Saintek Perikanan: Indonesian Journal of Fisheries Science and Technology
Available at http://ejournal.undip.ac.id/index.php/saintek
Vol. 22 No. 3 : 241 - 248, October 2026

AN ARTIFICIAL NEURAL NETWORK APPROACH FOR PREDICTING PELAGIC FISH HABITATS IN FISHERIES MANAGEMENT AREAS (WPP) 573 USING MULTI-PARAMETER SATELLITE DATA

Rosyid Paundra Gamawan*[1], Nur Mohammad Farda[2], Prima Widayani[2], Favian Mafazi Giska Putra[3], Widia Pangestika[4], Ailsa Afra Mawarid[5], and Irfani Anis[1]
[1]Master Program in Remote Sensing, Faculty of Geography, Gadjah Mada University, Indonesia
[2]Department of Geographic Information Science, Faculty of Geography, Gadjah Mada University, Indonesia
[3]Research Center for Hydrodynamics Technology, National Research and Innovation Agency, Indonesia
[4]Nutrition Study Program, Faculty of Medicine and Health Sciences, Sultan Ageng Tirtayasa University, Indonesia
[5]Food Technology Laboratory, Integrated Laboratory Unit, Diponegoro University, Indonesia
Email: rosyidpaundragamawan@mail.ugm.ac.id

ABSTRACT

Fisheries Management Area (WPP) 573 is one of the upwelling regions of the Indian Ocean supporting both neritic small pelagic and oceanic pelagic fisheries, and is affected by complex climate variability. This research modeled spatial probabilities of fish pelagic habitat using an optimized shallow architecture of a Feedforward Network, specifically a Multi-layer Perceptron integrated with oceanographic satellite data. This research utilized 492 validated fishing coordinates from Global Fishing Watch (GFW) collected throughout 2025 as presence labels. At the same time, oceanographic variables included sea surface temperature, distance to the nearest coral reef, chlorophyll-a, salinity, sea surface elevation, current velocity, and bathymetry. To prioritize model generalization and mitigate overfitting, this research implemented the MLP with a single hidden layer containing three neurons, optimized via the L-BFGS algorithm with strong regularization. The final model with spatially- constrained pseudo-absences, environmental distance filtering obtained a True Skill Statistic (TSS) of 0.68, a 5-fold cross-validation accuracy of 83.95% and a presence precision of 95%, indicating that the model possess a good ability to discriminate among habitats and it is robust to variation. The permutation importance highlighted SST as the strongest predictor (0.1386), followed by salinity (0.0976) and current velocity (0.0866). Dynamic habitat probability maps generated reveal fish catch potential zones that correspond to the biophysical conditions of the water mass. In summary, this study demonstrates that MLP-based prediction, reinforced by stringent parameter optimization, is a very suitable approach to identify pelagic fish habitats, thus assisting in fishing operations and sustainable fishery management in WPP 573.

Keywords: Pelagic Fish Habitat, Multilayer Perceptron, Satellite Oceanography, Global Fishing Watch, WPP 573

INTRODUCTION

WPP 573, a constituent of the eastern Indian Ocean upwelling system, maintains pelagic fisheries that prosper in neritic and oceanic pelagic environments modified by coastal upwelling system dynamics, where productivity is upwelling driven through the monsoon winds and climate variability, such as the manifestation of the Indian Ocean Dipole modifying the strength of upwelling, stratification, and primary production (Abrahams *et al*., 2021; Narvekar *et al*., 2021; Vinayachandran *et al*., 2021). Comparable climate-driven shifts in upwelling have already altered pelagic distribution and biomass in regions such as West Africa and the Arabian Sea, threatening small-scale fisheries and the nutritional security of coastal communities reliant on pelagic catch for protein and micronutrients (Sarré *et al*., 2024; Diogoul *et al*., 2021; Sardenne *et al*., 2026). WPP 573 faces similar vulnerability, where climate-induced stock shifts directly affect catch stability and fisher livelihoods.

Traditional statistical approaches such as GAM–KDE integration have successfully identified tuna hotspots in Indonesian waters (e.g., WPP 713; Zainuddin *et al*., 2023), but parametric and semi-parametric models struggle with the strong multicollinearity and non-linear interactions among oceanographic variables (SST, salinity, current velocity) across multiple spatio-temporal scales (Cai *et al*., 2022; Stacey *et al*., 2021). Machine learning approaches—particularly Multi-Layer Perceptrons (MLP)—better capture these non-linear ecological thresholds and synergistic environmental interactions, offering a more robust basis for delineating pelagic habitat envelopes (Pickens *et al*., 2021; Ramampiandra *et al*., 2023). While MaxEnt remains a benchmark for presence-only modeling, it is increasingly prone to overfitting without careful complexity control, and alternatives such as Random Forest and boosted trees can match or exceed its performance under proper spatial cross-validation (Karp *et al*., 2022; Stacey *et al*., 2021). Random Forest, however, tends to overfit local spatial patterns and amplify spatial autocorrelation, limiting transferability to new regions without rigorous validation (Cai *et al*., 2022). MLPs, by contrast, extract latent spatial structures that point-based models overlook and have been shown to outperform classical regression in explaining biodiversity variation globally (Cai *et al*., 2022; Deneu *et al*., 2021).

In the Indian Ocean, SST, salinity, and sea surface height are key predictors of tuna and pelagic distribution, with monsoon-driven Ekman transport and Kelvin/Rossby wave dynamics producing SST and SSH anomalies linked to nutrient enrichment and elevated chlorophyll-a (Abrahams *et al*., 2021;

Received: 05-06-2026, Last Revised: 03-08-2026, Accepted: 07-08-2026, Available Online: 05-09-2026, Published: 05-09-2026

Vinayachandran *et al*., 2021; Wu & Wang, 2022) patterns consistent with the oceanographic characteristics of WPP 573. Vessel tracking data (AIS/VMS) are now a standard "biosensor" for identifying fishing grounds and pelagic habitat, as fishing effort is often more stable than CPUE as a habitat indicator (Paolo *et al*., 2024; Welch *et al*., 2022; Jemeļjanova *et al*., 2024). Because oceanographic variables differ widely in scale and contain noise, robust scaling and regularization are recommended for model stability (Chen *et al*., 2022; Jiang & Zhou, 2025), while permutation importance is an established interpretability tool ensuring predictions align with sound ecological reasoning (Srinivasu *et al*., 2024; Zhou *et al*., 2024).

Given the importance of pelagic habitat modeling for sustainable fisheries planning, this study evaluates an MLP's ability to classify pelagic habitats in WPP 573, identify the most influential oceanographic predictors via Permutation Importance, and assess model precision in delineating catch zones for operational use by fishers and managers. While MLPs are already used across marine SDM workflows (Coro *et al*., 2024; Selvaraj *et al*., 2022), the novelty here lies in a deliberately shallow architecture tuned with L-BFGS and high regularization (alpha = 5.0) to counter spatial overfitting, paired with a custom QGIS-based GUI toolbox bridging shallow-learning modeling and operational spatial fisheries management (Bald *et al*., 2023; Kass *et al*., 2021). Academically, this study contributes empirical evidence on MLP reliability for marine habitat modeling in the Indian Ocean; practically, it offers a predictive tool to improve catch accuracy and support fisheries management in WPP 573.

RESEARCH METHODS

Research Time and Location

Fish presence data were obtained from Global Fishing Watch (GFW), which derives commercial vessel movements from AIS/VMS signals. The data consist of 2025 fishing locations for 492 vessels (in WPP 573) obtained from the Global Fishing Watch AIS data, representing gear-specific vessel trajectories within eight gear types: drift gillnets, set gillnets, purse seines, other purse seines, trawls, other trawls, line gears and other line gears. To minimize the effect of spatial autocorrelation due to clustering of coastal data, spatial thinning was applied to the data by limiting the number of records to a maximum of one presence record per 8,905.6 m grid cell. GFW's ANN-based processing of AIS/VMS data (Paolo *et al*., 2024; Wu & Wang, 2022) and the use of activity points as catch-location proxies are consistent with established effort-distribution modeling practice (Karp *et al*., 2022; Rufener *et al*., 2021).

Pre-Processing (Vessel and Satellite Data)

Coordinates were projected to WGS 84 (EPSG:4326), converted to vector points in QGIS, and labeled "1" (presence) as the MLP target variable, following standard protocols for integrating VMS with bathymetric and oceanographic data (Kalina *et al*., 2022; McCarthy *et al*., 2022). Although GFW/AIS data is an imperfect proxy for biology, constrained by fuel costs and regulations, harbor noise was reduced by eliminating vessel activity from the data within 5 km of ports. Furthermore, by adding distance to port as a control predictor to the MLP model, we ensure that zero recorded effort reflects operational constraints, and not actual habitat unsuitability (McDonald *et al*., 2024; Ashrafi & Abe, 2021); however, sustained fishing concentration reliably identifies productive pelagic grounds (Kerry *et al*., 2022; Guiet *et al*., 2024), consistent with presence-only species distribution modeling that accounts for preferential sampling.

Seven environmental and geographical factors (Table 1), including Sea Surface Temperature (°C), Chlorophyll-a (mg/m³), Bathymetry (m), Distance to Nearest Coral Reef (km), Salinity (PSU), Sea Surface Elevation (m), and Current Velocity (m/s) were calculated as annual median composites, following a robust compositing approach shown to statistically filter transient cloud and shadow contamination while preserving persistent spatial signal (Simonetti *et al*., 2021). This preprocessing homogenizes spatial input and mitigates the heavy, persistent cloud cover that limits optical and infrared satellite observation during the Northwest Monsoon in Indonesian waters and producing a spatially continuous and cloud-free grid consistent with the use of static, long-term averaged environmental predictors to characterize baseline habitat suitability in species distribution modelling (Milanesi *et al*., 2020) to represent the macro-spatial template of persistent habitat potential in WPP 573. All rasters were standardized to EPSG:4326 and resampled to a common grid resolution of 8,905.6 m using Nearest Neighbor interpolation to preserve sharp gradients (e.g., reef boundaries, thermal fronts) without blurring, consistent with established satellite bathymetry and habitat-modeling practice (Babbel *et al*., 2021; Kass *et al*., 2021; McCarthy *et al*., 2022). Environmental values were extracted at each vessel coordinate in QGIS to build multivariate habitat profiles for WPP 573. A custom Python-based GUI toolbox was developed to standardize input selection and support replication across regions or timeframes, reflecting current trends toward automated raster–vector workflows in fisheries modeling (Barth *et al*., 2022; Behivoke *et al*., 2021).

Table 1. Satellite Variables and Their Utilization in Literature

Variable Type	Source and Primary Use	Practice in Literature	Source
SSE, Current, Salinity	HYCOM / Altimetry – Pelagic dynamics	Mapping habitats and stock dynamics	(Jang *et al*., 2022; Soltanolkotabi *et al*., 2017; Vinayachandran *et al*., 2021)
SST, CHL	MODIS – Productivity, Upwelling	Foundation of tuna habitat models	(Abrahams *et al*., 2021; Sarré *et al*., 2024; Dahms and Killen, 2023; Embury *et al*., 2024)
Bathymetry, Coral Reefs	ETOPO / Coral Atlas – Habitat structure	SDB approach and coastal applications	(Babbel *et al*., 2021; Caballero & Stumpf, 2023; McCarthy *et al*., 2022)

Model Development

Valid pixel extractions (excluding null/−9999 values) formed the tabular training dataset, following standard MLP/ANN spatial modeling protocols (Koldasbayeva *et al*., 2024; Zhu *et al*., 2022). Outliers were removed using the IQR method (1.5×IQR) to exclude sensor anomalies (Rustam *et al*., 2022). To keep the ratios of presence and pseudo-absence equal, 492 pseudo-absence points ('0') were drawn from areas of no fishing activity. In line with established background-sampling schemes in SDM, a 1.1 SD environmental envelope filter was applied as a data augmentation approach to limit the background points from looking like the real fishing grounds. (Zbinden *et al*., 2024; Descombes *et al*., 2022). Gaussian noise (0.12), tuned via cross-validation, was added to reduce overfitting to specific coordinates (Koldasbayeva *et al*., 2024; Bejani & Ghatee, 2021).

To avoid overfitting of the model to the moderately sized training data, the choice of hyperparameters was informed by the 5-fold cross-validation performance among a number of candidate configurations for the number of hidden neurons (3 to 10) and for the L2 regularization penalty (alpha = 0.1 to 5.0). The best configuration was determined as a shallow MLP with seven input neurons and a single hidden layer with three neurons (Banda & Kumarasamy, 2024; Chen *et al*., 2022; Rustam *et al*., 2022; Wu & Wang, 2022; Zhu *et al*., 2022), using tanh activation function with the L-BFGS solver for stable convergence (Bejani & Ghatee, 2021; Soltanolkotabi *et al*., 2017), strong L2 regularization (alpha = 5.0), and 1,500 iterations to further suppress spatial-noise overfitting (Lee *et al*., 2022; Wang *et al*., 2023). The inputs were scaled by using RobustScaler (Rustam *et al*., 2022; Zhu *et al*., 2022), and 5-fold cross-validation was used to verify the stability of the model.

Model Evaluation

Performance was assessed via 5-fold cross-validation using Accuracy, Precision, Recall, F1-score, and True Skill Statistic (TSS) yielding accuracy consistently above 80%—within the range considered good for ecological modeling (Koldasbayeva *et al*., 2024; Galappaththi *et al*., 2021; Zhu *et al*., 2022). Precision for the presence class was prioritized to minimize false positives and field-verification costs (Alotaibi & Nassif, 2024). Permutation Importance was applied to quantify variable contributions by measuring accuracy loss under random shuffling (Kadagi *et al*., 2021; Zhou *et al*., 2024), mirroring sensitivity-analysis practice in ecological SDM (Galappaththi *et al*., 2021; Chen *et al*., 2022). The trained model was then applied pixel-wise across WPP 573 using predict_proba() to generate a habitat suitability raster (.tif), consistent with established HSI mapping approaches in MLP-based pelagic distribution studies (Pickens *et al*., 2021; Koldasbayeva *et al*., 2024).

RESULT AND DISCUSSION

The MLP model's habitat suitability classification was assessed using a comprehensive evaluation matrix on a balanced sample of 492 points. Five-fold cross-validation produced a mean generalization accuracy of 83.95%, closely matching the training accuracy of 85.57% and an overall accuracy of 0.86; this minimal gap between training and validation performance, together with a True Skill Statistic (TSS) of 0.68, indicates that overfitting—a common obstacle in applying neural networks to environmental data—was effectively mitigated. Full precision, recall, and F1-score results are detailed in Table 2.

Table 2. MLP Model Performance Metrics

Metric	Presence (Label 1)	Absence (Label 0)	Overall Average
Precision	0.95	0.77	0.86
Recall	0.73	0.95	0.84
F1-Score	0.83	0.85	0.84
Cross-Val Accuracy	-	-	83.95%

For the "Presence" class (Label 1.0), the model achieved precision = 0.95, recall = 0.73, and F1-score = 0.83. This high precision is particularly valuable for fisheries applications, as it confirms accurate catch locations and minimizes false positives, while the comparatively lower recall reflects a deliberately conservative predictive stance that favors reliable hotspot identification over exhaustive but noisier habitat detection. For the "Absence" class (Label 0.0), the model achieved an F1-score of 0.85, with recall = 0.95 and precision = 0.77. Macro-average scores (precision = 0.86, recall = 0.84, F1-score = 0.84) were nearly identical to the weighted averages, indicating that the MLP architecture is well balanced and unbiased—supporting its reliability as a spatial decision-support tool for fisheries management in WPP 573.

When these metrics were aggregated, the system maintained a high macro average, yielding a precision of 0.87, a recall of 0.86, and an F1-score of 0.85. The structural stability of the model is further underscored by the identical weighted average scores across all three metrics. The perfect alignment between macro and weighted averages indicates that the MLP architecture is effectively balanced and unbiased, providing a dependable tool for spatial fisheries management in WPP 573.

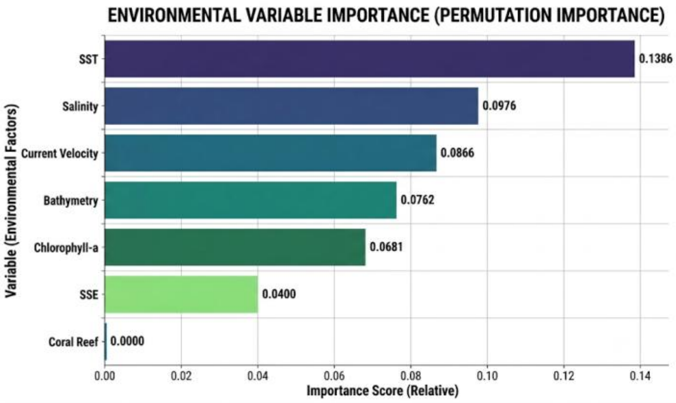

Figure 1. Environmental Variable Contribution Based on Permutation Importance

Sea Surface Temperature (SST) ranks as the model's most dominant predictor, according to the Permutation Importance analysis (Figure 1). The final output of the MLP model was high-resolution spatial prediction maps for WPP 573. By applying the predict_proba() function across the rasterized environmental layers, the model produced a continuous probability surface of habitat suitability. The resulting maps visually synthesize the dominance of SST and salinity; areas with higher probability scores (approaching 1.0) consistently align with specific thermal fronts and current-driven upwelling zones. This spatial projection transforms complex multivariate interactions into an actionable visual tool, effectively delineating "high-potential" fishing grounds where the convergence of physical variables creates optimal metabolic conditions. This outcome validates the model's utility as a dynamic decision-support system for operational fisheries management. This zero-value contribution internally validates the model's spatial selectivity.

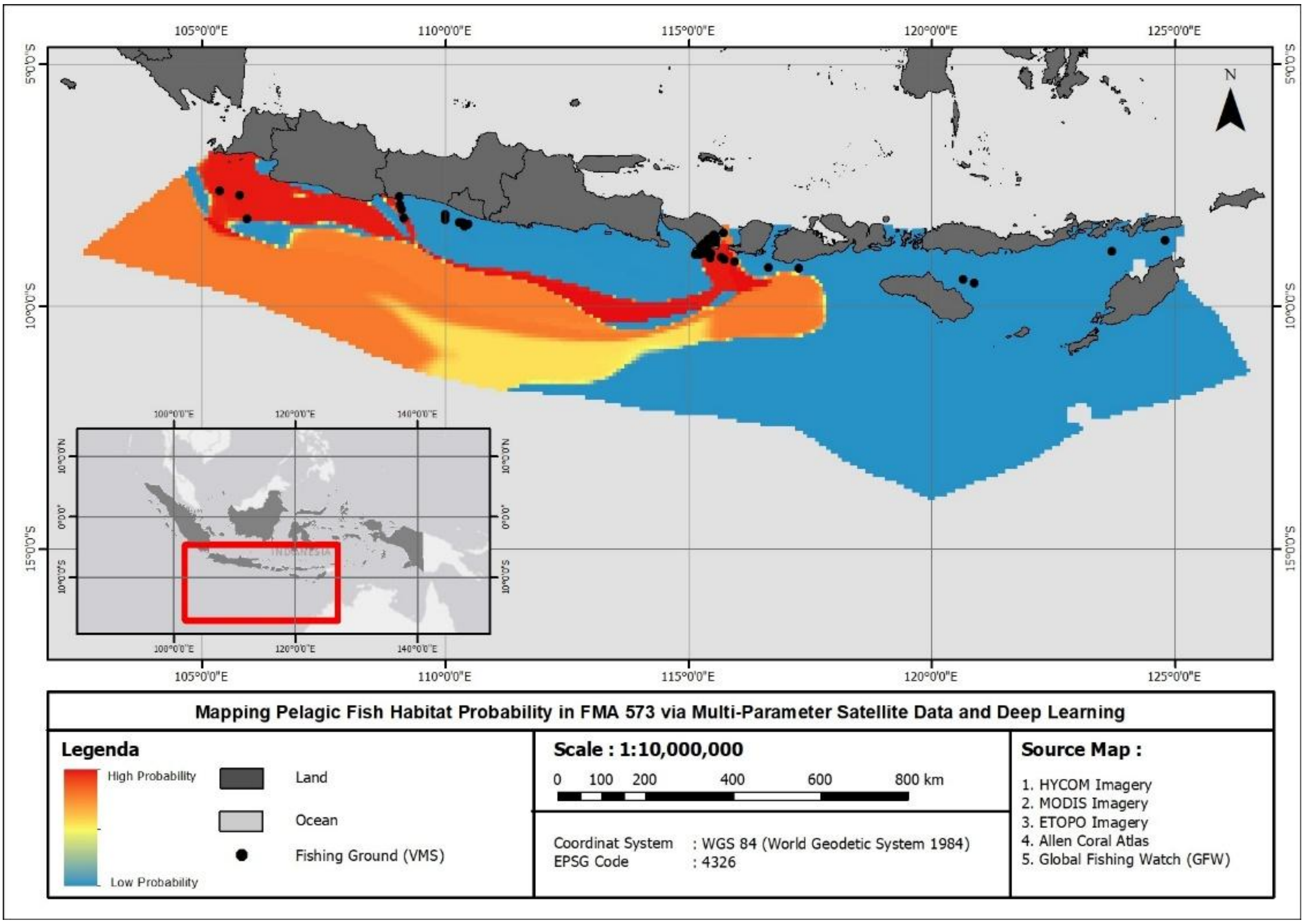

Figure 2. Spatial Prediction Map of Pelagic Habitat Probability in WPP 573.

The spatial output of MLP model shows that potential habitats are distributed in a very organised manner in the entire study area. A clear cluster of high probability zones (red shaded zones) is observed along the coastal and inner shelf waters off southern Java, Bali, and Nusa Tenggara. The spatial patterns illustrated in Figure 2 delineate an optimal habitat suitability corridor along the southern coastline of these islands. Sea Surface Temperature (SST) and current velocity dynamics at the ideal thresholds for pelagic fish aggregation are effectively captured by the concentration of high-probability zones at specific oceanographic convergences. By synthesizing complex multivariate interactions into this actionable visualization, the model provides a robust empirical basis for identifying potential fishing grounds. These results suggest that operational efficiency for the fishing industry can be significantly enhanced by focusing efforts within these identified corridors, thereby reducing search time and minimizing associated operational costs.

The spatial analysis in Figure 2 further reveals that intermediate to high-probability zones (indicated by orange gradients) are extensively distributed from the waters of Banten to Bali, covering a broader spatial extent toward the southern Indian Ocean. This widespread distribution suggests a larger corridor of favorable conditions throughout the central portion of the study area.

Model Performance

Five-fold cross-validation yielded a mean accuracy of 83.95%, close to the training accuracy of 85.57%—a gap under 2% indicating stable convergence rather than overfitting, consistent with standards for reliable spatial MLP projections (Koldasbayeva *et al*., 2024; Galappaththi *et al*., 2021; Zhu *et al*., 2022). This stability supports the parsimonious architecture (one hidden layer, three neurons) combined with high regularization (alpha = 5.0), which together forced the model to learn broad ecological patterns rather than memorize noise, consistent with established guidance on complexity control in shallow neural networks (Park & Lek, 2016; Bejani & Ghatee, 2021). The result supports confident large-scale spatial deployment across WPP 573. The Presence class achieved a Precision of 0.95, indicating that 95% of predicted habitat areas matched historical VMS catch locations—a deliberate strategic emphasis in industrial fisheries, where false positives translate directly into wasted fuel and labor (Alotaibi & Nassif, 2024). The corresponding Recall of 0.75 reflects a conservative trade-off: roughly 25% of potential habitat may be missed, but the areas the model does recommend are highly reliable and actionable for fishers.

Environmental Variable Contribution

Permutation Importance identified Sea Surface Temperature (SST, score 0.1386) as the dominant predictor, consistent with a meta-analysis of 224 marine fish distribution studies showing physical variables (SST, salinity, SSH) outweigh biological variables such as chlorophyll-a for pelagic guilds (Pickens *et al*., 2021), and with permutation-based hierarchies reported for Northeast Atlantic sardines (SST ~31%, salinity ~21%, current–coastline interaction ~19%; Lima *et al*., 2022). This physical dominance reflects Southeast Monsoon-driven upwelling in WPP 573: alongshore winds induce offshore Ekman transport that draws cool, saline, nutrient-rich water to the surface, creating thermal/salinity fronts that pelagic fish track as cues for prey aggregation, while current velocity concentrates planktonic food into retention zones—patterns also documented off South Java and the Lesser Sunda Islands (Mandal *et al*., 2022; Simanjuntak and Lin, 2022). In contrast, the null importance of Coral Reef (0.0000) is consistent with known guild divisions, with reef associated predictors being important for reef fish but not for offshore pelagic species (Pickens *et al*., 2021). Rather than being a flaw, this zero-value score provides a powerful internal validation of the MLP model spatial selectivity and ecological realism, indicating that the network autonomously inhibited and removed non-informative benthic features without necessitating manual feature pruning and thus properly distinguished benthic-coastal from pelagic environment.

Pelagic Habitat Probability

Overlaying training points (VMS/catch locations) on the prediction surface confirmed that most presence points fall within or near high-probability zones, mirroring standard HSI validation practice for tuna and sardine habitat models (Alotaibi and Nassif, 2024; Lima *et al*., 2022; Jiang and Zhou, 2025). These hotspots coincide with current convergences and thermal fronts and are intended to delineate priority operational zones rather than direct tonnage estimates, consistent with global pelagic modeling standards for decision support in fisheries operations and spatial planning. Because the model is calibrated on 2025 annual medians to capture a macro-spatial baseline, predicted hotspots do not reflect intra-annual seasonal shifts driven by monsoon cycles (e.g., Southeast upwelling vs. Northwest monsoon) and areexpected to shift under interannual climate anomalies such as El Niño/Positive IOD (intensified upwelling, expanded suitability) versus La Niña/Negative IOD (suppressed upwelling, contracted suitability)—though the direction of these shifts is not uniform across species or regions: negative IOD has been shown to shrink mackerel tuna habitat off Java (Koropitan *et al*., 2021), while increasing small pelagic production elsewhere in the eastern Indian Ocean (Lumban-Gaol *et al*., 2021). This sensitivity means the model will transfer imperfectly to anomaly years, consistent with evidence that MLPs show the largest performance variation among SDM algorithms under changing climate regimes (Lee *et al*., 2022). Given an estimated pelagic potential of ~978,581 tons in WPP 573 (KKP data), these suitability maps support two functions: directing fishing effort toward high-probability zones within TAC limits (Alotaibi & Nassif, 2024; Jiang & Zhou, 2025), and supporting adaptive management by linking future shifts in SST or current velocity to potential habitat relocation (Lima *et al*., 2022; Deneu *et al*., 2021).

CONCLUSION

This study concludes that an optimized shallow Multi-Layer Perceptron (MLP) architecture, only three neurons and strong regularization (Alpha 5.0), effectively classifies pelagic habitats in WPP 573 with stable generalization accuracy of 83.95%. The minimal gap between training (85.57%) and testing performance ensures the model's robustness against overfitting, demonstrating that a streamlined network is sufficient for capturing complex, non-linear oceanographic relationships. Sea Surface Temperature (SST) emerged as the most significant predictor (0.1386), followed by salinity (0.0976) and current velocity dynamics (0.0866), indicating that physical water mass characteristics primarily drive fish distribution in this region. Crucially, the model achieved a 95% precision for the presence class, providing a reliable, low-risk tool for delineating potential catch zones and mitigating operational costs for fishers by minimizing "false positive" hotspots. While integrating these predictions into a custom QGIS Toolbox enhances practical marine management, future studies must incorporate independent fisheries datasets and validate the model against inter-annual climate variability, such as ENSO and IOD, before full-scale operational deployment.

ACKNOWLEDGEMENT

This research is part of projects titled "Establishment of the Integrated Ocean Fisheries Technology Training Center and the Enhancement of Capacity Building in Indonesia (1192000-2023-006)" which are funded by the Ministry of Oceans and Fisheries, Korea.

REFERENCES

Abrahams, A., Schlegel, R., and Smit, A. (2021). Variation and change of upwelling dynamics detected in the world's eastern boundary upwelling systems. *Frontiers in Marine Science,* 8, 626411.

Alotaibi, E., and Nassif, N. (2024). Artificial intelligence in environmental monitoring: in-depth analysis. *Discover Artificial Intelligence*, 4, 1981.

Ashrafi, T. A., and Abe, K. (2021). Intra- and inter-temporal effort allocation and profit-maximizing strategy of trawl fishery. *Ices Journal of Marine Science*.

Babbel, B., Parrish, C., and Magruder, L. (2021). ICESat-2 elevation retrievals in support of satellite-derived bathymetry for global science applications. *Geophysical Research Letters*, 48, e2020GL090629.

Bald, L., Gottwald, J., and Zeuss, D. (2023). spatialMaxent: Adapting species distribution modeling to spatial data. *Ecology and Evolution*, 13.

Banda, T., and Kumarasamy, M. (2024). Artificial neural network (ANN)-based water quality index (WQI) for assessing spatiotemporal trends in surface water quality—A case study of South African river basins. *Water*, 16, 111485.

Barth, A., Álvera-Azcárate, A., Troupin, C., and Beckers, J. (2022). DINCAE 2.0: multivariate convolutional neural network with error estimates to reconstruct sea surface temperature satellite and altimetry observations. *Geoscientific Model Development*, 15, 2183–2202.

Behivoke, F., Étienne, M., Guitton, J., Randriatsara, R., Ranaivoson, E., and Léopold, M. (2021). Estimating fishing effort in small-scale fisheries using GPS tracking data and random forests. *Ecological Indicators*, 121, 107321.

Bejani, M., and Ghatee, M. (2021). A systematic review on overfitting control in shallow and deep neural networks. *Artificial Intelligence Review*, 54, 6391–6438.

Caballero, I., and Stumpf, R. (2023). Confronting turbidity, the major challenge for satellite-derived coastal bathymetry. *Science of the Total Environment*, 881, 161898.

Cai, L., Kreft, H., Taylor, A., Denelle, P., Schrader, J., Essl, F., ... Weigelt, P. (2022). Global models and predictions of plant diversity based on advanced machine learning techniques. *New Phytologist*, 237, 1432–1445.

Chen, J., Gong, X., Guo, X., Xing, K., Lu, K., Gao, H., and Gong, X. (2022). Improved perceptron of subsurface chlorophyll maxima by a deep neural network: A case study with BGC-Argo float data in the Northwestern Pacific Ocean. *Remote Sensing*, 14, 632.

Coro, G., Sana, L., and Bove, P. (2024). An open science automatic workflow for multi-model species distribution estimation. *International Journal of Data Science and Analytics*, 20, 1131 - 1150.

Dahms, C., and Killen, S. (2023). Temperature change effects on marine fish range shifts: A meta-analysis of ecological and methodological predictors. *Global Change Biology*, 29, 4459–4479.

Deneu, B., Servajean, M., Bonnet, P., Botella, C., Munoz, F., and Joly, A. (2021). Convolutional neural networks improve species distribution modelling by capturing the spatial structure of the environment. *PLoS Computational Biology*, 17, e1008856.

Descombes, P., Chauvier, Y., Brun, P., Righetti, D., Wüest, R. O., Karger, D., Zurell, D., and Zimmermann, N. (2022). Strategies for sampling pseudo-absences for species distribution models in complex mountainous terrain. *bioRxiv*.

Diogoul, N., Brehmer, P., Demarcq, H., Ayoubi, S., Thiam, A., Sarré, A., ... Perrot, Y. (2021). On the robustness of an eastern boundary upwelling ecosystem exposed to multiple stressors. *Scientific Reports*, 11, 1–13.

Embury, O., Merchant, C., Good, S., Rayner, N., Høyer, J., Atkinson, C., ... Donlon, C. (2024). Satellite-based time-series of sea-surface temperature since 1980 for climate applications. *Scientific Data*, 11, 326.

Galappaththi, E., Susarla, V., Loutet, S., Ichien, S., Hyman, A., and Ford, J. (2021). Climate change adaptation in fisheries. *Fish and Fisheries*, 22, 1141–1159.

Guiet, J., Bianchi, D., Scherrer, K., Heneghan, R., and Galbraith, E. (2024). BOATSv2: new ecological and economic features improve simulations of high seas catch and effort. *Geoscientific Model Development*.

Jang, E., Kim, Y., Im, J., Park, Y., and Sung, T. (2022). Global sea surface salinity via the synergistic use of SMAP satellite and HYCOM data based on machine learning. *Remote Sensing of Environment*, 273, 112980.

Jemeļjanova, M., Kmoch, A., and Uuemaa, E. (2024). Adapting machine learning for environmental spatial data - A review. *Ecological Informatics*, 81, 102634.

Jiang, B., and Zhou, W. (2025). Fishing operation type recognition based on multi-branch convolutional neural network using trajectory data. *PeerJ Computer Science*, 11, e3020.

Kadagi, N., Wambiji, N., Fennessy, S., Allen, M., and Ahrens, R. (2021). Challenges and opportunities for sustainable development and management of marine recreational and sport fisheries in the Western Indian Ocean. *Marine Policy*, 124, 104351.

Kalina, J., Tumpach, J., and Holeňa, M. (2022). On combining robustness and regularization in training multilayer perceptrons over small data. *Proceedings of the 2022 IJCNN*, 1–8.

Karp, M., Brodie, S., Smith, J., Richerson, K., Selden, R., Liu, O., ... Jacox, M. (2022). Projecting species distributions using fishery-dependent data. *Fish and Fisheries*, 24, 212–231.

Kass, J., Muscarella, R., Galante, P., Bohl, C., Pinilla-Buitrago, G., Boria, R., ... Anderson, R. (2021). ENMeval 2.0: Redesigned for customizable and reproducible modeling of species' niches and distributions. *Methods in Ecology and Evolution*, 12, 1602–1608.

Kerry, C., Exeter, O. M., and Witt, M. (2022). Monitoring global fishing activity in proximity to seamounts using automatic identification systems. *Fish and Fisheries*.

Koldasbayeva, D., Tregubova, P., Gasanov, M., Zaytsev, A., Petrovskaia, A., and Burnaev, E. (2024). Challenges in data-driven geospatial modeling for environmental research and practice. *Nature Communications*, 15, 1–14.

Koropitan, A., Kholilullah, I., and Yusfiandayani, R. (2021). Modeling Mackerel Tuna (Euthynnus affinis) Habitat in Southern Coast of Java: Influence of Seasonal Upwelling and Negative IOD. *HAYATI Journal of Biosciences*.

Lee, W., Song, J.-W., Yoon, S., and Jung, J.-M. (2022). Spatial Evaluation of Machine Learning-Based Species Distribution Models for Prediction of Invasive Ant Species Distribution. *Applied Sciences*, 12.

Lima, A., Baltazar-Soares, M., Garrido, S., Riveiro, I., Carrera, P., Piecho-Santos, A., ... Silva, G. (2022). Forecasting shifts in habitat suitability across the distribution range of a temperate small pelagic fish under different scenarios of climate change. *Science of the Total Environment*, 804, 150167.

Lumban-Gaol, J., Siswanto, E., Mahapatra, K., Natih, N. M., Nurjaya, I., Hartanto, M. T., Maulana, E., Adrianto, L., Rachman, H. A., Osawa, T., Rahman, B. M. K., and Permana, A. (2021). Impact of the Strong Downwelling (Upwelling) on Small Pelagic Fish Production during the 2016 (2019) Negative (Positive) Indian Ocean Dipole Events in the Eastern Indian Ocean off Java. *Climate*, 9.

Mandal, S., Susanto, R., and Ramakrishnan, B. (2022). On Investigating the Dynamical Factors Modulating Surface Chlorophyll-a Variability along the South Java Coast. *Remote Sensing*, 14, 1745.

McCarthy, M., Otis, D., Hughes, D., and Müller-Karger, F. (2022). Automated high-resolution satellite-derived coastal bathymetry mapping. *International Journal of Applied Earth Observation and Geoinformation*, 107, 102693.

McDonald, G. G., Bone, J., Costello, C., Englander, G., and Raynor, J. (2024). Global expansion of marine protected areas and the *redistribution of fishing effort. Proceedings of the National Academy of Sciences, 121.*

Milanesi, P., Della Rocca, F., & Robinson, R. A. (2020). Integrating dynamic environmental predictors and species occurrences: Toward true dynamic species distribution models. Ecology and Evolution, 10*(2), 1087–1092.*

Narvekar, J., Chowdhury, R., Gaonkar, D., Kumar, P., and Kumar, P. (2021). Observational evidence of stratification control of upwelling and pelagic fishery in the eastern Arabian Sea. Scientific Reports, 11, 7523.

Paolo, F., Kroodsma, D., Raynor, J., Hochberg, T., Davis, P., Cleary, J., ... Halpin, P. (2024). Satellite mapping reveals extensive industrial activity at sea. *Nature*, 625, 85–91.

Park, Y., and Lek, S. (2016). Artificial neural networks: Multilayer perceptron for ecological modeling. In S. E. Jørgensen (Ed.), *Developments in Environmental Modelling* (Vol. 28, pp. 123–140). Elsevier.

Pickens, B., Carroll, R., Schirripa, M., Forrestal, F., Friedland, K., and Taylor, J. (2021). A systematic review of spatial habitat associations and modeling of marine fish distribution: A guide to predictors, methods, and knowledge gaps. *PLoS ONE*, 16, e0251818.

Ramampiandra, E., Scheidegger, A., Wydler, J., and Schuwirth, N. (2023). A comparison of machine learning and statistical species distribution models: Quantifying overfitting supports model interpretation. *Ecological Modelling*, 481, 110353.

Rufener, M., Kristensen, K., Nielsen, J., and Bastardie, F. (2021). Bridging the gap between commercial fisheries and survey data to model the spatiotemporal dynamics of marine species. *Ecological Applications*, 31, e02453.

Rustam, F., Ishaq, A., Kokab, S., Díez, I., Mazón, J., Rodríguez, C., and Ashraf, I. (2022). An artificial neural network model for water quality and water consumption prediction. *Water*, 14, 213359.

Sardenne, F., Munaron, J. M., Geja, Y., Jaffrezic, E., Vagner, M., Pecquerie, L., and van der Lingen, C. (2026). Long-chain omega-3 fatty acids and metallic elements in small pelagic fish from South Africa from a human consumption perspective. *Journal of Food Composition and Analysis*, 152, 109018.

Sarré, A., Demarcq, H., Keenlyside, N., Krakstad, J., Ayoubi, S., Jeyid, A., ... Brehmer, P. (2024). Climate change impacts on small pelagic fish distribution in Northwest Africa: trends, shifts, and risk for food security. *Scientific Reports*, 14, 10565.

Selvaraj, J., Rosero-Henao, L. V., and Cifuentes-Ossa, M. A. (2022). Projecting future changes in distributions of small-scale pelagic fisheries of the southern Colombian Pacific Ocean. *Heliyon*, 8.

Simanjuntak, F., and Lin, T.-H. (2022). Monsoon Effects on Chlorophyll-a, Sea Surface Temperature, and Ekman Dynamics Variability along the Southern Coast of Lesser Sunda Islands and Its Relation to ENSO and IOD Based on Satellite Observations. Remote Sensing, 14, 1682.

Simonetti, D., Pimple, U., Langner, A., & Marelli, A. (2021). Pan-tropical Sentinel-2 cloud-free annual composite datasets. Data in Brief, 39, 107488.

Soltanolkotabi, M., Javanmard, A., and Lee, J. (2017). Theoretical insights into the optimization landscape of over-parameterized shallow neural networks. *IEEE Transactions on Information Theory*, 65, 742–769.

Srinivasu, P., Lakshmi, J., Gudipalli, A., Narahari, S., Shafi, J., Woźniak, M., and Ijaz, M. (2024). XAI-driven CatBoost multi-layer perceptron neural network for analyzing breast cancer. *Scientific Reports*, 14, 27620.

Stacey, N., Gibson, E., Loneragan, N., Warren, C., Wiryawan, B., Adhuri, D., ... Fitriana, R. (2021). Developing sustainable small-scale fisheries livelihoods in Indonesia: Trends, enabling and constraining factors, and future opportunities. *Marine Policy*, 132, 104654.

Vinayachandran, P., Masumoto, Y., Roberts, M., Hugget, J., Halo, I., Chatterjee, A., ... Hood, R. (2021). Reviews and syntheses: Physical and biogeochemical processes associated with upwelling in the Indian Ocean. *Biogeosciences*, 18, 5967–6013.

Wang, S., Shin, M., and Bai, R. (2023). Generative Quantile Regression with Variability Penalty. *Journal of Computational and Graphical Statistics*, 33, 1202 - 1213.

Welch, H., Clavelle, T., White, T., Cimino, M., Van Osdel, J., Hochberg, T., ... Hazen, E. (2022). Hot spots of unseen fishing vessels. *Science Advances*, 8, eabq2109.

Wu, J., and Wang, Z. (2022). A hybrid model for water quality prediction based on an artificial neural network, wavelet transform, and long short-term memory. *Water*, 14, 040610.

Zainuddin, M., Safruddin, S., Farhum, A., Budimawan, B., Hidayat, R., Selamat, M., ... Ihsan, Y. (2023). Satellite-based ocean color and thermal signatures defining habitat hotspots and the movement pattern for commercial skipjack tuna in Indonesia Fisheries Management Area 713, Western Tropical Pacific. *Remote Sensing*, 15, 1268.

Zbinden, R., Van Tiel, N., Kellenberger, B., Hughes, L., and Tuia, D. (2024). On the selection and effectiveness of pseudo-absences for species distribution modeling with deep learning. *ArXiv, abs/2401.02989*.

Zhou, W., Yan, Z., and Zhang, L. (2024). A comparative study of 11 non-linear regression models highlighting autoencoder, DBN, and SVR, enhanced by SHAP importance analysis in soybean branching prediction. *Scientific Reports*, 14, 55243.

Zhu, M., Wang, J., Yang, X., Zhang, L., Ren, H., Wu, B., and Ye, L. (2022). A review of the application of machine learning in water quality evaluation. *Eco-Environment and Health*, 1, 107–116.