predictions of PFZs in the western waters of Sumatra. Combining GRRF with RF reduced redundancy, improved computational efficiency, and enhanced classification accuracy, thereby strengthening regional-scale identification of PFZs.

2.6. Estimate Fishing Zone

RF algorithm was used to estimate the spatial distribution of PFZs for yellowfin tuna. This algorithm is an ensemble ML method based on bootstrap aggregating, where multiple decision trees are built from randomly selected subsets of training data and combined to produce stable and robust predictions (Zhang *et al.,* 2022a; Salman *et al.,* 2024). The method reduces variance, improves generalization, and effectively captures both linear and nonlinear relationships in complex geospatial datasets without prior assumptions about data distribution (Talebi *et al.,* 2022). Additionally, RF has relatively low sensitivity to variable encoding, allowing diverse numerical predictor datasets to be processed with minimal preprocessing requirements. This characteristic is particularly beneficial for fisheries and oceanographic applications that integrate multiple environmental variables derived from heterogeneous EO products (Song *et al.,* 2023b). RF can also efficiently handle high-dimensional datasets while providing feature-importance estimates that facilitate interpretation of the relative contribution of individual environmental variables to model performance (Kawaguchi *et al.,* 2024). Compared with conventional statistical methods for fish catch prediction (Smoliński, 2017), RF has shown superior performance in handling complex geospatial datasets and achieving higher predictive accuracy (Lattanzi *et al.,* 2026).

In this study, model accuracy and overfitting control were evaluated through internal cross-validation using out-of-bag (OOB) error estimation combined with the 70/30 train–validation split described in Section 2.3. Since both procedures obtained training and test data from the same geographic domain without explicit spatial blocking, nearby fishing points could occur in the subsets. Consequently, the reported performance metrics may overestimate predictive accuracy relative to a spatially independent validation framework and should be interpreted as an upper bound of model performance (Section 4). Although RF is less interpretable than single decision trees, particularly when multiple environmental interactions are included (Wen et al., 2024), systematic hyperparameter optimization was performed to improve robustness and predictive capability. Before model development, essential parameters such as the number of trees (n_estimators), maximum tree depth (max_depth), minimum samples required to split a node (min_samples_split), and bootstrap sampling (bootstrap) were optimized to reduce the risks of overfitting and underfitting while enhancing model generalization (Gladju *et al.,* 2022; Hanif *et al.,* 2026). The optimal configuration consisted of n_estimators = 100, min_samples_split = 10, max_depth = 3, and bootstrap = True. These parameter values were used throughout model training and prediction. The optimized configuration improved predictive stability and generalization for both training and previously unseen data (Nikparvar & Thill, 2021). To capture spatiotemporal variability in tuna distribution, the reference dataset was analyzed under two configurations comprising four seasonal models corresponding to the monsoonal periods (DJF, MAM, JJA, and SON) and one annual model incorporating all reference points. This method generated five prediction maps, providing a comprehensive representation of seasonal and annual PFZs in the western waters of Sumatra. Since all models were developed from observations collected during a single year (2022), the predictions characterize environmental–catch relationships under the oceanographic conditions of that year rather than generalizable multi-year climatological patterns (Sections 4 and 5).

2.7. Model Evaluation and Accuracy Assessment

Model performance was evaluated using multiple statistical metrics, including test accuracy, recall, F1-score, RMSE, and MAE, to ensure robust and reliable prediction of PFZs. Classification metrics such as accuracy, precision, recall, and F1-score were used to evaluate the binary presence/background prediction of fishing suitability. Meanwhile, RMSE and MAE were additionally reported to quantify error in the underlying predicted probability values, providing a complementary continuous measure of model fit. During the exploratory phase, correlation analysis was conducted to examine interrelationships and multicollinearity among environmental variables, providing diagnostic insight before model training (Atwa *et al.,* 2024). Feature importance analysis was also applied to quantify the relative contribution of each oceanographic variable, thereby identifying key environmental drivers influencing the spatial distribution of yellowfin tuna and enhancing model interpretability. Classification performance was further assessed using a confusion matrix to derive class-level and overall accuracy metrics (Farhadpour *et al.,* 2024). However, AUC-ROC and precision–recall curves, which would provide a threshold-independent assessment of probabilistic performance, were not computed in this study and were recommended for future evaluations. For the classification metrics, the F1-score was used to evaluate the balance between precision and recall. Prediction errors were quantified using RMSE and MAE to capture both systematic and random deviations between observed and modeled outputs. Because model training and evaluation did not use spatially independent (blocked) cross-validation, the reported accuracy, precision, recall, and error metrics could be optimistic relative to predictive performance at genuinely independent locations. This methodological limitation is discussed further in Section 4. Model implementation and performance evaluation were conducted using the Scikit-Learn library in Python 3, ensuring methodological consistency and reproducibility (Miranda *et al.,* 2023).

3. Results and Discussion

3.1. Correlation and Importance of Environmental Variables in PFZ Prediction

Correlation analysis of the marine environmental variables showed several strong relationships among key oceanographic parameters (Figure 3), providing insight into the interactions between the physical and biological processes that structure pelagic habitats in the study area. The strongest association occurred between Chl-a and NPP (r = 0.87), indicating a close ecological relationship between phytoplankton biomass and carbon fixation in the water column. Due to this strong relationship, both variables contain substantially overlapping information, thereby increasing the possibility of multicollinearity when used simultaneously as predictors in RF. This effect is partially mitigated by GRRF feature-selection procedure (Section 2.5), which penalizes redundant predictors. Consequently, the importance of Chl-a and NPP should be interpreted jointly rather than as fully independent contributions. These results are consistent with established ecological theory, where Chl-a serves as a proxy for standing phytoplankton biomass, while NPP represents photosynthetic production, causing a greater primary productivity in waters with elevated phytoplankton concentrations (Cloern *et al.,* 2014; Fontaine *et al.,* 2025). In addition to the strong relationship between the biological variables, SST and SSS showed a moderate positive correlation (r = 0.67). This indicated coherent water-mass characteristics associated with monsoon-driven mixing, stratification, and regional advection. However, relationships between the biological variables and vertical current velocity were weak (Chl-a: r = 0.19; NPP: r = 0.11), suggesting that primary productivity was influenced more strongly by surface hydrographic conditions and light availability compared to vertical mixing alone (Dalabehara & Sarma, 2021). Horizontal current velocity also showed only a weak correlation with salinity (r = 0.08), indicating a limited influence of horizontal advection on salinity patterns at the spatial scale examined. These results show that biological variables, particularly Chl-a and NPP, are more closely related than the physical variables, while current velocities exert a comparatively weaker influence on environmental variability. The observed pattern suggests that habitat suitability for pelagic fisheries in the western waters of Sumatra is primarily regulated by surface hydrographic conditions and phytoplankton availability (Lima *et al.,* 2022).

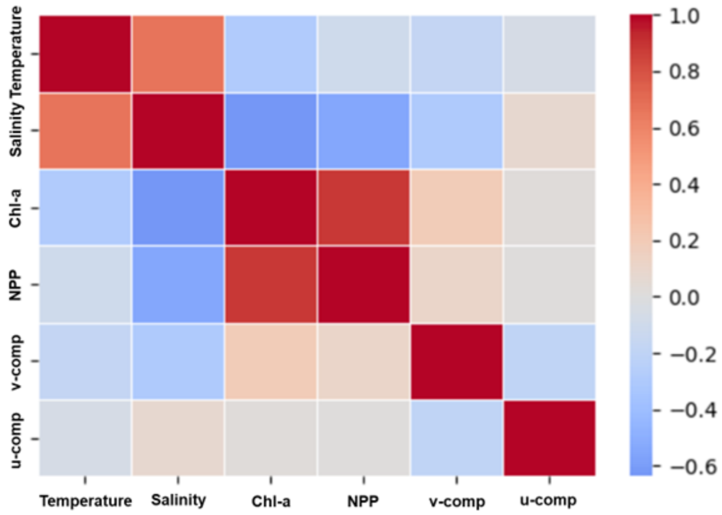

Figure 3. Correlation Matrix of Oceanographic Variables.

Understanding the relative contribution of environmental predictors is crucial for identifying the ecological drivers influencing the spatial distribution of tuna fishing grounds in dynamic tropical marine systems. The feature importance analysis (Figure 4) shows that oceanographic variables contribute unequally to model performance. This indicates that RF model captures habitat suitability and underlying ecological structures within the dataset. Based on the results, NPP is identified as the most influential predictor (importance = 0.34), underscoring its role as a primary indicator of food availability. NPP integrates nutrient supply and photosynthetic carbon fixation, making it a robust proxy for prey abundance that supports pelagic predators such as yellowfin tuna (Letscher *et al.,* 2023). SST ranked second (0.23), showing the importance of thermal conditions in regulating tuna physiology, metabolism, vertical movement, and prey distribution, thereby acting as a key determinant of habitat selection (Chen, 2022). Salinity also showed a substantial contribution (0.19), reflecting the influence of water-mass characteristics on habitat suitability.

Chl-a contributed moderately (0.14), while vertical (0.07) and horizontal (0.02) current velocities showed smaller but non-negligible effects. These importance values represented single-model point estimates from one training run, although formal uncertainty quantification such as bootstrapped importance scores across repeated RF runs was not performed. Therefore, relative ranking of closely valued predictors, such as Chl-a against current velocity, should be interpreted with caution. The lower importance of current components suggests that circulation processes influence nutrient transport, plankton aggregation, and frontal dynamics. However, the impact of circulation processes on fishing ground suitability is significantly indirect compared to biologically driven variables. These results indicate that biological productivity, particularly NPP, in combination with key physical drivers such as SST and SSS, collectively influence the spatial distribution of potential fishing grounds, with primary production acting as the dominant constraint on habitat quality for higher trophic levels (Franzè *et al.,* 2023).

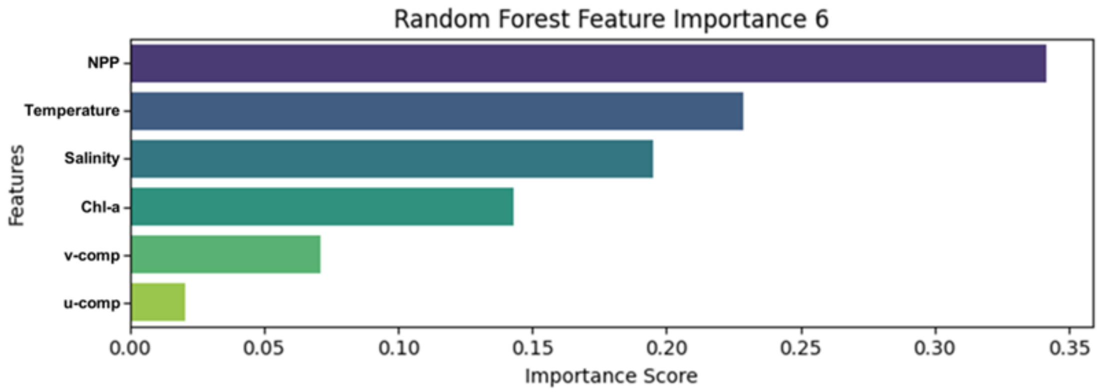

Figure 4. Variable Importance for Predicting Potential Fishing Grounds.

3.2. Seasonal Dynamics of Marine Environmental Conditions and Tuna Distribution

The spatial distributions of mean SST, SSS, and NPP during 2022 showed significant seasonal variability across the western waters of Sumatra (Figure 5), indicating the combined influence of monsoonal forcing, regional circulation, and upper-ocean processes on the physical–biogeochemical environment supporting pelagic fish. Based on the results, SST ranged from 29.0 to 30.0°C, with warmer conditions during MAM (March–May) and JJA (June–August), particularly along the coastal shelf and the Mentawai Island chain. This warming during the southeast monsoon is consistent with weakened convective mixing and reduced entrainment of cooler subsurface waters, leading to enhanced surface stratification (Clift *et al.,* 2022). In comparison, cooler SST during DJF (December–February) and SON (September–November) are associated with stronger convection and vertical mixing under the northwest monsoon regime. These seasonal thermal variations have been shown to influence habitat compression and the vertical distribution of tuna within the upper euphotic zone, thereby affecting accessibility to surface fishing gear (Matsubara *et al.,* 2024; Aoki *et al.,* 2025). Seasonal salinity varied between 32.0 and 34.0 PSU, with the lowest values observed during DJF in response to increased rainfall and terrestrial runoff. Higher salinity during JJA and SON corresponded to reduced precipitation and enhanced offshore advection (Zhang *et al.,* 2022b). These patterns are consistent with the alternating influence of the SJCC and ITF in regulating regional water-mass exchange (Purba *et al.,* 2021; Li *et al.,* 2023). Since the descriptions are based on a single annual cycle (2022), the observed differences should be regarded as indicative rather than statistically confirmed. Therefore, formal statistical comparisons, such as ANOVA or non-parametric alternatives, were not performed (Section 5).

NPP varied between 200 and 500 mg C m^{-2} day^{-1}, showing pronounced seasonal contrasts across the western Sumatra waters. High values of NPP during DJF and SON were primarily concentrated along the continental slope and offshore regions, consistent with periods of enhanced nutrient supply and upwelling-related processes. In comparison, lower NPP during MAM and JJA was consistent with warmer and more stratified surface conditions, indicating reduced vertical nutrient flux. The correspondence between seasonal NPP enhancement and large-scale circulation signatures suggests that basin-scale dynamics play a major role in structuring spatial bioproductivity gradients, in line with observations from the southeastern Indian Ocean (Navarro *et al.,* 2025). In this study, the spatial distribution of yellowfin tuna catch points showed a strong association with moderate SST of 29.2–29.8°C and elevated NPP of >350 mg C m^{-2} day^{-1}. The observed pattern suggests that yellowfin tuna preferentially occupy thermally favorable and highly productive habitats (Debiyanti *et al.,* 2025). This observation confirms the ecological principle that pelagic predators track regions where physical forcing enhances prey availability, supporting the value of bio-physical indicators for ML–based fisheries habitat modeling. However, the relationship is descriptive rather than statistically confirmed because formal spatial analyses are not performed. Furthermore, the fisheries-dependent catch records reflect the spatial distribution of fishing effort in addition to tuna occurrence, and the potential influence of effort-related bias is not explicitly modeled. The methodological limitations are discussed further in Section 4.

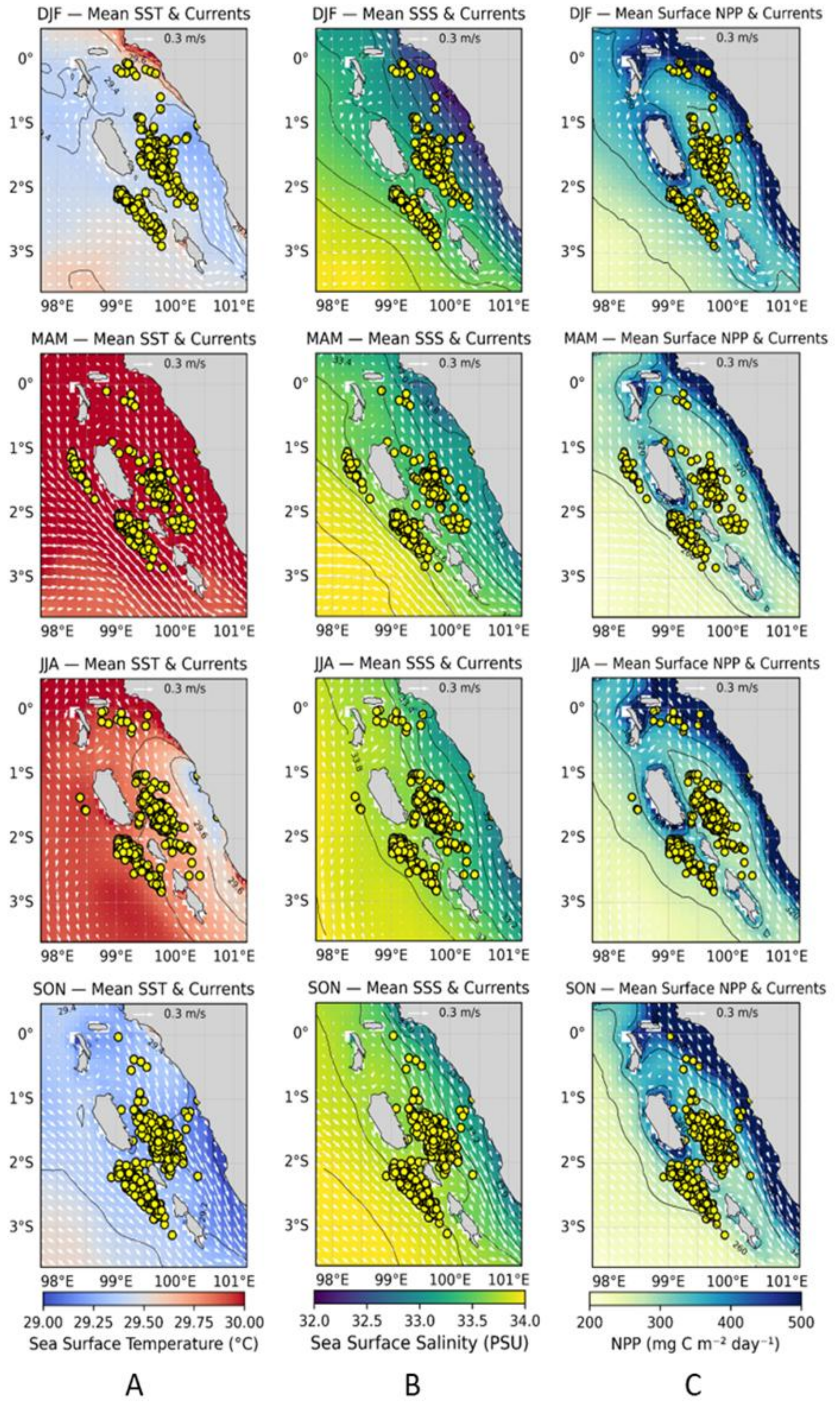

Figure 5. Seasonal variation of oceanographic conditions and yellowfin tuna (*Thunnus albacares*) catch locations in the western waters of Sumatra, Indonesia. Each row represents a monsoonal period: DJF

(December–February), MAM (March–May), JJA (June–August), and SON (September–November). The maps illustrate the mean distributions of (A) sea surface temperature (SST, °C), (B) sea surface salinity (SSS, PSU), and (C) net primary productivity (NPP, mg C m^{-2} day^{-1}), overlaid with averaged surface current vectors (m s^{-1}) and fishing reference points (yellow dots).

Temporal analysis further showed that intra-annual variability in yellowfin tuna catches during 2022 closely tracked seasonal fluctuations in SST, SSS, and NPP (Figure 6). Catch levels peaked during April–June and October–December, coinciding with the monsoon transition periods characterized by elevated primary productivity and favorable thermal conditions. This seasonal association is consistent with the dynamics of monsoon-dominated marine systems, where mixed-layer processes regulate nutrient availability and prey production (Ducklow *et al.,* 2022). In comparison, lower catches during periods of elevated SST (>30 °C) and reduced productivity, particularly during JJA, suggest a redistribution of yellowfin tuna to maintain favorable thermal conditions and access prey resources (Waller *et al.,* 2024). Seasonal boxplot analysis further supports these observations by showing higher median catches during DJF and SON under moderate SST and elevated NPP. Meanwhile, MAM and JJA have lower catch levels despite relatively stable salinity conditions. Seasonal salinity ranged from 32.8 to 34.0 PSU, with slightly higher values during JJA and SON, reflecting the influence of monsoon-driven circulation. Since the temporal patterns are derived from a single annual cycle (2022), the observed seasonal relationships should be regarded as indicative rather than representative of long-term climatological variability. These results indicate that yellowfin tuna distribution is jointly regulated by biological productivity and monsoon-driven hydrodynamics. Consequently, predictable seasonal environmental forcing provides a strong foundation for ML–based operational prediction of PFZs in the western waters of Sumatra (Ismail & Al-Shehhi, 2023).

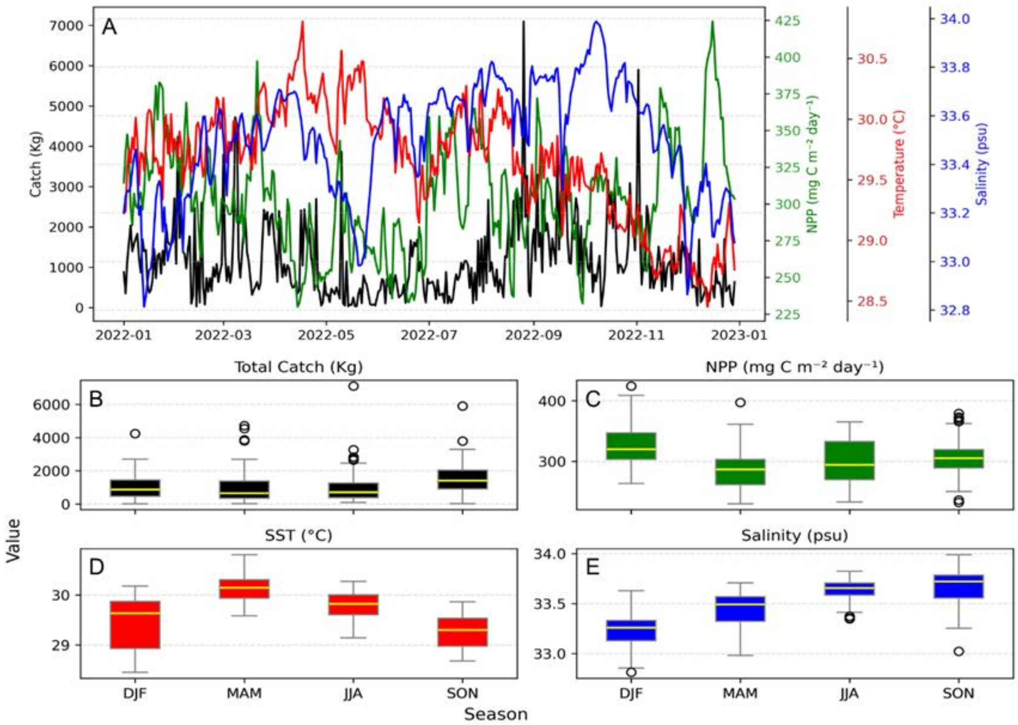

Figure 6. Temporal variation of yellowfin tuna (*Thunnus albacares*) catches and key oceanographic parameters in the western Sumatra waters during 2022. (A) Time series of total tuna catch (black line) overlaid with sea surface temperature (SST, red), salinity (SSS, blue), and net primary productivity (NPP, green). (B–E) Seasonal boxplots showing the distribution of total catch (B), NPP (C), SST (D), and SSS (E) across four monsoonal periods: DJF (December–February), MAM (March–May), JJA (June–August), and SON (September–November).

3.3. Model Performance Evaluation and Spatial Prediction of PFZs

Evaluation of statistical and ML performance is essential for assessing model generalization beyond the training dataset and the ability to capture the nonlinear ecological relationships that characterize pelagic fisheries environments. In this study, RF evaluation showed strong predictive capability for identifying potential fishing areas from oceanographic variables (Figure 7), and the corresponding performance metrics are summarized in Table 3. The classifier showed high discriminative ability, indicating that the environmental variability contained sufficient spatial and temporal structure to support reliable learning of habitat patterns associated with yellowfin tuna. Furthermore, the annual model achieved an accuracy of 0.91, precision of 0.86, recall of 0.97, and an F1-score of 0.91, with low prediction errors (RMSE = 0.29; MAE = 0.08). As discussed in Sections 2.6 and 2.7, these metrics were derived without spatially independent validation and might overestimate predictive performance at genuinely independent locations. Therefore, the results should be interpreted as measures of relative model fit under the current train–validation design rather than as unbiased estimates of operational predictive skill. The high recall further indicates that the classifier effectively identified true positive fishing grounds, a characteristic that is particularly valuable in operational fisheries applications where missed detections have greater consequences than false-positive predictions.

Table 3. Modeling Performance Report with Random Forest (RF) Prediction.

Periods	Test Accuracy	Precision	Recall	F1-Score	RMSE	MAE
DJF	0.91	0.88	0.96	0.92	0.28	0.08
MAM	0.87	0.80	0.97	0.88	0.35	0.12
JJA	0.90	0.87	0.96	0.91	0.30	0.09
SON	0.92	0.86	0.99	0.92	0.28	0.07
Yearly	0.91	0.86	0.97	0.91	0.29	0.08

Seasonal performance was similarly consistent, with the highest accuracy obtained during SON at 0.92 and recall at 0.99, while the lowest accuracy was obtained during MAM at 0.87 with a precision of 0.80. The reduced precision during MAM indicates high spatial noise and low environmental contrast, particularly when stratification and productivity gradients are weaker compared to transitional monsoon periods. Despite slight seasonal variations, the model showed reliable classification capability and stable predictive accuracy throughout the year. This stability across seasons shows the robustness of RF in handling heterogeneous and multivariate oceanographic datasets, outperforming many conventional statistical methods when dealing with high-dimensional nonlinear processes. Furthermore, the performance consistency strengthens the feasibility of adopting RF-based PFZ models for continuous or near-real-time operational fisheries forecasting in dynamic Indonesian waters, particularly where environmental signals related to tuna presence show strong seasonal modulation.

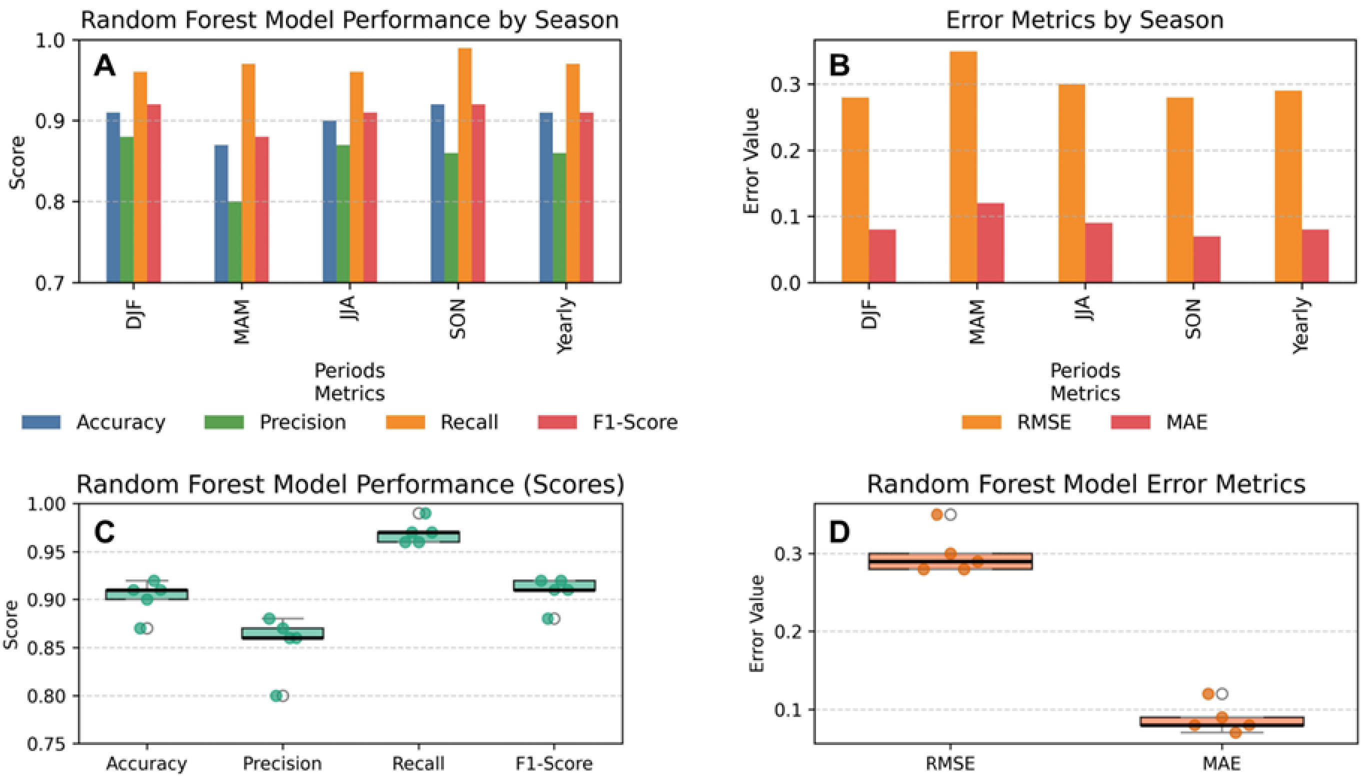

Figure 7. Performance evaluation of the Random Forest model across seasonal and annual periods. Panels A and B present bar charts showing classification metrics (Accuracy, Precision, Recall, and F1-Score) and error metrics (RMSE and MAE) for DJF, MAM, JJA, SON, and Yearly datasets. Panels C and D show the distribution of the same performance and error metrics using boxplots to highlight variability and consistency across model runs. Overall, the model demonstrates stable performance with high classification scores and relatively low error values, indicating robust predictive capability across seasons.

Spatial interpretation of the predicted PFZs is essential to evaluate the correlation of the modeled habitat suitability with regional oceanographic processes and the ecological behavior of yellowfin tuna. The seasonal prediction maps (Figure 8) for DJF, MAM, JJA, SON, and the annual composite show the combined influence of primary productivity, SST, salinity, and current structure on habitat suitability in the western waters of Sumatra. Based on observations, high-probability fishing zones were concentrated mainly along the continental slope and offshore frontal areas. This indicates the importance of mesoscale circulation features, including upwelling cells, current shear zones, and retention structures, that enhance prey aggregation and create energetically favorable feeding habitats for large pelagic predators (Braun *et al.,* 2022). The prediction maps represent spatial patterns of habitat suitability but do not include measures of prediction uncertainty or confidence. Therefore, future applications should incorporate pixel-level uncertainty estimates, such as the variance of predictions across RF trees, to strengthen confidence in operational PFZ forecasting.

Seasonal differences in habitat suitability were evident, with DJF and SON showing the most extensive high-suitability areas, corresponding to elevated NPP and greater observed catches. In comparison, MAM showed more fragmented and spatially restricted suitable habitats, indicating stronger water-column stratification and reduced nutrient supply that limit prey availability. The annual prediction identified persistent core habitats that remained favorable across the monsoonal cycle, suggesting the presence of stable oceanographic features such as eddy boundaries and coastal–current interaction zones. These persistent habitats represent priority fishing grounds for monitoring and sustainable management, emphasizing the value of GeoAI-based predictions for supporting long-term fisheries management in the western waters of Sumatra.

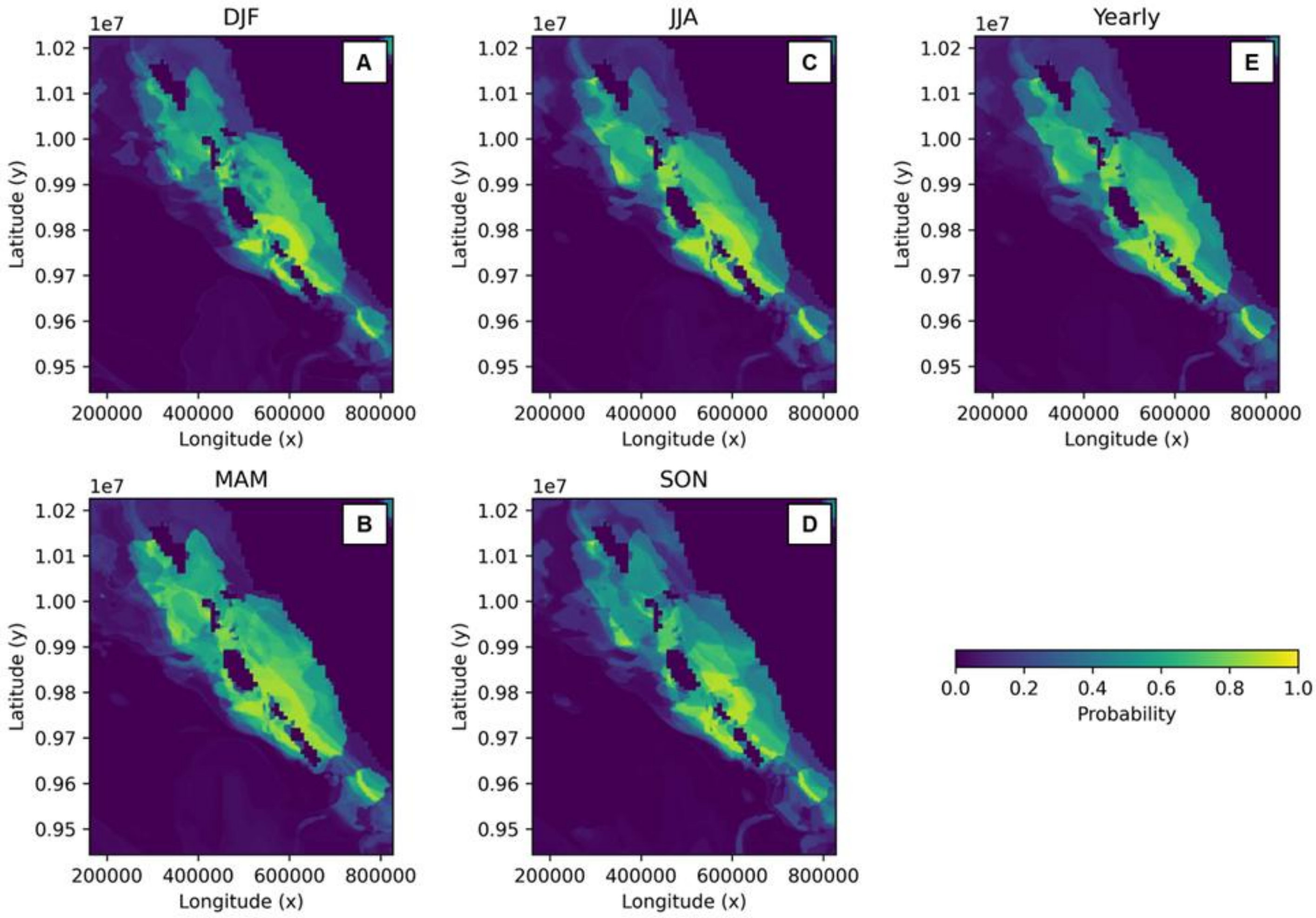

Figure 8. Spatial prediction maps of seasonal fishing grounds during (A) DJF, (B) MAM, (C) JJA, (D) SON, and (E) Yearly derived from the Random Forest model using multi-parameter oceanographic inputs.

4. Discussion

The results show that GeoAI-based ML methods, particularly GRRF and RF, are effective in identifying and predicting PFZs for yellowfin tuna in the highly dynamic western Sumatra waters.

The high predictive accuracy achieved by both models shows their robustness in integrating multiple oceanographic variables and capturing nonlinear ecological relationships that are often inadequately represented by conventional statistical methods. This capability is particularly important in monsoon-influenced systems, where ecological responses emerge from complex interactions and threshold effects rather than linear drivers.

Among the environmental predictors, NPP and Chl-a contributed most strongly to tuna occurrence. This is consistent with previous results in the Indian Ocean that identify productivity-related variables as reliable indicators of feeding habitats (Heltria *et al.,* 2024; Zhang *et al.,* 2023). The strong correlation between NPP and Chl-a (r = 0.87) shows the central role of phytoplankton biomass and photosynthetic activity in sustaining lower trophic levels that indirectly regulate tuna distribution through prey availability. As discussed in Section 3.1, this correlation also indicates that high individual contributions of NPP and Chl-a to model predictions cannot be fully disentangled. The observed preference for moderate SST (29–29.8°C) further supports the importance of optimal thermal conditions in enhancing prey aggregation and metabolic efficiency (Sambah *et al.,* 2021). However, extreme SST conditions can reduce catch probability by altering stratification and nutrient fluxes (Satar *et al.,* 2024).

Although salinity and current velocity contributed less to model predictions, both variables remained ecologically significant. Spatial clustering of tuna along regions influenced by the SJCC and the ITF shows the role of mesoscale mixing and localized upwelling in enhancing nutrient availability and biological productivity (Xing *et al.,* 2023). Furthermore, seasonal displacement of PFZs across monsoonal phases indicates that habitat suitability is dynamically modulated by circulation variability rather than being spatially fixed. Higher catches during April–June and October–December coincide with transitional monsoon periods when vertical mixing intensifies and SST–NPP interactions reach optimal balance, supporting the monsoon–productivity correlation mechanisms described by Utari *et al.* (2019) and Fadhilah *et al.* (2022). However, several high-probability zones, particularly along offshore frontal boundaries, remain underexploited, suggesting a spatial mismatch between resource availability and fishing effort.

Comparison of the two ML methods showed that GRRF slightly outperformed standard RF in predictive stability and variable selection efficiency. By penalizing redundant predictors, GRRF improved interpretability without compromising accuracy. This serves as an advantage in marine ecological modeling where multicollinearity among environmental variables is common (Panzeri *et al.,* 2021; Pourzangbar *et al.,* 2023). Furthermore, the integration of SST, salinity, Chl-a, NPP, and current velocity enabled the GeoAI framework to capture synergistic physical–biological interactions shaping tuna distribution. The visualization of probabilistic PFZ maps combined with environmental contours and spatial bounding indicators (Figure 9) also enhances interpretability for operational fisheries applications by explicitly showing where key ecological drivers converge.

Despite the significant contribution of this study, several methodological limitations should be considered when interpreting the results. First, the analysis was based on a single year (2022) of fisheries and environmental data, capturing the seasonal variability observed during the fishing period. Consequently, the analysis did not account for interannual variability associated with large-scale climate models such as the El Niño–Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD), which influence productivity and yellowfin tuna availability in the area (Fadhilah *et al.,* 2022). Second, the fisheries-dependent catch records partly showed the spatial distribution of handline fishing effort, gear type, and fisher behavior rather than tuna occurrence alone. Due to this limitation, the source of sampling bias was not explicitly quantified or corrected in the analysis. Third, model training and evaluation did not incorporate spatially independent validation, leading to performance overestimation by reported accuracy metrics (Table 3) at unsampled zones. These limitations indicate that the proposed PFZ maps should be regarded as a proof-of-concept demonstration of the GRRF–RF framework for the western waters of Sumatra rather than as an operationally validated forecasting product. Therefore, future studies are recommended to incorporate multi-year datasets, fisheries-independent observations, additional environmental predictors, and spatially explicit validation frameworks to improve model robustness and operational reliability.

This study indicates that the western Sumatra region shows favorable conditions for yellowfin tuna aggregation, particularly along the Mentawai slope and adjacent offshore waters where productivity and hydrodynamic mixing are strongest. The GRRF–RF–derived PFZ maps provide spatially explicit guidance that can improve fishing efficiency by reducing search time and fuel consumption. Additionally, the identification of persistently suitable but underutilized zones shows opportunities for adaptive spatial expansion of fishing activities and optimized fleet routing. These results offer valuable inputs for spatially informed management strategies supporting sustainable tuna fisheries in Indonesia's eastern Indian Ocean, while emphasizing that further model refinement and independent validation are required before operational implementation.

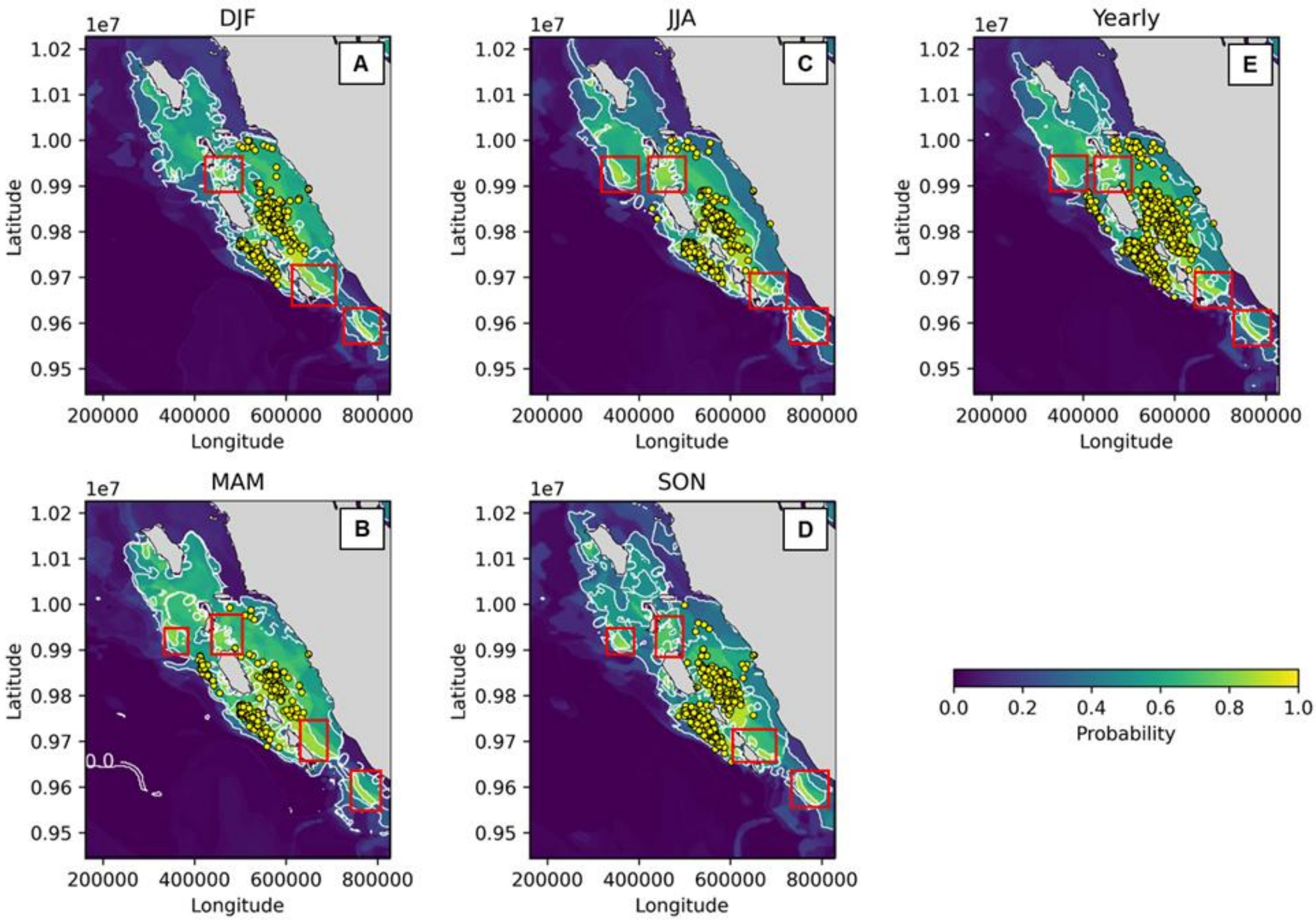

Figure 9. Spatial prediction maps of seasonal fishing grounds during (A) DJF, (B) MAM, (C) JJA, (D) SON, and (E) Yearly derived from the Random Forest model and overlaid with fishing catch locations (yellow points), environmental contour lines, and identified high-probability regions (red bounding boxes) that remain underutilized by current fishing activities.

5. Conclusion

In conclusion, this study successfully applies a GeoAI-based ML framework to predict PFZs for yellowfin tuna in the western waters of Sumatra. By integrating GRRF and RF algorithms with multi-source marine environmental data, the analysis achieves strong predictive performance and identifies key environmental drivers of tuna distribution. However, the reported accuracy is subject to sampling and validation limitations, which require confirmation through spatially independent and multi-year validation before being treated as a generalizable measure of model skill. Among multi-source oceanographic variables, Chl-a and NPP are dominant predictors, showing the ecological importance of productivity-driven processes. SST and current dynamics modulate habitat suitability through their influence on thermal structure and nutrient availability. These results confirm the hypothesis that pelagic predators in monsoon-influenced tropical systems primarily respond to bottom-up ecological controls, where primary productivity supports localized prey enhancement influenced by hydrodynamic forcing.

The results provide initial insights into the capacity of GeoAI methods to represent complex, nonlinear relationships between environmental variability and tuna habitat suitability, extending previous studies in the region that depend largely on empirical or correlation-based methods. The single-year scope of the dataset and the absence of spatially independent validation constrain the generalizability of the results. This indicates that the data obtained should be considered an initial, region-specific demonstration of the GRRF–RF framework rather than a validated operational forecasting tool. From a practical perspective, PFZ maps contribute to more eco-efficient fishing strategies and support sustainable resource use, particularly for artisanal and small-scale fisheries with limited oceanographic information. Therefore, future studies are recommended to expand the temporal domain through multi-year observations to account for interannual climate variability, including the ENSO and the IOD. Spatially blocked cross-validation and threshold independent performance metrics, such as AUC–ROC, should also be incorporated to provide more rigorous model evaluation. Additionally, future studies should explicitly account for fishing-effort bias in fisheries-dependent catch data and explore hybrid deep learning architectures to enhance spatiotemporal generalization. Integrating these GeoAI-based predictions with near–real-time satellite observations and fisheries monitoring data would serve as a promising direction toward developing adaptive, environmentally informed decision-support tools for sustainable tuna fisheries management in Indonesia and the wider Indian Ocean zone.

Acknowledgements

This study used multi-parameter Earth Observation (EO) datasets provided by the Copernicus Marine Environment Monitoring Service (CMEMS), which supported the oceanographic analysis. Historical fishing records of yellowfin tuna (Thunnus albacares) were obtained from the Bungus Ocean Fishing Port (PPS Bungus). The authors are grateful to the anonymous reviewers for their constructive comments and suggestions, which contributed to the improvement of this study.

Author Contributions

Conceptualization: Yuliardi, A.Y., Hanif, M.; **methodology**: Yuliardi, A.Y., Hanif, M., Heltria. S.; **investigation**: Rahmalia, D. A., Napitupulu, G.; **writing—original draft preparation**: Yuliardi, A.Y., Hanif, M., Napitupulu, G.; **writing—review and editing**: Heltria. S., Rahmalia, D. A.; **visualization**: Yuliardi, A.Y., Hanif, M.. All authors have read and agreed to the published version of the manuscript.

Conflict of interest

All authors declare that they have no conflicts of interest.

Data availability

Data is available upon Request.

Funding

This research received no external funding.

References

Abdi, A. M. (2020). Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. *GIScience & Remote Sensing,* 57(1), 1-20. doi: 10.1080/15481603.2019.1650447

Ahmad, H., Jose, F., Bhuyan, M. S., Islam, M. N., & Dash, P. (2024). Seasonal influence of freshwater discharge on spatio-temporal variations in primary productivity, sea surface temperature, and euphotic zone depth in the northern Bay of Bengal. *Acta Oceanologica Sinica,* 43(6), 1-14. doi: 10.1007/s13131-023-2254-y

Al Iqromi, M. F., Harahap, A., Heltria, S., Ramdhani, F., Gelis, E. R. E., & Syarifuddin, H. (2025). Study on The Distribution of Longtail Tuna (Thunnus Tonggol) Catch Based on Chlorophyll-A Using Aqua Modis Imagery in West Sumatera. *Grouper,* 16(1), 189-203.

Aoki, Y., Kitagawa, T., & Kiyofuji, H. (2025). Surfacing and diving behavior associated with thermal physiology in oceanic habitats of skipjack tuna (Katsuwonus pelamis) in the western north Pacific Ocean. *Frontiers in Physiology,* 16, 1462940. doi: 10.3389/fphys.2025.1462940

Atwa, W., Almazroi, A. A., & Ayub, N. (2024). Reliable renewable energy forecasting for climate change mitigation. *PeerJ Computer Science, 10*, e2067. doi: 10.7717/peerj-cs.2067

Boutayeb, A., Lahsen-Cherif, I., & El Khadimi, A. (2025). When Machine Learning Meets Geospatial Data: A Comprehensive GeoAI Review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 18. doi: 10.1109/JSTARS.2025.3568715

Braun, C. D., Arostegui, M. C., Thorrold, S. R., Papastamatiou, Y. P., Gaube, P., Fontes, J., & Afonso, P. (2022). The functional and ecological significance of deep diving by large marine predators. *Annual Review of Marine Science,* 14(1), 129-159. doi: 10.1146/annurev-marine-032521-103517

Breiman, L. (2001). Random forests. *Machine Learning,* 45(1), 5–32. doi: 10.1023/A:1010933404324

Brown, A., Purwati, P., & Isnaniah, I. (2023). Availability of fish catch production yellowfin tuna (Thunnus albacares) landed in the Bungus Ocean Fishery Port of West Sumatra Province. *South East Asian Marine Sciences Journal,* 1(1), 1-7.

Chabalala, Y., Adam, E., & Ali, K. A. (2023). Exploring the effect of balanced and imbalanced multi-class distribution data and sampling techniques on fruit-tree crop classification using different machine learning classifiers. *Geomatics,* 3(1), 70-92. doi: 10.3390/geomatics3010004

Chen, X. (2022). Shoaling and migration of fish and their relationships with environment. In Theory and method of fisheries forecasting. *Singapore: Springer Nature Singapore,* 39-85. doi: 10.1007/978-981-19-2956-4_3

Clift, P. D., Betzler, C., Clemens, S. C., Christensen, B., Eberli, G. P., France-Lanord, C., ... & Wan, S. (2022). A synthesis of monsoon exploration in the Asian marginal seas. *Scientific Drilling,* 31, 1-29. doi: 10.5194/sd-31-1-2022

Cloern, J. E., Foster, S. Q., & Kleckner, A. E. (2014). Phytoplankton primary production in the world's estuarine-coastal ecosystems. *Biogeosciences,* 11(9), 2477-2501. doi: 10.5194/bg-11-2477-2014

Dalabehara, H. B., & Sarma, V. V. S. S. (2021). Physical forcing controls spatial variability in primary production in the Indian Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography,* 183, 104906. doi: 10.1016/j.dsr2.2020.104906

Debiyanti, D., Gunawan, D., & Budiyanto, S. (2025). Patterns of oceanographic factor distribution and tuna fishing potential: Spatial and temporal analysis. *Remote Sensing Technology in Defense and Environment,* 2(1), 60-77. doi: 10.61511/rstde.v2i1.2025.1196

Ducklow, H., Cimino, M., Dunton, K. H., Fraser, W. R., Hopcroft, R. R., Ji, R., ... & Sosik, H. M. (2022). Marine pelagic ecosystem responses to climate variability and change. *BioScience,* 72(9), 827-850. doi: 10.1093/biosci/biac050

Fadhilah, A., Leidonald, R., Susetya, I. E., & Dewinta, A. F. (2022). Analysis of the relationship between the ENSO phenomenon, Net Primary Productivity and catches of yellowfin tuna in Sibolga Waters, Indonesia. *Biodiversitas Journal of Biological Diversity,* 23(9).

Farhadpour, S., Warner, T. A., & Maxwell, A. E. (2024). Selecting and interpreting multiclass loss and accuracy assessment metrics for classifications with class imbalance: Guidance and best practices. *Remote Sensing,* 16(3), 533. doi: 10.3390/rs16030533

Fontaine, D. N., Marrec, P., Menden-Deuer, S., Sosik, H. M., & Rynearson, T. A. (2025). Time series of phytoplankton net primary production reveals intense interannual variability and size-dependent chlorophyll-specific productivity on a continental shelf. *Limnology and Oceanography,* 70(1), 203-216. doi: 10.1002/lno.12749

Franzè, G., Anderson, S. I., Kling, J. D., Wilburn, P., Hutchins, D. A., Litchman, E., ... & Menden-Deuer, S. (2023). Interactive effects of nutrients and temperature on herbivorous predation in a coastal plankton community. *Limnology and Oceanography*, 68, S144-S157. doi: 10.1002/lno.12289

Gao, S., Hu, Y., & Li, W. (2023). Introduction to geospatial artificial intelligence (GeoAI). *Handbook of geospatial artificial intelligence*, 3-16. doi: 10.1201/9781003308423

Geerts, M., Vanden Broucke, S., & De Weerdt, J. (2024). GeoRF: a geospatial random forest. *Data Mining and Knowledge Discovery*, 38(6), 3414-3448. doi: 10.1007/s10618-024-01046-7

Gladju, J., Kamalam, B. S., & Kanagaraj, A. (2022). Applications of data mining and machine learning framework in aquaculture and fisheries: A review. Smart *Agricultural Technology,* 2, 100061. doi: 10.1016/j.atech.2022.100061

Hanif, M., Apichontrakul, S., & Razi, P. (2024). Surface deformation monitoring and forecasting of sinabung volcano using interferometry synthetic aperture radar and forest-based algorithm. *Remote Sensing Applications: Society and Environment,* 36, 101288. doi: 10.1016/j.rsase.2024.101288

Hanif, M., Apichontrakul, S., Suryani, R., & Ramadhan, R. (2026). Landslide susceptibility prediction on mount Marapi using Interferometric Synthetic aperture radar (INSAR) integrated with multi-model machine learning approaches. *Geography, Environment, Sustainability,* 19(2), 29-40. doi: 10.24057/2071-9388-2026-4016

Hao, R., Zhao, Y., Zhang, S., & Deng, X. (2025). *Deep Learning for Ocean Forecasting: A Comprehensive Review of Methods, Applications, and Datasets.* IEEE Transactions on Cybernetics. doi: 10.1109/TCYB.2025.3539990

Heltria, S., Ramdhani, F., & Yuliardi, A. Y. (2024). Suitability Of Spatial-Temporal Habitat Of Yellowfin Tuna (Thunnus Albacares) Using Maximum Entropy Model In West Sumatra. *Marine Fisheries: Journal of Marine Fisheries Technology and Management,* 15(2), 147-158. doi: 10.29244/jmf.v15i2.52940

Hosen, B., Rahaman, M., Kumar, S., Sagar, L., & Akhtar, M. N. (2023). Leveraging artificial intelligence and big data for advanced spatial analytics and decision support systems in geography. *Malaysian Applied Geography,* 1(2), 62-67. doi: 10.26480/magg.02.2023.62.67

Iranzad, R., & Liu, X. (2025). A review of random forest-based feature selection methods for data science education and applications. International Journal of Data Science and Analytics, 20(2), 197-211. https://doi.org/10.1007/s41060-024-00509-w

Islam, M. K., & Moontaha, T. J. (2025). Satellite-based monitoring of chlorophyll-a and sea surface temperature for identifying potential fishing zones (PFZs) for Hilsha fish: a case study in the Bay of Bengal. *Remote Sensing for Agriculture, Ecosystems, and Hydrology XXVII,* 13666, 371-388. doi: 10.1117/12.3070008

Ismail, K. A., & Al-Shehhi, M. R. (2023). Applications of biogeochemical models in different marine environments: a review. *Frontiers in Environmental Science,* 11, 1198856. doi: 10.3389/fenvs.2023.1198856

Izquierdo-Verdiguier, E., & Zurita-Milla, R. (2020). An evaluation of Guided Regularized Random Forest for classification and regression tasks in remote sensing. *International Journal of Applied Earth Observation and Geoinformation,* 88, 102051. doi: 10.1016/j.jag.2020.102051

John, S. S., Kulkarni, S. A., Begam, M. F., & Hiremath, A. (2025). Forecasting of Potential Fishing Zones in the Arabian Sea Using Satellite Imagery. In 2025 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT). IEEE, 544-549. doi: 10.1109/IAICT65714.2025.11101450

Kawaguchi, O. (2024). Machine-learning and thresholding algorithms to automatically predict fishing effort of small-scale trawl fishery. *Fisheries Science,* 90(2), 123-137. doi: 10.1007/s12562-023-01734-1

Lattanzi, P., Vasapollo, C., Samarão, J., Galdelli, A., Mendo, T., Rufino, M., ... & Tassetti, A. N. (2026). Predicting fishing vs. not-fishing in small-scale fisheries: A sample vessel tracking dataset and a reproducible machine learning approach. *SoftwareX,* 34, 102579. doi: 10.1016/j.softx.2026.102579

Lellouche, J. M., Greiner, E., Le Galloudec, O., Garric, G., Regnier, C., Drevillon, M., ... & Le Traon, P. Y. (2018). Recent updates to the Copernicus Marine Service global ocean monitoring and forecasting real-time 1/12∘ high-resolution system. *Ocean Science,* 14(5), 1093-1126. doi: 10.5194/os-14-1093-2018

Letscher, R. T., Moore, J. K., Martiny, A. C., & Lomas, M. W. (2023). Biodiversity and stoichiometric plasticity increase pico-phytoplankton contributions to marine net primary productivity and the biological pump. Global Biogeochemical Cycles, 37(8), e2023GB007756. https://doi.org/10.1029/2023GB007756

Li, J., Li, Y., Guo, Y., Li, G., & Wang, F. (2023). Decadal variability of sea surface salinity in the Southeastern Indian Ocean: Roles of local rainfall and the Indonesian throughflow. *Frontiers in Marine Science,* 9, 1097634. doi: 10.3389/fmars.2022.1097634

Lima, A. R., Garrido, S., Riveiro, I., Rodrigues, D., Angélico, M. M., Gonçalves, E. J., ... & Silva, G. (2022). Seasonal approach to forecast the suitability of spawning habitats of a temperate small pelagic fish under a high-emission climate change scenario. *Frontiers in Marine Science,* 9, 956654. doi: 10.3389/fmars.2022.956654

Lisna, Febryanti H. R., Heltria S., Toni Y., Ramadan F., Ramdhani F., 2024 Estimation of the fishing ground potential based on oceanography parameters (GIS-based) of yellowfin tuna (Thunnus albacares) in West Sumatra waters. *AACL Bioflux,* 17(6):3175-3189.

Matsubara, N., Aoki, Y., Aoki, A., & Kiyofuji, H. (2024). Lower thermal tolerance restricts vertical distributions for juvenile albacore tuna (Thunnus alalunga) in the northern limit of their habitats. *Frontiers in Marine Science,* 11, 1353918.

Miranda, F. M., Köhnecke, N., & Renard, B. Y. (2023). Hiclass: a python library for local hierarchical classification compatible with scikit-learn. *Journal of Machine Learning Research,* 24(29), 1-17. doi: 10.48550/arXiv.2112.06560

Nadeem, A., Rana, A. D., Batool, S. A., Azhar, A., Ishfaq, U. B. E., & Hameed, A. (2025). Satellite based assessment of potential fishing zones (PFZs) within the exclusive economic zone (EEZ) of Pakistan. *Continental Shelf Research,* 286, 105410. doi: 10.1016/j.csr.2025.105410

Navarro, J., Lonin, S., Linero-Cueto, J., & Romero-Balcucho, C. (2025). Hydrodynamic Modelling of the Guajira Upwelling System (Colombia). *Applied Sciences,* 15(20), 11000. doi: 10.3390/app152011000

Nóbrega, M. F., de Lira, M. G., Oliveira, M. A., & Oliveira, J. E. L. (2023). Interactions between oceanographic variables and population structure of the yellowfin tuna Thunnus albacares (Bonnaterre, 1788) in the Western Central Atlantic. *Fisheries Oceanography,* 32(2), 213-228. doi: 10.1111/fog.12624

Ni, B., & Wang, X. (2024). Geographic research in the AI era: Applications, methods, and prospects. *Geographical Research Bulletin,* 3, 103-124. doi: 10.50908/grb.3.0_103

Nikparvar, B., & Thill, J. C. (2021). Machine learning of spatial data. *ISPRS International Journal of Geo-Information,* 10(9), 600. doi: 10.3390/ijgi10090600

Nurholis., Gaol, J. L., Syah, A. F., & Dewi, A. K. (2020). GIS-based spatio-temporal analysis on yellow fin tuna catch in Eastern Indian Ocean off Sumatera. *IOP Conference Series: Earth and Environmental Science*, 429(1), 012041. doi: 10.1088/1755-1315/429/1/012041

Nurholis., Sukoraharjo, S. S., Lumban-Gaol, J., & Syah, A. F. (2021). Yellowfin tuna fishing dynamic during low IOD positive and negative at eastern Indian ocean; study case at western waters of sumatera. In IOP Conference Series: *Earth and Environmental Science,* 695(1), 012010. doi: 10.1088/1755-1315/695/1/012010

Panzeri, D., Libralato, S., Carlucci, R., Cipriano, G., Bitetto, I., Spedicato, M. T., ... & Solidoro, C. (2021). Defining a procedure for integrating multiple oceanographic variables in ensemble models of marine species distribution. In 2021 International Workshop on Metrology for the Sea; Learning to Measure Sea Health Parameters (MetroSea). *IEEE*, 360-365. doi: 10.1109/MetroSea52177.2021.9611559

Pourzangbar, A., Jalali, M., & Brocchini, M. (2023). Machine learning application in modelling marine and coastal phenomena: a critical review. *Frontiers in Environmental Engineering,* 2, 1235557. doi: 10.3389/fenve.2023.1235557

Purba, N. P., Pranowo, W. S., Ndah, A. B., & Nanlohy, P. (2021). Seasonal variability of temperature, salinity, and surface currents at 0 latitude section of Indonesia seas. *Regional Studies in Marine Science,* 44, 101772. doi: 10.1016/j.rsma.2021.101772

Rahaman, H., Srinivasu, U., Panickal, S., Durgadoo, J. V., Griffies, S. M., Ravichandran, M., ... & Wang, Q. (2020). An assessment of the Indian Ocean mean state and seasonal cycle in a suite of interannual CORE-II simulations. *Ocean Modelling,* 145, 101503. doi: 10.1016/j.ocemod.2019.101503

Sahib, W. M., Alhuseen, Z. A. A., Saeedi, I. D. I., Abdulkadhem, A. A., & Ahmed, A. (2025). Leveraging machine learning for enhanced cybersecurity: an intrusion detection system. *Service Oriented Computing and Applications,* 19(2), 107-124. doi: 10.1007/s11761-024-00435-6

Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random forest algorithm overview. *Babylonian Journal of Machine Learning,* 2024, 69-79. doi: 10.58496/BJML/2024/007

Sambah, A. B., Muamanah, A., Harlyan, L. I., Lelono, T. D., Iranawati, F., & Sartimbul, A. (2021). Sea surface temperature and chlorophyll-a distribution from Himawari satellite and its relation to yellowfin tuna in the Indian Ocean. *Aquaculture, Aquarium, Conservation & Legislation*, *14*(2), 897-909. http://www.bioflux.com.ro/docs/2021.897-909.pdf

Sarangi, R. K., Jishad, M., Sharma, R., Das, A., Mali, K., Ramalingam, L., ... & Prakash, C. (2024). Multiple ocean parameter-based potential fishing zone (PFZ) location generation and validation in the Western Bay of Bengal. *Environmental Monitoring and Assessment,* 196(1), 98. doi: 10.1007/s10661-023-12259-6

Sari, N. R., Hartoko, A., & Fitri, A. D. (2024). Potential big eye tuna (Thunnus obesus) fishing zone in western Sumatra Sea, Indonesia based on Sub-surface Temperature. *International Journal of Multidisciplinary Research and Growth Evaluation,* 5(2). doi: 10.54660/.IJMRGE.2024.5.2.72-76

Satar, M. N., Marshal, W., & Akhir, M. F. (2024). The Effects of Climate Change on Ocean Upwelling and Productivity in Regional Seas. In Coastal Sustainability: Insights from Southeast Asia and Beyond. *Cham: Springer Nature Switzerland,* 21-49. doi: 10.1007/978-3-031-75749-5_2

Siddique, I. (2024). Machine learning empowered geographic information systems: Advancing Spatial analysis and decision making. *World Journal of Advanced Research and Reviews,* 22(1), 10-30574. doi: 10.30574/wjarr.2024.22.1.1200

Singh, Y., Pandey, V. K., Mishra, S. K., & Kushwaha, P. (2024). Understanding environmental drivers and trends in chlorophyll-a and related factors across equatorial Indian Ocean regions: implications for marine ecosystem dynamics and vulnerability. *Marine Biology Research,* 20(9-10), 365-379. doi: 10.1080/17451000.2024.2371327

Siregar, E. S. Y., Simbolon, D., Wahju, R. I., Yulianto, I., & Siregar, V. P. (2025). Yellowfin Tuna (Thunnus Albacares) Fisheries Sustainability: A Weight-Length Analysis In The West Sumatera Waters. *Jurnal Teknologi Perikanan Dan Kelautan,* 16(3), 282-293. doi: 10.24319/jtpk.16.282-293

Smoliński, S., & Radtke, K. (2017). Spatial prediction of demersal fish diversity in the Baltic Sea: comparison of machine learning and regression-based techniques. *ICES Journal of Marine Science,* 74(1), 102-111. doi: 10.1093/icesjms/fsw136

Song, T., Pang, C., Hou, B., Xu, G., Xue, J., Sun, H., & Meng, F. (2023a). A review of artificial intelligence in marine science. *Frontiers in Earth Science,* 11, 1090185. doi: 10.3389/feart.2023.1090185

Song, L., Li, T., Zhang, T., Sui, H., Li, B., & Zhang, M. (2023b). Comparison of machine learning models within different spatial resolutions for predicting the bigeye tuna fishing grounds in tropical waters of the Atlantic Ocean. *Fisheries Oceanography,* 32(6), 509-526. doi: 10.1111/fog.12643

Srivastava, N., & Saxena, N. (2023). Applications of artificial intelligence and machine learning in geospatial data. In Emerging Trends, Techniques, and Applications in Geospatial Data Science. *IGI Global,* 196-219. doi: 10.4018/978-1-6684-7319-1.ch010

Talebi, H., Peeters, L. J., Otto, A., & Tolosana-Delgado, R. (2022). A truly spatial random forests algorithm for geoscience data analysis and modelling. *Mathematical Geosciences,* 54(1), 1-22. doi: 10.1007/s11004-021-09946-w

Utari, P. A., Setiabudidaya, D., Khakim, M., & Iskandar, I. (2019). Dynamics of the South Java Coastal Current revealed by RAMA observing network. *Terrestrial, Atmospheric & Oceanic Sciences,* 30(2). doi: 10.3319/TAO.2018.12.14.01

Vinston Raja, R., & Ashok Kumar, K. (2022). Fisher Scoring with Condition-Based Ensemble Supervised Learning Classification Technique for Prediction in PFZ. *Journal of Uncertain Systems,* 15(03), 2241009. doi: 10.1142/S1752890922410094

Waller, M. J., Humphries, N. E., Womersley, F. C., Loveridge, A., Jeffries, A. L., Watanabe, Y., ... & Sims, D. W. (2024). The vulnerability of sharks, skates, and rays to ocean deoxygenation: Physiological mechanisms, behavioral responses, and ecological impacts. *Journal of Fish Biology,* 105(2), 482-511. doi: 10.1111/jfb.15830

Wessel, P., Luis, J. F., Uieda, L. A., Scharroo, R., Wobbe, F., Smith, W. H., & Tian, D. (2019). The generic mapping tools version 6. *Geochemistry, Geophysics, Geosystems,* 20(11), 5556-5564. doi: 10.1029/2019GC008515

Xing, Q., Yu, H., Wang, H., Ito, S. I., & Chai, F. (2023). Mesoscale eddies modulate the dynamics of human fishing activities in the global midlatitude ocean. *Fish and Fisheries,* 24(4), 527-543. doi: 10.1111/faf.12742

Yuliardi, A. Y., Afiyah, N. N., Napitupulu, G., Munandar, B., Haryati, A., Husni, I. A., & Koropitan, A. F. (2026). Oceanographic Variability and its Association with Fisheries Productivity in Lamongan, Indonesia. *Thalassas: An International Journal of Marine Sciences*, 42(2), 57. doi: 10.1007/s41208-026-01088-2

Zhang, J., Fan, D., He, H., Xiao, B., Xiong, Y., & Shi, J. (2023). Forecasting Albacore (Thunnus alalunga) fishing grounds in the South Pacific based on machine learning algorithms and ensemble learning model. *Applied Sciences,* 13(9), 5485. doi: 10.3390/app13095485

Zhang, Y., Liu, J., & Shen, W. (2022a). A review of ensemble learning algorithms used in remote sensing applications. Applied Sciences, 12(17), 8654. https://doi.org/10.3390/app12178654

Zhang, Z., Wang, J., & Yuan, D. (2022b). Mixed layer salinity balance in the eastern tropical Indian Ocean. *Journal of Geophysical Research: Oceans,* 127(6), e2021JC018229. doi: 10.1029/2021JC018229

Zhao, Q., Peng, S., Wang, J., Li, S., Hou, Z., & Zhong, G. (2024). Applications of deep learning in physical oceanography: a comprehensive review. *Frontiers in Marine Science,* 11, 1396322. doi: 10.3389/fmars.2024.1396322