

Article

Fishing Area Prediction Using Scene-Based Ensemble Models

Adillah Alfatinah¹, Hone-Jay Chu^{1,*} , Tatas^{1,2}  and Sumriti Ranjan Patra¹ 

¹ Department of Geomatics, National Cheng Kung University, Tainan City 701401, Taiwan; adillah.alfatinah@gmail.com (A.A.); p68097022@gs.ncku.edu.tw (T.); sumritiranjanpatra@gmail.com (S.R.P.)
² Civil Infrastructure Engineering Department, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia
* Correspondence: honejaychu@geomatics.ncku.edu.tw

Abstract: This study utilized Chlorophyll-a, sea surface temperature (SST), and sea surface height (SSH) as the environmental variables to identify skipjack tuna catch hotspots. This study conducted statistical methods (decision tree, DT, and generalized linear model, GLM) as ensemble models that were employed for predicting skipjack area for each time slice. Using spatial historical data, each model was trained for one of the ensemble model sets. For prediction, the correlations of historical and new inputs were applied to select the predictive model. Using the scene-based model with the highest input correlation, this study further identified the fishing area of skipjack tuna in every case whether the alterations in their environment affected their abundance or not. Overall, the performance achieved over 83% for correlation coefficients (CC) based on the accuracy assessment. This study concluded that DT appears to perform better than GLM in predicting skipjack tuna fishing areas. Moreover, the most influential environmental variable in model construction was sea surface temperature (SST), indicating that the presence of skipjack tuna was primarily influenced by regional temperature.

Keywords: decision tree; generalized linear model; ensemble prediction; skipjack tuna; fisheries



Citation: Alfatinah, A.; Chu, H.-J.; Tatas; Patra, S.R. Fishing Area Prediction Using Scene-Based Ensemble Models. *J. Mar. Sci. Eng.* **2023**, *11*, 1398. <https://doi.org/10.3390/jmse11071398>

Academic Editor: Dariusz Kucharczyk

Received: 17 June 2023
Revised: 6 July 2023
Accepted: 8 July 2023
Published: 11 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Tuna is one of the largest and most commercially important groups of fish species, known for its specialization in various areas. Tuna species play a fundamental role in the open ocean ecosystem as both predator and prey within the intricate food web [1]. Among major tuna fisheries, the most important and significant catches within the species are the skipjack tuna (*Katsuwonus pelamis*), which accounted for more than half of the total global catch [2–4]. The skipjack tuna, characterized by its small size, is the most abundantly caught species of tuna both in terms of total numbers and weight [1]. In addition to its abundant resources, skipjack tuna stands out for its exceptional productivity when compared to other tuna species [5]. Due to these reasons, skipjack tuna has held a high importance level in global fisheries. However, recently, the world fishery area is facing increased pressure from a range of factors, including overfishing, climate change, pollution, and habitat degradation [6]. This problem held great importance due to the connection between the support received by the fishing industry and the overexploitation of marine resources [7,8].

Determining a possible fishing spot improves fishing efficiency and helps to facilitate the fisherman to maintain the total catches of skipjack tuna. Determining the possible fishing area for skipjack tuna is desirable by considering their life preference, which is much related to the marine environment condition. Their life preferences can be analyzed through the relationship between their presence with marine environmental parameters such as sea surface temperature (SST). The skipjack tuna is primarily found in habitats characterized by warm water masses with high oxygen content due to the limited thermoregulation capabilities. The habitat is often situated near cooler waters (below the thermocline),

allowing the skipjack tuna to release any excess metabolic heat [9]. Skipjack habitat is associated with an SST range between 18° and 30 °C [10], with a preferred range in the equatorial area ranging from 28.5 °C to 30.5 °C [11]. Dueri et al. [12] and Lehodey et al. [13] conducted a more complex investigation that derives skipjack habitat preferences as input data for population models such as SST and oxygen using a biogeochemical model. Mugo et al. [9] conducted a more detailed analysis of skipjack habitat by using generalized additive models to associate SST, chlorophyll-a (Chl-a) concentration, sea surface height (SSH) anomalies, and eddy kinetic energy at weekly and monthly time scales. Hsu et al. [14] found the suitability index based on SST, SSH, sea surface salinity, mixed layer depth, chlorophyll, and finite-size Lyapunov exponents. Putri et al. [15] aimed to analyze the relationship between satellite data on SST and Chl-a anomalies, and skipjack tuna catch distribution. Mugo et al. [16] predicted habitat suitability maps using SST, SSH, Chl-a, diffuse attenuation coefficient, and surface wind speed. Thus, we analyze the relationship between the possible skipjack presence with the environmental variables such as Chl-a, SSH, and SST. Moreover, the ensemble model is a machine-learning technique that combines multiple models to produce an optimal predictive model [17]. Here, each case in the ensemble model is reproduced using historical environmental data. This study uses models, including the generalized linear model (GLM) and decision tree (DT) to develop the presence prediction of skipjack tuna. The utilization of DT to identify and classify objects was implemented first by Hunt et al. [18]. DT is supervised learning used for classification and regression [19]. Previous studies compared the predictive accuracy of statistical and rule-based methods [20,21]. GLMs are commonly used for prediction modeling because of their strong statistical ability to realistically model ecological relationships [22]. GLM provides a less restrictive form than classic multiple regressions by providing error distributions since the variance is not constant or normally distributed. A comparative analysis will be conducted between DT and GLM to identify the most suitable predictive model.

Fishing area prediction of skipjack tuna is conducted by identifying the hotspot of the tuna fishing activities using environmental variables, e.g., Chl-a, SST, and SSH. This study uses statistical approaches such as GLM and DT to generate a favorable fishing area for skipjack tuna on a weekly scale. Based on ensemble model sets, each model is trained using spatial historical data. For scene-based prediction, the correlations of historical and new inputs are applied to select the most suitable predictive model. This study further estimates the spatial distribution of skipjack tuna fishing spots using this scene-based approach.

2. Materials and Methods

2.1. Study Area

The study area for this study is located at 140° to 179.91° W, −10° to 30° N which is revealed in Figure 1 below. The area was part of the Western and Central Pacific Ocean (WCPO) and situated near the equator line. The warmest pool of oceanic surfaces lay in this area and played an essential role in the Earth's climate. Annual SST in the western Pacific Ocean was above 28 °C [23]. Skipjack tuna are usually found year-round concentrated in warmer tropical waters of the WCPO, with that distribution expanding seasonally into subtropical waters to the north and south, particularly on the western side of the WCPO [24]. A previous study carried out by Hampton [24] stated that the catch is highly concentrated in the equatorial zone due to the abundant catch of skipjack tuna in this area.

2.2. Dataset

This study dataset consists of fish catch points and environmental variables such as Chl-a, SST, and SSH in 2016 and 2017. The complete dataset comprises 48 cases, encompassing the model development and the prediction phases. This study developed the model using a weekly dataset (46 cases) in 2017. For scene-based prediction, the new dataset consisted of weeks 2 and 14 in 2016 (2 cases).

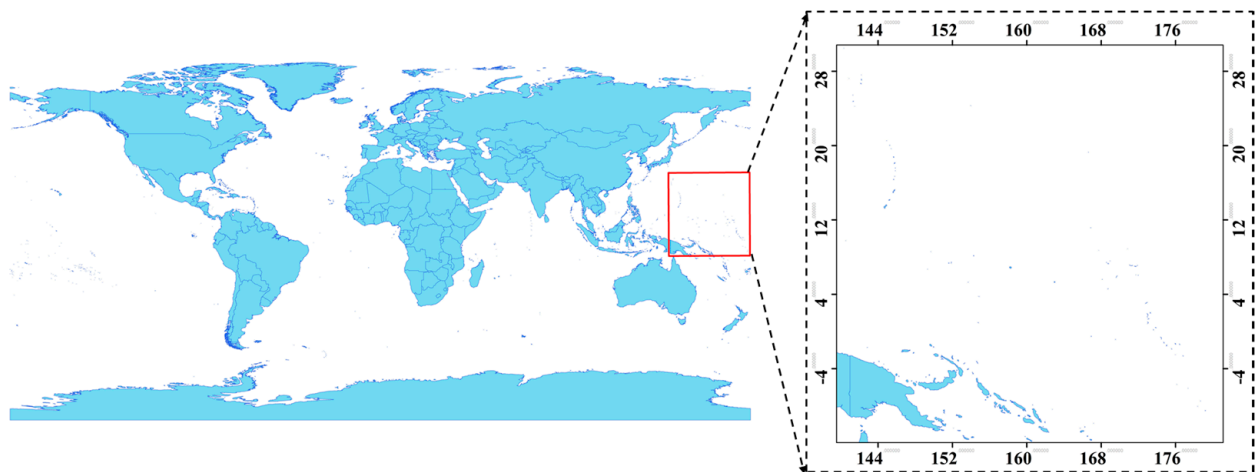


Figure 1. Study area at 140° to 179.91° W, −10° to 30° N.

2.2.1. Fishing Catch Points

Fishing catch points were obtained from the location record of skipjack tuna fishing activities within the study area (Figure 2). Fishing catch points were obtained from the Overseas Fisheries Development Council of Taiwan (dataset can be accessed at www.ofdc.org.tw, accessed on 10 July 2020). The recorded points consisted of fishing date, latitude, longitude, fishing activity code, vessel school type, start and end time of fishing activities, and the total catch.

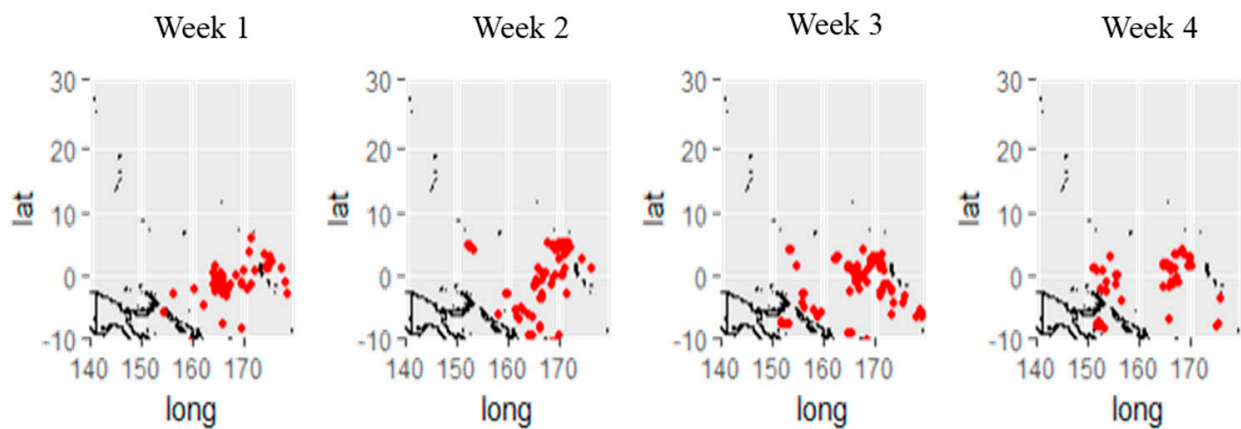


Figure 2. Fish catch point for the first four weeks in 2017.

2.2.2. Environmental Factors from Remote Sensing

According to Laevastu and Hayes [25], the distribution of pelagic fish tends to migrate to fertile water. The skipjack tuna catch results are more influenced by the combination of two oceanographic variables, which are SST and Chl-a [11,25]. Aside from those variables, this study also considers the fact of skipjack tuna habit to inhabit the upper layer of the ocean. In consideration of that, this study also includes SSH as one of the environmental variables for predicting skipjack tuna fishing areas. Table 1 shows details of satellite imagery dataset information about Chl-a, SST, and SSH. The spatial resolution of data is 9 km using a resampling process based on bilinear interpolation. The input image time series are selected based on the closest time of weekly fishing catches.

Table 1. Imagery dataset details.

Var	Information Details
Chl-a	MODIS-Aqua Level 3 SMI; Temporal: 8 Days; Spatial: 9 km; Unit: mgm^{-3}
SST	MODIS Level 3 SMI; Temporal: 8 Days; Spatial: 9 km; Unit: $^{\circ}\text{C}$
SSH	Global Ocean Analysis; Temporal: Daily mean; Spatial: $0.083^{\circ} \times 0.083^{\circ}$; Unit: Meter (m)

2.3. Method

Figure 3 shows scene-based ensemble prediction. The relation between the possible skipjack presence and environmental variables using machine learning is analyzed. Firstly, a machine learning model is constructed using training datasets, which leads to the generation of the skipjack presence map. Using weekly spatial data, every model is trained for ensemble model sets. To understand their performance from DT and GLM, this study used confusion matrix methods to calculate correlation coefficients (CC), Cohen’s kappa, and area under curve (AUC) to analyze the ability of both models for predicting the fish probability maps. After training, an additional dataset is incorporated into the scene-based model to predict the final skipjack presence maps. For fish-presence prediction, the correlations of historical and new inputs are applied to select the most suitable predictive model. Using the scene-based model with the highest input correlation, this study will further plot the spatial distribution of skipjack tuna in any case and understand whether the alteration in their environment affects their abundance or not.

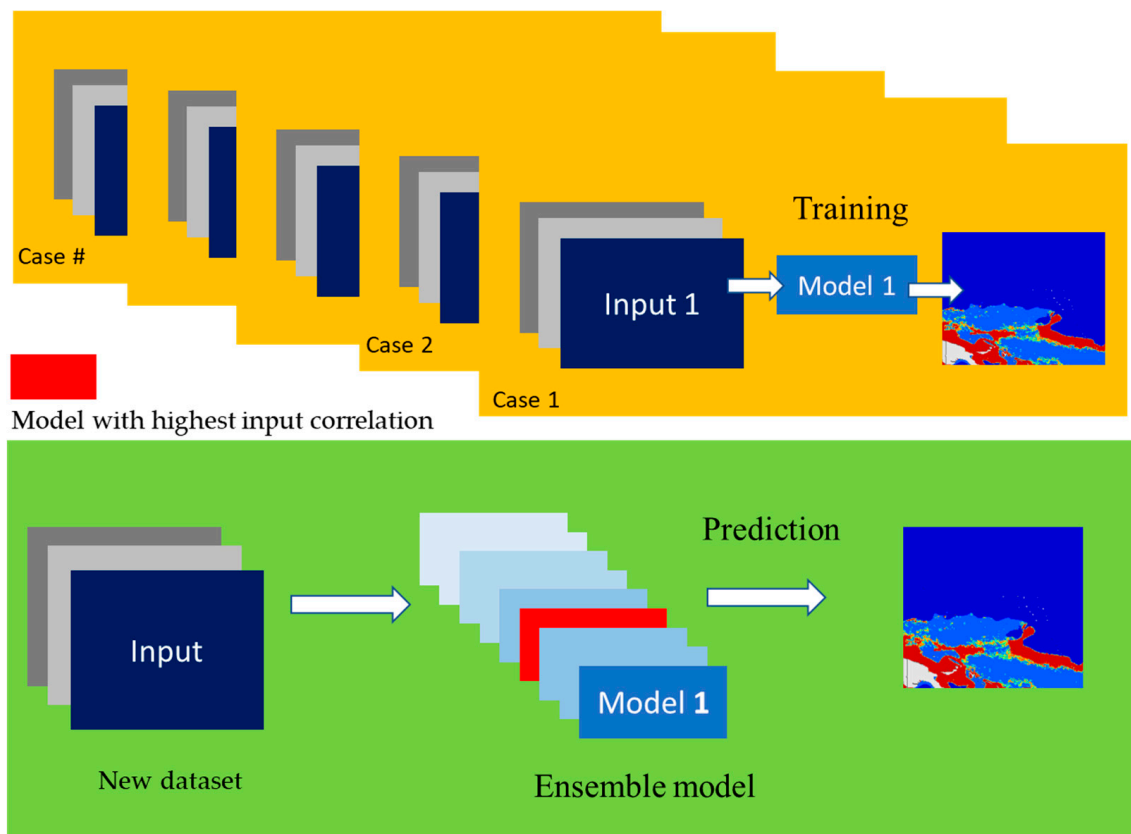


Figure 3. Ensemble model training, and scene-based prediction (training: the DT or GLM model involved in training data resulting in skipjack presence maps. In each case, the model was trained for one of the ensemble model sets; prediction: adding a new dataset into the scene-based model to predict the skipjack presence map).

2.3.1. Fishing Catch Point Density Map

To assume the possible fishing area, this study uses a density map based on the fishing catch point in 2017. The kernel density method in ArcMap 10.5 was used to generate the density maps. Kernel density estimation (KDE) is a technique for the estimation of probability density function (PDF) that analyzes the studied probability distribution. The kernel technique produces smooth estimation, uses all sample points locations, and more convincingly suggests multimodality. In the commonly used practical application, kernel estimation uses the symmetric kernel function, although the asymmetric function has recently been increasingly used [26,27].

For a random sample y_1, y_2, \dots, y_T , the kernel estimator of the PDF $f'_T(y)$ at point y can be expressed as:

$$f'_T(y) = \frac{1}{nh} \sum_{i=1}^T K\left(\frac{y - y_i}{h}\right) \tag{1}$$

where n is the number of observations, h is the bandwidth that determines the smoothness of the estimate, and $K(\cdot)$ is the kernel [28,29]. In other words, the kernel transforms the point location of i y into an interval centered around y i [27].

In this study, the fishing points were used as samples to generate the density maps. Furthermore, the density map will help to assume the fish presence area. The dense area in density maps indicated the area with the most gathered fishing catch points and vice versa. Moreover, this study assumed the dense area as the fishing area, considering the abundance of fishing catch points here, which implied that massive amounts of fishing activities occurred. Meanwhile, the area with lower density is assumed to be an area with the least fishing activities. Following that, for this study, we will set the dense area as the fish area (FA) and the lower density area as the no fish area (NFA). To obtain the prediction model, it is necessary to generate the training samples dataset as well as the testing samples dataset. The training samples dataset is employed to construct the prediction model; meanwhile, the testing samples dataset is employed to test the model's performance with the intention that a processing method explained in the following section be carried out.

2.3.2. Fishing Area Prediction Model Construction

This study applied DT and GLM techniques to construct predictive models. All calculations were performed under the statistical software "R" and its contributed packages [21]. The model utilizes the R Software Environment, employing the 'CART' package for DT and the 'glm' package for GLM. The processing was conducted on a computer system equipped with an Intel(R) Core™ i5-3550 CPU operating at 3.30 GHz and 8.00 GB of RAM. With this computer specification, the runtime for the process ranged from approximately 3 to 5 h. CARTs are nonparametric classification techniques that are implemented using a binary recursive partition algorithm [19]. We already prepared the training samples dataset such as Chl-a, SST, and SSH values. The classified fish area, which consists of class 1 for FA and class 2 for NFA, will be assigned as the response variables, while Chl-a, SST, and SSH values assigned as explanatory variables. GLM is executed in this study as the alternative model to compare the performance of DT for the prediction. The response variables for GLM will be carried out by classified fish area based on explanatory or predictor variables.

A GLM is made up of a linear predictor:

$$\eta_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi} \tag{2}$$

A link function that describes how the mean, $E(Y_i) = \mu_i$, depends on the linear predictor shown in Equation (3) below,

$$g(\mu_i) = \eta_i \tag{3}$$

In a link function, a logistic regression model is used to predict fish area occurrence.

Where x_{1i}, \dots, x_{pi} are environmental variables in this study, e.g., Chl-a, SST, and SSH. $\beta_0, \beta_1, \dots, \beta_p$ are model parameters in Equation (2).

2.3.3. Scene-Based Prediction

This study generated 46 weekly models for ensemble models. To know which model gives the best performance when applied to the new dataset, we analyzed the correlation between the environmental variables. The correlation matrix is one of the foundations of factor analysis and has found its way into such diverse areas, playing an essential role in multivariate analysis since, by itself, it captures the pairwise degrees of relationship between different components of a random vector [30,31]. The model, which has the highest correlation with the new and historical dataset, conducts further as the main prediction model.

2.3.4. Accuracy Assessment

The confusion matrix is needed to calculate most of the measures of classification accuracy from the prediction model [32]. It quantified the prediction performance of the model as the percentage samples where the model correctly predicts the FA and NFA class using the testing samples dataset. From the values in the matrix of confusion matrix, this study, therefore, calculated alternative performance measures including CC [32,33], Cohen’s kappa, and the area under the receiving operating characteristic curve (AUC) [32,34]. The CC is the ratio of samples correctly classified by the prediction model. Cohen’s kappa is a statistic that measures the agreement of two categorical items [35]. The kappa index also considers both omission and commission errors and can be used to assess whether prediction model performance differs from expectation based on chance alone [33]. In this study, the agreement that was measured was between the observed and predicted FA and NFA classes. Cohen’s kappa values were used to determine which model was better for fishing area prediction (Table 2). A previous study conducted by McKenna and Castiglione [36] used the categorization presented in the following Table to assess the significance of Cohen’s kappa values. Following that, this study also used this categorization system for determining Cohen’s kappa significance. The receiver operating characteristic (ROC) curve is a graphical method that represents the relation between the false positive fraction and the sensitivity for a range of thresholds [34]. If the prediction was possibly expected by chance, the relation would be 45°. Good model performance is characterized when the curve passes close to the upper left corner of the plot. The area between the 45° line and the curve that measures discrimination was called AUC. The AUC quantifies the ability of a model to discriminate between the area where the fish is present (FA) versus those where it is absent (NFA).

Table 2. Cohen’s kappa value categorization.

Cohen’s Kappa Value	Agreement Categorization
<0.01	No agreement
0.01–0.20	Slight agreement
0.21–0.40	Fair agreement
0.41–0.60	Moderate agreement
0.61–0.80	Substantial agreement
>0.80	Almost perfect agreement

3. Results

3.1. Fishing Catch Point Density Maps

Figure 4 shows the weekly density map of fish catch points in 2017 from Week 13 to Week 16. The unit of the density map is the total catch points within one pixel. The brown color area in density maps was the dense area, while the yellow area was the low-density area. The dense area in density maps indicated the area with the most fishing catch points gathered and vice versa. This dense area is assumed to be the fishing potential area, considering the abundance of fishing catch points here, which implied massive amounts of

fishing activities occurred. Furthermore, the density pattern appears to be more clustered in maps. The point distributions in these weeks were more scattered than in another week. Considering the KDE used the points distribution location to generate the density map, this circumstance might happen because of the scattered points distribution in these weeks. Following the goal of this study, we predict and generate the possible fish presence area probability map constructed by the DT and GLM. The map reveals the possibility of the percentage of fish existence by considering environmental factors consistent in every area calculated by the models. The result displayed an almost zero density pattern in week 33. We assume that this condition was caused by a limited number of points occurring in this week, producing a very low-density pattern in the density map.

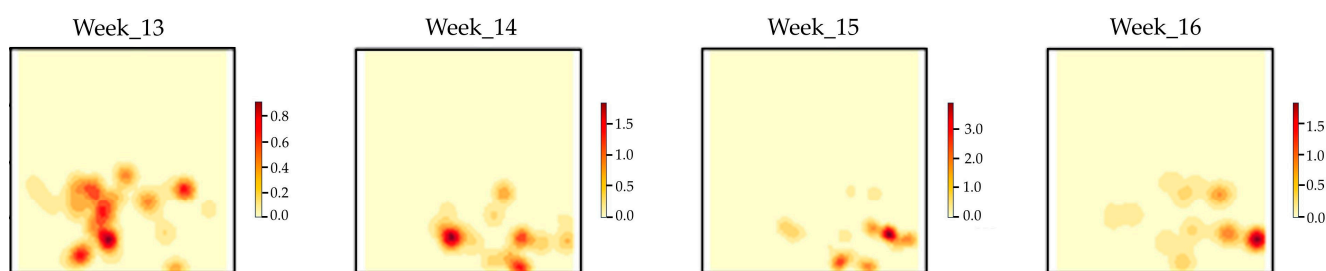


Figure 4. Weekly density map of fish catch points in 2017 from Week 13 to Week 16.

3.2. Fishing Area Model in DT and GLM

Most spatial patterns of fishing spots are similar when compared to DT and GLM in these weeks (Figure 5). All CC achieved more than 80% in DT and GLM, except week 33 in DT (Table A1). The DT belonging to the supervised classification was processed to generate the fishing area prediction. By considering the environmental factors condition in every area as the explanatory variables and the fish area assumption as the response variables in model construction, the DT can predict the fish possibility. Ensemble models in DT were obtained. Using the input dataset, we constructed, weekly, 46 models and generated the fish probability map. We found SST to be the most influential variable in weekly processing. The SST became the most influential variable from 46 weeks of model construction, followed by SSH and Chl-a, respectively (Table A2). Aside from that, we also generated the fish probability maps for weekly processing using the constructed models. From the maps, we were able to discover much white area around the possibility maps result. The white area indicated the null value area that was caused by the null value in the Chl-a satellite imagery. Moreover, we discovered a particular result happened because there was a lack of fishing catch points within a week, indicating less fishing catch activity occurring at that time. Following that, in the training samples preparation, we obtained fewer samples for the fish area class, causing the DT to assume that there were no fish areas detected.

Mostly, GLM achieved less accuracy than DT. This study executed GLM techniques as an alternative model because GLM was one of the simple parametric approaches for prediction. The purpose of this comparison is to evaluate which prediction method might be more robust and accurate for fishing area prediction. This study applied a simple GLM with linear terms to all the cases. GLM proceeds using the same training and testing sample dataset with the DT, with the response variables being carried out by classified fish area assumption while the remaining will be explanatory or predictor variables. In GLM, we discover SST almost has high significance every week (Table A3). Similar to fish maps probability generated by DT, we also found the null value of Chl-a affecting the map result. Meanwhile, GLM was able to classify and generate a probability map.

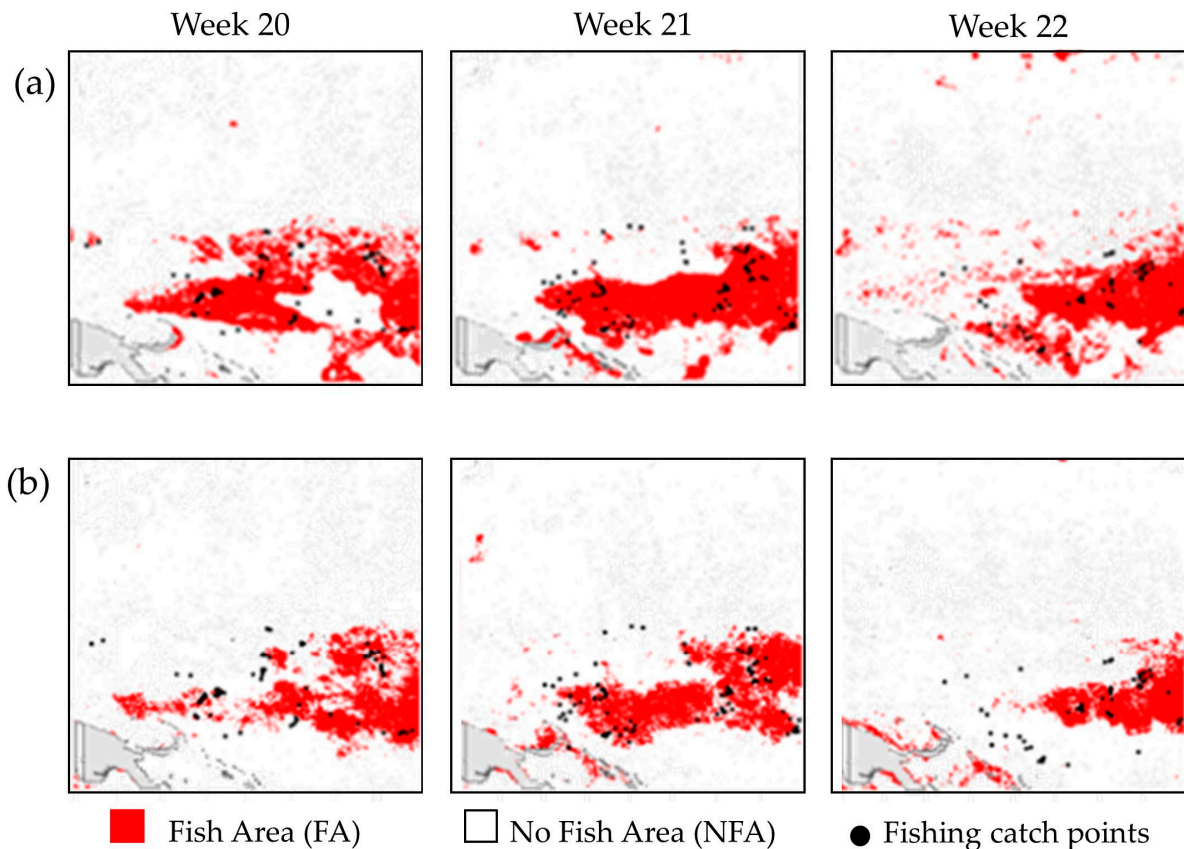


Figure 5. Fish presence area maps with catch points for (a) DT and (b) GLM from Weeks 20 to 22.

3.3. Scene-Based Prediction in Fishing Area

To discover the best prediction model among the generated models, the correlation was used between the environmental dataset in those 46 weeks with the new datasets. The week in which environmental variables have a high correlation with the new dataset will be chosen as the main prediction model for the weekly processing.

Since the environmental variables change every week, we, thus, intended to find a prediction model that might work well in different weeks. The correlation matrix between the environmental variables indicated in the week with the highest correlation. In the previous section, DT has a better performance than GLM in predicting fish presence. Thus, we applied the prediction model generated by DT to our new dataset. To know whether the model can give an excellent performance using the new dataset, an accuracy assessment was conducted. The accuracy assessment presented by CC, Cohen’s kappa, and the AUC value from the prediction results is shown in the following Table 3. Based on the CC value, the model can achieve more than 80% accuracy. This finding indicated that the prediction model was appropriate to be the main prediction model for weekly processing. Figure 6 plotted the new prediction maps with the fishing catch points. The red area represents the predicted FA while the blue area presented the NFA class. The maps appeared to mostly land in the predicted FA, indicating that the prediction model can generate the prediction maps accurately.

Table 3. Accuracy assessment results for new datasets.

	CC	Kappa	AUC
Test 1	0.8814	0.7315	0.8455
Test 2	0.8308	0.6311	0.8087

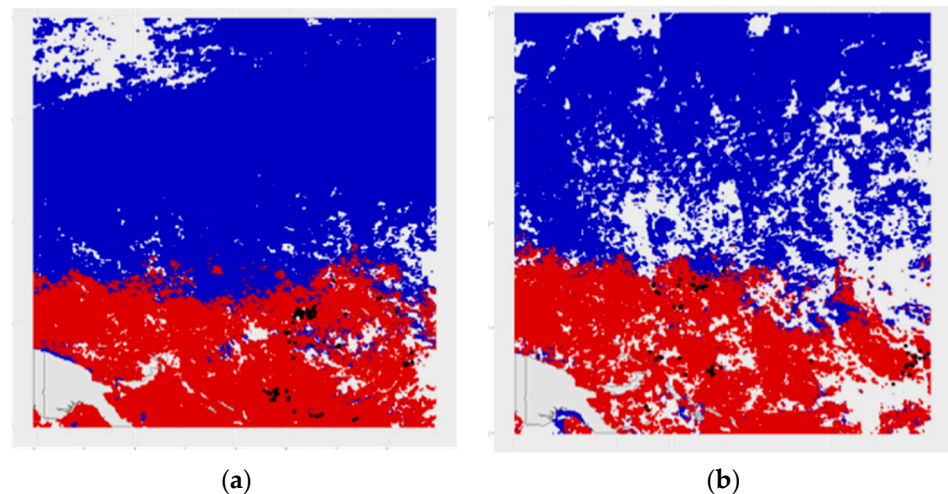


Figure 6. Fishing area prediction maps for new datasets in (a) Test 1; (b) Test 2 (NFA: blue; FA: red).

4. Discussion

4.1. Ensemble Model

The application of machine-learning algorithms to fishing area prediction is appropriate for improving estimation accuracy [37,38]. Machine learning is an effective approach to evaluating relations between the potential fishing area and corresponding environmental parameters. The GLM as the logistic method fits a single line to divide the area precisely into two. The GLM bisects the variables into smaller and smaller regions until it fits the condition needed. GLM determines the significance level statistically between the variable and the term which, in this case, is the fish presence class. The GLM showed a more stable result for the variable significance in its model construction. Meanwhile, GLM considers the significance of each variable statistically in the term, making it possible for the variable to be significant in all cases. Moreover, DT establishes the variable's rule in the model construction by fitting all the variables into a suitable condition, causing the importance of the variables to vary in each case. For the consideration of both models' performances, we discover both DT and GLM to be appropriate for predicting skipjack tuna fishing areas. Furthermore, the performance of DT appears to be better than that of GLM in all cases based on their accuracy assessment result. DT's ability to partition the variables into groups until it fits with the class term and becomes a major reason for DT to perform best for the prediction based on the classification in this study. However, DT fails in the very low-density pattern of fishing area prediction.

We only applied one-year weekly data for the ensemble model. More ensemble models with historical data will generate a more accurate spatial map of the fish potential zone. The model proves to be able to predict the weekly fish presence with high accuracy but may get lower accuracy along with the longer period. This condition makes sense due to the uncertainty in their environmental variables at time scales.

4.2. Tuna and Environmental Variables

A compelling case was found when there was a massive decrement in the skipjack tuna catch number. A strong El-Niño occurred, causing the SST to become an anomaly in the study case region [39]. Following that, we consider more analysis of the change in SST with the predicted fishing area. Our study case was located in the Western Equatorial Pacific which has the warmest surface temperatures in the world and is commonly known as the 'warm pool' with SST > 29 °C [40]. This area also has a weak seasonal variation of the warm pool, which means the temperature only slightly changes over seasons. Skipjack tuna spreading in this area is hypothesized to correlate with the position of the warm pool–cold tongue convergence zone state [13]. In the Northeast Indian Monsoon, NEM season, the predicted fish area was located mostly in the coastal area with an SST range

around 29–30 °C. We also discovered part of the fishing area is located in the offshore area, where the position of the convergence zone in this season with SST is around 29 °C. In the IM1 (inter-monsoon 1) season, the fishing area was shifting to the offshore area mostly located in the convergence area, while the cold tongue appears to be more apparent during this season. During the rapid decrease of the total skipjack tuna catch number from IM2 into the southwest monsoon, SWM season, the Chl-a concentration was not changed seriously. The Chl-a increased rapidly in the IM2 season from 0.093 to 0.129 mgm⁻³, inversely proportional to the decreasing total catch number. Based on the predicted result, the fish gathers around the area with Chl-a concentration ranging from 0.09 to 0.10 mgm⁻³. In addition, the fishing spots were not far from strong SST fronts (its gradient) [41]. The high-gradient SST front was highly associated with the fishing area. Future work will suggest that the SST front, subsurface temperature, or near-surface salinity can be considered as a predictor in the model [14,42].

5. Conclusions, Limitations, and Future Research

The research goal of this study is to predict skipjack tuna fishing areas by investigating the relation between skipjack tuna presence and marine environmental parameters using an ensemble model. The predictive tuna fishing area for all cases in this research was successfully obtained by using DT and GLM methods. Both models can generate the predicted maps but DT appeared to perform better than GLM in this research. In the accuracy assessment, most performance achieved more than 80% for CC. Following these results, DT is more appropriate to use in predicting skipjack tuna fishing areas. Furthermore, the correlation for environmental variables between new and historical input datasets is analyzed to discover the suitable prediction model. This prediction was proven to acquire a good performance based on the optimal-scene ensemble model.

In the scene-based model, this result supports our assumption that the environmental variables have an effective impact on the skipjack tuna area prediction. The strong correlation between the prediction and the environmental variables, particularly sea surface temperature (SST) and chlorophyll-a concentration (Chl-a), is indicative of their significant influence on fish feeding habits. This observation validates the robust association between the presence of skipjack and the environmental conditions prevalent in the region. This scene-based model is appropriate to use for potential fishing zone prediction. This model is only available based on spatial environmental data, e.g., satellite-image time series. For future research works, it is recommended to adopt fine spatial-resolution datasets to provide fishing areas at finer resolutions. This would allow us to obtain more detailed fishing area maps, enhancing their usefulness for localized research studies. The impact of marine environmental parameters on the presence of skipjack tuna will be thoroughly examined in the future.

Author Contributions: Conceptualization, H.-J.C.; methodology, A.A. and S.R.P.; validation, S.R.P. and T.; formal analysis, A.A.; data curation, A.A. and T.; writing—original draft preparation, A.A. and T.; writing—review and editing, H.-J.C.; visualization, A.A. and S.R.P.; supervision, H.-J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Chlorophyll-a and SST were obtained from MODIS-Aqua Level 3 SMI whereas SSH was obtained from the Copernicus Marine Environment Monitoring Service.

Acknowledgments: The authors would like to thank the editors and anonymous reviewers for providing suggestions for paper improvement. We are grateful for the data providers and the support from the SATU joint research scheme and the SDGs joint research project in NCKU.

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

Appendix A

Table A1. Model Performances in DT and GLM with CC (Correlation Coefficients), Kappa, and AUC (Area Under Curve).

Weeks	DT			GLM		
	CC	Kappa	AUC	CC	Kappa	AUC
Week 1	0.922	0.823	0.925	0.845	0.638	0.849
Week 2	0.913	0.807	0.894	0.874	0.712	0.816
Week 3	0.956	0.905	0.958	0.857	0.689	0.840
Week 4	0.946	0.882	0.941	0.929	0.845	0.932
Week 5	0.911	0.804	0.907	0.899	0.774	0.869
Week 6	0.885	0.869	0.917	0.879	0.729	0.843
Week 7	0.898	0.733	0.872	0.879	0.734	0.900
Week 8	0.983	0.925	0.976	0.922	0.838	0.974
Week 9	0.919	0.612	0.924	0.859	0.707	0.912
Week 10	0.938	0.803	0.947	0.918	0.826	0.947
Week 11	0.911	0.870	0.888	0.931	0.847	0.903
Week 12	0.933	0.833	0.958	0.878	0.733	0.891
Week 13	0.876	0.669	0.872	0.832	0.626	0.854
Week 14	0.846	0.759	0.863	0.901	0.781	0.908
Week 15	0.848	0.736	0.88	0.840	0.628	0.764
Week 16	0.92	0.735	0.935	0.860	0.680	0.893
Week 17	0.899	0.808	0.874	0.871	0.708	0.841
Week 18	0.874	0.722	0.845	0.859	0.678	0.813
Week 19	0.9	0.725	0.907	0.813	0.585	0.840
Week 20	0.873	0.807	0.925	0.878	0.726	0.833
Week 21	0.935	0.824	0.923	0.915	0.814	0.928
Week 22	0.91	0.791	0.919	0.802	0.603	0.889
Week 23	0.885	0.835	0.914	0.885	0.745	0.897
Week 24	0.887	0.793	0.881	0.850	0.658	0.831
Week 25	0.817	0.614	0.781	0.824	0.608	0.841
Week 26	0.859	0.778	0.905	0.890	0.753	0.887
Week 27	0.836	0.614	0.845	0.815	0.580	0.827
Week 28	0.894	0.741	0.902	0.813	0.600	0.867
Week 29	0.868	0.671	0.89	0.795	0.573	0.881
Week 30	0.835	0.761	0.893	0.897	0.771	0.881
Week 31	0.809	0.594	0.852	0.766	0.445	0.715
Week 32	0.823	0.656	0.891	0.839	0.624	0.805
Week 33	0.333	0.000	0.5	1.000	1.000	1.000
Week 34	0.914	0.798	0.944	0.829	0.616	0.881
Week 35	0.9	0.802	0.884	0.876	0.727	0.892
Week 36	0.921	0.785	0.907	0.835	0.616	0.750

Table A1. *Cont.*

Weeks	DT			GLM		
	CC	Kappa	AUC	CC	Kappa	AUC
Week 37	0.871	0.709	0.864	0.821	0.589	0.811
Week 38	0.903	0.782	0.916	0.866	0.707	0.884
Week 39	0.914	0.857	0.919	0.905	0.797	0.923
Week 40	0.857	0.734	0.829	0.881	0.739	0.899
Week 41	0.848	0.683	0.846	0.856	0.675	0.827
Week 42	0.92	0.808	0.929	0.902	0.785	0.912
Week 43	0.927	0.781	0.933	0.863	0.697	0.863
Week 44	0.902	0.806	0.892	0.872	0.745	0.967
Week 45	0.911	0.798	0.925	0.876	0.738	0.938
Week 46	0.888	0.743	0.897	0.885	0.737	0.905

Table A2. Variable Importance of DT (variable importance: a score to quantify how useful they are at predicting a target variable).

Weeks	Chl-a	SST	SSH
Week 1	12%	43%	45%
Week 2	17%	52%	31%
Week 3	13%	47%	40%
Week 4	23%	42%	35%
Week 5	22%	47%	31%
Week 6	9%	47%	45%
Week 7	20%	65%	15%
Week 8	22%	53%	25%
Week 9	27%	48%	25%
Week 10	17%	59%	25%
Week 11	22%	56%	22%
Week 12	23%	43%	34%
Week 13	22%	56%	22%
Week 14	25%	58%	17%
Week 15	26%	49%	25%
Week 16	25%	46%	29%
Week 17	25%	47%	29%
Week 18	38%	29%	33%
Week 19	27%	39%	34%
Week 20	25%	21%	54%
Week 21	44%	20%	37%
Week 22	42%	18%	40%
Week 23	23%	29%	48%
Week 24	17%	18%	64%

Table A2. *Cont.*

Weeks	Chl-a	SST	SSH
Week 25	24%	3%	72%
Week 26	21%	13%	65%
Week 27	18%	22%	60%
Week 28	48%	13%	39%
Week 29	51%	17%	31%
Week 30	31%	22%	47%
Week 31	27%	16%	57%
Week 32	10%	18%	72%
Week 33	28%	16%	55%
Week 34	21%	20%	59%
Week 35	13%	49%	39%
Week 36	53%	8%	39%
Week 37	32%	21%	47%
Week 38	43%	28%	29%
Week 39	45%	23%	32%
Week 40	13%	56%	31%
Week 41	27%	45%	28%
Week 42	25%	46%	29%
Week 43	20%	43%	38%
Week 44	12%	55%	33%
Week 45	26%	45%	29%
Week 46	19%	50%	31%

Table A3. Coefficients, and Significance Level in GLM (t-tests are used to determine if a particular variable is statistically significant).

Weeks	Variables		
	Chl-a	SST	SSH
Week 1	0.921	1.58×10^{-10} ***	0.221
Week 2	3.42×10^{-14} ***	$<2 \times 10^{-16}$ ***	0.002 **
Week 3	2.55×10^{-8} ***	$<2 \times 10^{-16}$ ***	1.96×10^{-6} ***
Week 4	$<2 \times 10^{-16}$ ***	$<2 \times 10^{-16}$ ***	0.145
Week 5	0.120	$<2 \times 10^{-16}$ ***	0.476
Week 6	0.727	1.65×10^{-13} ***	4.47×10^{-5} ***
Week 7	0.757	$<2 \times 10^{-16}$ ***	0.033 *
Week 8	0.044 *	2.54×10^{-11} ***	1.00×10^{-5} ***
Week 9	0.211	$<2 \times 10^{-16}$ ***	3.48×10^{-5} ***
Week 10	0.102	$<2 \times 10^{-16}$ ***	0.006 **
Week 11	0.216	$<2 \times 10^{-16}$ ***	1.15×10^{-11} ***
Week 12	0.836	$<2 \times 10^{-16}$ ***	5.85×10^{-9} ***
Week 13	0.769	$<2 \times 10^{-16}$ ***	2.27×10^{-5} ***

Table A3. *Cont.*

Weeks	Variables		
	Chl-a	SST	SSH
Week 14	0.688	$<2 \times 10^{-16}$ ***	0.029 *
Week 15	0.591	$<2 \times 10^{-16}$ ***	0.002 **
Week 16	0.814	1.96×10^{-15} ***	$<2 \times 10^{-16}$ ***
Week 17	$<2 \times 10^{-16}$ ***	9.08×10^{-15} ***	0.389
Week 18	2.37×10^{-8} ***	$<2 \times 10^{-16}$ ***	6.07×10^{-13} ***
Week 19	0.467	$<2 \times 10^{-16}$ ***	0.013 *
Week 20	0.177	$<2 \times 10^{-16}$ ***	$<2 \times 10^{-16}$ ***
Week 21	$<2 \times 10^{-16}$ ***	$<2 \times 10^{-16}$ ***	$<2 \times 10^{-16}$ ***
Week 22	$<2 \times 10^{-16}$ ***	1.55×10^{-10} ***	$<2 \times 10^{-16}$ ***
Week 23	3.07×10^{-6} ***	6.71×10^{-16} ***	1.96×10^{-11} ***
Week 24	1.21×10^{-7} ***	2.49×10^{-6} ***	1.03×10^{-15} ***
Week 25	0.646	0.491	$<2 \times 10^{-16}$ ***
Week 26	$<2 \times 10^{-16}$ ***	$<2 \times 10^{-16}$ ***	$<2 \times 10^{-16}$ ***
Week 27	$<2 \times 10^{-16}$ ***	$<2 \times 10^{-16}$ ***	$<2 \times 10^{-16}$ ***
Week 28	$<2 \times 10^{-16}$ ***	$<2 \times 10^{-16}$ ***	2.44×10^{-13} ***
Week 29	$<2 \times 10^{-16}$ ***	$<2 \times 10^{-16}$ ***	$<2 \times 10^{-16}$ ***
Week 30	2.07×10^{-8} ***	2.14×10^{-6} ***	0.064
Week 31	1.21×10^{-7} ***	$<2 \times 10^{-16}$ ***	$<2 \times 10^{-16}$ ***
Week 32	$<2 \times 10^{-16}$ ***	0.001 **	$<2 \times 10^{-16}$ ***
Week 33	0.112	0.848	0.845
Week 34	$<2 \times 10^{-16}$ ***	$<2 \times 10^{-16}$ ***	$<2 \times 10^{-16}$ ***
Week 35	8.05×10^{-9} ***	$<2 \times 10^{-16}$ ***	$<2 \times 10^{-16}$ ***
Week 36	2.26×10^{-8} ***	$<2 \times 10^{-16}$ ***	$<2 \times 10^{-16}$ ***
Week 37	0.00356**	$<2 \times 10^{-16}$ ***	$<2 \times 10^{-16}$ ***
Week 38	1.98×10^{-14} ***	$<2 \times 10^{-16}$ ***	$<2 \times 10^{-16}$ ***
Week 39	$<2 \times 10^{-16}$ ***	9.67×10^{-14} ***	0.016 *
Week 40	0.013 *	$<2 \times 10^{-16}$ ***	2.21×10^{-5} ***
Week 41	4.4×10^{-11} ***	$<2 \times 10^{-16}$ ***	0.217
Week 42	1.77×10^{-8} ***	$<2 \times 10^{-16}$ ***	0.881
Week 43	0.733	$<2 \times 10^{-16}$ ***	0.066
Week 44	0.344	3.95×10^{-14} ***	0.083
Week 45	0.055	$<2 \times 10^{-16}$ ***	0.001 **
Week 46	0.748	$<2 \times 10^{-16}$ ***	0.700

*** is one whose *p*-value < 0.001; ** is *p* < 0.01; * is *p* < 0.05; . is *p* < 0.1.

References

- Galland, G.; Rogers, A.; Nickson, A. Netting billions: A Global Valuation of Tuna. The Pew Charitable Trusts. 2016. Available online: <http://hdl.handle.net/1957/60217> (accessed on 10 July 2020).
- Mahaliyana, A.S.; Jinadasa, B.K.K.K.; Liyanage, N.P.P.; Jayasinghe, G.D.T.M.; Jayamanne, S.C. Nutritional Composition of Skipjack Tuna (*Katsuwonus pelamis*) Caught from the Oceanic Waters around Sri Lanka. *Am. J. Food Nutr.* **2015**, *3*, 106–111. [CrossRef]
- Miyake, M.P.; Guillotreau, P.; Sun, C.-H.; Ishimura, G. *Recent Developments in the Tuna Industry*; FAO: Rome, Italy, 2010.

4. Cimino, M.A.; Anderson, M.; Schramek, T.; Merrifield, S.; Terrill, E.J. Towards a fishing pressure prediction system for a western Pacific EEZ. *Sci. Rep.* **2019**, *9*, 461. [[CrossRef](#)] [[PubMed](#)]
5. Indian Ocean Tuna Commission. *Executive Summary of The Status of The Skipjack Tuna Resource*; Indian Ocean Tuna Commission: Victoria, Seychelles, 2005.
6. Stuart, V.; Platt, T.; Sathyendranath, S. The future of fisheries science in management: A remote-sensing perspective. *ICES J. Mar. Sci.* **2011**, *68*, 644–650. [[CrossRef](#)]
7. Armas, E.; Arancibia, H.; Neira, S. Identification and Forecast of Potential Fishing Grounds for Anchovy (*Engraulis ringens*) in Northern Chile Using Neural Networks Modeling. *Fishes* **2022**, *7*, 204. [[CrossRef](#)]
8. Brandoli, B.; Raffaetà, A.; Simeoni, M.; Adibi, P.; Bappee, F.K.; Pranovi, F.; Rovinelli, G.; Russo, E.; Silvestri, C.; Soares, A.; et al. From multiple aspect trajectories to predictive analysis: A case study on fishing vessels in the Northern Adriatic sea. *Geoinformatica* **2022**, *26*, 551–579. [[CrossRef](#)]
9. Mugo, R.; Saitoh, S.I.; Nihira, A.; Kuroyama, T. Habitat characteristics of skipjack tuna (*Katsuwonus pelamis*) in the western North Pacific: A remote sensing perspective. *Fish. Oceanogr.* **2010**, *19*, 382–396. [[CrossRef](#)]
10. Barkley, R.A.; Neill, W.H.; Gooding, R.M. Skipjack Tuna, *Katsuwonus Pelamz*s, Habitat Based on Temperature and Oxygen Requirements. *Fish. Bull.* **1978**, *76*, 653–662.
11. Zainuddin, M.; Nelwan, A.; Farhum, S.A.; Hajar, M.A.I.; Kurnia, M. Characterizing Potential Fishing Zone of Skipjack Tuna during the Southeast Monsoon in the Bone Bay-Flores Sea Using Remotely Sensed Oceanographic Data. *Int. J. Geosci.* **2013**, *4*, 259–266. [[CrossRef](#)]
12. Dueri, S.; Faugeras, B.; Maury, O. Modelling the skipjack tuna dynamics in the Indian Ocean with APECOSM-E: Part 1. Model formulation. *Ecol. Modell.* **2012**, *245*, 41–54. [[CrossRef](#)]
13. Lehodey, P.; Andre, J.M.; Bertignac, M.; Hampton, J.; Stoens, A.; Menkes, C.; Memery, L.; Grima, N. Predicting skipjack tuna forage distributions in the equatorial Pacific using a coupled dynamical bio-geochemical model. *Fish. Oceanogr.* **1998**, *7*, 317–325. [[CrossRef](#)]
14. Hsu, T.Y.; Chang, Y.; Lee, M.A.; Wu, R.F.; Hsiao, S.C. Predicting skipjack tuna fishing grounds in the Western and Central Pacific Ocean based on high-spatial-temporal-resolution satellite data. *Remote Sens.* **2021**, *13*, 861. [[CrossRef](#)]
15. Putri, A.R.S.; Zainuddin, M. Application of remotely sensed satellite data to identify Skipjack Tuna distributions and abundance in the coastal waters of Bone Gulf. *IOP Conf. Ser. Earth Environ. Sci.* **2019**, *241*, 012012. [[CrossRef](#)]
16. Mugo, R.; Saitoh, S.I.; Igarashi, H.; Toyoda, T.; Masuda, S.; Awaji, T.; Ishikawa, Y. Identification of skipjack tuna (*Katsuwonus pelamis*) pelagic hotspots applying a satellite remote sensing-driven analysis of ecological niche factors: A short-term run. *PLoS ONE* **2020**, *15*, e0237742. [[CrossRef](#)] [[PubMed](#)]
17. Dietterich, T.G. Ensemble methods in machine learning. In *International Workshop on Multiple Classifier Systems*; Springer: Berlin/Heidelberg, Germany, 2000; pp. 1–15.
18. Hunt, E.B. *Experiments in Induction*; Hunt, E.B., Marin, J., Stone, P.J., Eds.; Academic Press: New York, NY, USA, 1966.
19. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; Routledge: New York, NY, USA, 1984. [[CrossRef](#)]
20. Nelder, J.A.; Wedderburn, R.W. Generalized linear models. *J. R. Stat. Soc. Ser. A* **1972**, *135*, 370–384. [[CrossRef](#)]
21. Faraway, J.J. *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2016. [[CrossRef](#)]
22. Elith, J.; Graham, C.H.; Anderson, R.P.; Dudík, M.; Ferrier, S.; Guisan, A.; Hijmans, R.J.; Huettmann, F.; Leathwick, J.R.; Lehmann, A.; et al. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **2006**, *29*, 129–151. [[CrossRef](#)]
23. DeGaridel-Thoron, T.; Rosenthal, Y.; Bassinot, F.; Beaufort, L. Stable sea surface temperatures in the western Pacific warm pool over the past 1.75 million years. *Nature* **2005**, *433*, 294–298. [[CrossRef](#)]
24. Hampton, J. *Tuna Fisheries Status and Management in the Western and Central Pacific Ocean*; Oceanic Fisheries Programme; Secretariat of the Pacific Community: Noumea, New Caledoni, 2010; p. 23. Available online: <https://tuvalu-data.sprep.org/resource/tuna-fisheries-status-and-management-western-and-central-pacific-ocean-hampton-spc-2011> (accessed on 10 July 2020).
25. Laevastu, T.; Hayes, M.L. *Fisheries Oceanography and Ecology*; Fishing News Books Ltd.: Surrey, UK, 1981; Available online: <https://books.google.com.tw/books?id=-mNvQgAACAAJ> (accessed on 10 July 2020).
26. Scaillet, O. Density estimation using inverse and reciprocal inverse Gaussian kernels. *J. Nonparametr. Stat.* **2004**, *16*, 217–226. [[CrossRef](#)]
27. Weglarczyk, S. Kernel density estimation and its application. *ITM Web Conf.* **2018**, *23*, 37. [[CrossRef](#)]
28. Izenman, A.J. *Modern Multivariate Statistical Techniques*; Springer: London, UK, 2008; Available online: <https://link.springer.com/book/10.1007/978-0-387-78189-1> (accessed on 10 July 2020).
29. Wang, X.; Tsokos, C.P.; Saghafi, A. Improved parameter estimation of Time Dependent Kernel Density by using Artificial Neural Networks. *J. Financ. Data Sci.* **2018**, *4*, 172–182. [[CrossRef](#)]
30. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2018.
31. Gupta, A.K.; Johnson, B.E.; Nagar, D.K. Testing Equality of Several Correlation Matrices. *Rev. Colomb. Estadística* **2013**, *36*, 239–260.
32. Pham-Gia, T.; Choulakian, V. Distribution of the Sample Correlation Matrix and Applications. *Open J. Stat.* **2014**, *4*, 330–344. [[CrossRef](#)]

33. Leclere, J.; Oberdorff, T.; Belliard, J.; Leprieur, F. A comparison of modeling techniques to predict juvenile 0+ fish species occurrences in a large river system. *Ecol. Inform.* **2011**, *6*, 276–285. [[CrossRef](#)]
34. Manel, S.; Williams, H.C.; Ormerod, S.J. Evaluating presence–absence models in ecology: The need to account for prevalence. *J. Appl. Ecol.* **2001**, *38*, 921–931. [[CrossRef](#)]
35. Thuiller, W.; Araujo, M.B.; Lavorel, S. Generalized models vs. classification tree analysis: Predicting spatial distributions of plant species at different scales. *J. Veg. Sci.* **2003**, *14*, 669–680. [[CrossRef](#)]
36. Alexander, R.E. A Comparison of GLM, GAM, and GWR Modeling of Fish Distribution and Abundance in Lake Ontario. Ph.D. Thesis, University of Southern California, Los Angeles, CA, USA, 2016.
37. McKenna, J.E., Jr.; Castiglione, C. Model distribution of Silver Chub (*Macrhybopsis storeriana*) in western Lake Erie. *Am. Midl. Nat.* **2014**, *171*, 301–310. [[CrossRef](#)]
38. Gladju, J.; Kamalam, B.S.; Kanagaraj, A. Applications of data mining and machine learning framework in aquaculture and fisheries: A review. *Smart Agric. Technol.* **2022**, *2*, 100061. [[CrossRef](#)]
39. Zhao, S.; Zhang, S.; Liu, J.; Wang, H.; Zhu, J.; Li, D.; Zhao, R. Application of machine learning in intelligent fish aquaculture: A review. *Aquaculture* **2021**, *540*, 736724. [[CrossRef](#)]
40. Lynham, J.; Nikolaev, A.; Raynor, J.; Vilela, T.; Villaseñor-Derbez, J.C. Impact of two of the world’s largest protected areas on longline fishery catch rates. *Nat. Commun.* **2020**, *11*, 979. [[CrossRef](#)]
41. Nicol, S.; Menkes, C.; Jurado-Molina, J.; Lehodey, P.; Usu, T.; Kumasi, B.; Muller, B.; Bell, J.; Tremblay-Boyer, L.; Briand, K. Oceanographic characterisation of the Pacific Ocean and the potential impact of climate variability on tuna stocks and tuna fisheries. *SPC Fish. Newsl.* **2014**, *145*, 37–48.
42. Kim, J.; Na, H.; Park, Y.G.; Kim, Y.H. Potential predictability of skipjack tuna (*Katsuwonus pelamis*) catches in the Western Central Pacific. *Sci. Rep.* **2020**, *10*, 3193. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.